

**Національна академія педагогічних наук України  
Інститут обдарованої дитини**

**Комп'ютерні онтології та їх використання у  
навчальному процесі. Теорія і практика  
МОНОГРАФІЯ**

**Київ 2013**

УДК 004.89:681.581:374:7.011:930.85  
ББК 32.965  
К63

*Рекомендовано до друку вченою радою Інституту обдарованої дитини  
НАПН України (протокол № 2 від 27.02.2013 р.).*

**Рецензенти:**

Широков В.А. – академік Національної академії наук України, доктор технічних наук

Гуржий А.М. – академік Національної академії педагогічних наук, професор, доктор технічних наук

**Автори:**

Довгий С.О., Величко В. Ю., Глоба Л. С., Стрижак О. Є., Андрущенко Т. І., Гальченко С. А, Гончар А. В., Гуляєв К. Д., Кудляк В.М., Ляшук К.В., Палагін О. В., Петренко М. Г., Попова М. А., Сидоренко В. І., Слюсаренко О. О., Стус Д.В., Терновой М. Ю.

**К63 Комп'ютерні онтології та їх використання у навчальному процесі. Теорія і практика. : Монографіям / С. О. Довгий, В. Ю. Величко, Л. С. Глоба, О. Є. Стрижак., Т. І. Андрущенко, С. А. Гальченко, А. В. Гончар, К. Д. Гуляєв, В. М. Кудляк, К. В. Ляшук, О. В. Палагін, М. Г. Петренко, М. А. Попова, В. І. Сидоренко, О. О. Слюсаренко, Д. В. Стус, М. Ю. Терновой. – К. : Інститут обдарованої дитини, 2013. – 310 с.**

ISBN 978-966-2633-15-3

У монографії викладаються основи створення інформаційно-програмних та методичних засобів побудови онтологічних моделей опису об'єктів та процесів предметних областей при формуванні загальної мережі знань. Описуються засади та засоби вирішення актуальних проблем підвищення ефективності підтримки процесу дослідницької діяльності талановитої молоді, що включають науково-методологічні засади та сучасні інформаційні технології, які забезпечують розробку формалізованої системи знань в конкретних предметних галузях. Визначаються методи проведення міждисциплінарних досліджень з використанням певних онтологій.

Наведено приклади та фрагменти програм опрацювання онтологічних даних (ПООД), що входять до інструментального комплексу і підтримують процес автоматизованої побудови онтологічних структур.

ISBN 978-966-2633-15-3

**УДК 004.89:681.581:374:7.011:930.85  
ББК 32.965**

© Інститут обдарованої дитини, 2013

## ЗМІСТ

ВСТУП.....	9
РОЗДІЛ 1. КЕРУВАННЯ ЗНАННЯМИ - НЕВІД'ЄМНА ЧАСТИНА НАВЧАЛЬНОГО ПРОЦЕСУ .....	11
1.1. Аналіз систем керування знаннями.....	11
1.2. Принципи створення систем керування знаннями.....	16
1.3. Основні принципи організації дослідницької діяльності .....	19
1.4. Огляд систем та пошукових платформ, які використовуються в світі для створення систем керування знаннями .....	21
1.5. Місце систем керування знаннями в освітніх проектах.....	37
РОЗДІЛ 2. КОНЦЕПТУАЛЬНІ АСПЕКТИ ПОБУДОВИ МОДЕЛЕЙ ОНТОЛОГІЧНОГО ОПИСУ ОБ'ЄКТІВ І ПРОЦЕСІВ.....	47
2.1. Дослідження предметно-орієнтованих знань.....	47
2.2. Концептуальна парадигма роботи із знаннями.....	52
2.3. Проблеми, методи та засоби обробки природномовних об'єктів .....	62
2.4. Формалізовані підходи добування та представлення знань .....	65
2.5. Формальна постановка задачі добування знань з ПМТ .....	66
2.6. Комп'ютерні онтології ПдО.....	67
2.7. Лінгвістичні онтології.....	72
2.8. Системна інтеграція лінгвістичних і предметних онтологій.....	85
2.9. Методологія розробки онтології ПдО.....	90
РОЗДІЛ 3. ЛІНГВІСТИЧНІ МОДЕЛІ ОБРОБКИ ПРИРОДНОМОВНИХ ТЕКСТІВ (ПМТ).....	102
3.1. Загальний підхід до проблеми аналізу ПМТ.....	102
3.2. Етапи і моделі аналізу ПМТ.....	105
3.3. Загальні принципи та нормаль на постановка задачі аналізу ПМТ... ..	108
3.4. Методологічні засади проектування ОКІС з обробкою предметно- орієнтованих знань.....	119
3.5. Використання онтологічного опису предметних областей при створенні інформаційних систем та роботи з контентом .....	129
3.6. Застосування теорії онтологій предметних областей для структурування інформації, яка представлена в Інтернеті .....	131
3.7. Інформаційна модель етапів лінгвістичного аналізу.....	135
3.8. Алгоритми семантичної інтерпретації морфологічних властивостей одиниць природної мови .....	143
3.9. Моделі представлення багатослівних термінів предметної області... ..	146
3.10. Алгоритм автоматичного виділення багатослівних термінів з природномовних об'єктів предметної області .....	151
РОЗДІЛ 4. МЕТОДОЛОГІЧНІ ОСНОВИ СТВОРЕННЯ СИСТЕМ УПРАВЛІННЯ ЗНАННЯМИ .....	154
4.1. Особливості систем управління знаннями .....	154
4.2. Систематизація і аналіз методів обробки інформації в системах управління знаннями.....	156

4.3. Обробка інформації на основі підсистеми нечіткого виведення загального виду .....	160
4.4. Проблеми використання неупорядкованої бази знань .....	164
4.5. Методи та способи завдання функцій належності.....	168
<b>РОЗДІЛ 5. МЕТОДИ ОБРОБКИ ІНФОРМАЦІЇ В СИСТЕМАХ УПРАВЛІННЯ ЗНАННЯМИ .....</b>	<b>174</b>
5.1. Зведення бази нечітких знань до деревоподібної структури.....	174
5.2. Модифікація підсистеми нечіткого виведення .....	183
5.3. Модифікація схеми нечіткого виведення .....	183
5.4. Модифікація структури підсистеми нечіткого виведення .....	186
5.5. Представлення верхньої частини бази нечітких знань як класичної логічної .....	188
5.6. Метод обробки інформації на основі деревоподібних баз нечітких знань зі змішаною схемою виведення .....	192
<b>РОЗДІЛ 6. МЕТОДИКА СТВОРЕННЯ ДЕРЕВОПОДІБНОЇ БАЗИ НЕЧІТКИХ ЗНАНЬ .....</b>	<b>196</b>
6.1. Підхід до побудови та настроювання бази нечітких знань.....	196
6.2. Представлення бази нечітких знань у вигляді нейромережі .....	198
6.3. Формування структури бази нечітких знань .....	201
6.4. Настроювання функцій належності термів лінгвістичних змінних нижнього рівня .....	207
<b>РОЗДІЛ 7. ЗАСОБИ ОБРОБКИ ОБРОБКИ ІНФОРМАЦІЇ У СИСТЕМАХ УПРАВЛІННЯ ЗНАННЯМИ .....</b>	<b>209</b>
7.1. Основні положення гнучкої технології створення програмних модулів обробки інформації .....	209
7.2. Інструментальні засоби гнучкої технології .....	210
7.3. Прив'язка до предметної області.....	212
7.4. Автоматична побудова запиту до бази даних .....	219
<b>РОЗДІЛ 8. МЕТОДИКА ПРОВЕДЕННЯ КОНТЕНТ-АНАЛІЗУ ЛЕКСИКИ ХУДОЖНЬОГО ТВОРУ НА ОСНОВІ ВИКОРИСТАННЯ ФУНКЦІЙ СИСТЕМИ КОНСПЕКТ (ЗА ТЕКСТАМИ ЗБІРКИ Т. Г. ШЕВЧЕНКА «КОБЗАР»).....</b>	<b>226</b>
8.1. Програма обробки лінгвістичних даних “Конспект” .....	226
8.2. Методика обробки текстів у середовищі системи Конспект.....	233
<b>РОЗДІЛ 9. МЕТОДИКА ПОБУДОВИ ОНТОЛОГІЇ (НА ПРИКЛАДІ ОПISУ ЖИТТЕДІЯЛЬНОСТІ ТАРАСА ГРИГОРОВИЧА ШЕВЧЕНКО).</b>	<b>240</b>
9.1. Система обробки предметних знань “Конфор” .....	240
9.2. Таксономічне відображення.....	251
9.3. Методика застосування.....	262
Висновки .....	277
ЛІТЕРАТУРА.....	278

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ СКОРОЧЕНЬ І ТЕРМІНІВ

АВІ	автоматичне визначення інструменту
БД	база даних
БЗ	база знань
ВДПЛ	віртуальна дослідно-проектувальна лабораторія
ВКС	віртуальне комунікаційне середовище
ВМ	віртуальна машина
ВСЛ	віртуальне середовище лабораторії
ДН	Дистанційне навчання
ЕБ	електронна бібліотека
ЕК	електронний курс
ЕлК	електронна колекція
ЕнС	енциклопедичний словник
ЗОІС	знання-орієнтована інформаційна система
ІС	інтелектуальна інформаційна система
ІТ	інтелектуальна інформаційна технологія
ІКОП	інструментальний комплекс онтологічного призначення
ІнЗ	інструментальні засоби
ІК	інтерфейс користувача
ІР	інформаційний ресурс
ІТ	інформаційна технологія
КГ	концептуальні графи
КГО	керуюча графічна оболонка
КЛП	Когнітивний лінгвістичний процесор
КО	комп'ютерна онтологія
КС	комп'ютерна система
ЛБД	лексикографічна база даних
ЛКТ	лінгвістичний корпус текстів
ЛОМ	Локальна обчислювальна мережа
ЛО ПдО	Лінгвістична онтологія предметної області
ЛПТЛ	лінійна пропозиціональна темпоральна логіка
МВ	машина виведення
МНД	міждисциплінарні наукові дослідження
МПЗ	моделі представлення знань
МКС	мовна картина світу
МОІС	мовно-онтологічна інформаційна система
МОКС	мовно-онтологічна картина світу
МПЗ	мова представлення знань
НД	наукові дослідження
НОКС	навчальні онтолого-керовані системи
ОВР	Онтологія верхнього рівня
ОГ	онтологічний граф, ондограф

ОО	об'єктно-орієнтований
ОнП	онтологічний підхід
ОнС	Онтологічна система
ООП	об'єктно-орієнтований підхід
О ПдО	Онтологія предметної області
ОКІС	Онтолого-керована інформаційна система
ОКІС ПдО	онтолого-керована інформаційна система предметної області
О-модель	онтологічна модель
ОТ	обчислювальна техніка
ОП	обчислювальний пристрій
ПдД	предметна дисципліна
ПдО	предметна область
ПЛІС	програмовна логічна інтегральна схема
ПМ	Природна мова
ПМВ	природномовне висловлювання
ПІ	природний інтелект
ПМЛ	пропозиціональна модальна логіка
ПМО	природномовний об'єкт
ПМТ	природномовний текст
ПМ-	природномовний-
ПООД	програма опрацювання онтологічних даних
ПО ПдО	початкова онтологія предметної області
ПП	Прикладний процесинг
ППЗ	підтримка програмного забезпечення
ППП	підсистема прикладного процесингу
ППФ	правильно побудована формула
ПРМ	проблемно-розв'язуваний метод
ПрП	Проблемний простір
САПР	система автоматизованого проектування
СВД	симулятор з віддаленим доступом
СІІ	семантико-інформаційний інтерпретатор
СМ	семантична мережа
СКБД	система керування базою даних
ТД	текстовий документ
ФЛП	формально-логічне представлення
ШІ	штучний інтелект
EIGRP	Протокол внутрішньої маршрутизації між шлюзами
FOL	First Order Logic
IDEF	Integrated DEFinition
KIF	Knowledge Interchange Format
KP	Knowledge Processing
KR	Knowledge Representation
LSA	Оголошення про стан каналу ( <i>link-state advertisements</i> )

MDD	Model Driven Development
NLP	Natural Language Processing
ODA	Ontology Driven Architecture
OWL	Ontology Web Language
RDFS	Resource Description Framework Schema
RIP	Протокол маршрутної інформації

## **ПЕРЕДМОВА**

У монографії викладаються основи створення методичних та інформаційно-програмних засобів побудови онтологічних моделей опису функціональних об'єктів та процесів при формуванні загальної мережі знань для вирішення актуальних проблем підвищення ефективності підтримки процесу дослідницької діяльності талановитої молоді, що включають науково-методологічні засади та сучасні інформаційні технології, які забезпечують розробку формалізованої системи знань в конкретних предметних галузях, зокрема з використанням віртуальних дослідно-проектувальних лабораторій (ВДПЛ), які надаватимуть можливість учням здійснювати дистанційно та локально дослідження за допомогою аналітичного устаткування, аналіз та обробку отриманих результатів.

За змістом у монографії відображено:

методологія системно-онтологічного аналізу предметної області, інформаційних та лінгвістичних моделей обробки текстових документів із заданої предметної області, формалізації та представлення добутих знань;

архітектура і структура інструментального комплексу підтримки процесів онтологічного опису предметних галузей.

Наведено приклади та фрагменти програм опрацювання онтологічних даних (ПООД), що входять до інструментального комплексу і підтримують процес автоматизованої побудови онтологічних структур.



## ВСТУП

У теперішній час інтелектуальні інформаційні технології (ІТ) все глибше проникають у всі сфери науково-технічної діяльності людського суспільства. Центральною їх ланкою є технології інженерії знань, включаючи процеси керування знаннями (в тому числі і в навчальному процесі), тому успіхи даного напрямку багато в чому визначаються інтелектуальним рівнем і загальною ефективністю комп'ютерних систем (КС).

Особливістю інтелектуальних інформаційних систем (ІС) є інтеграційні процеси, які впливають на розвиток їх архітектури та функціональних можливостей. Системна інтеграція міждисциплінарних знань як таких, а також технології їх ефективного формування, представлення, обробки та використання є сильними стимулюючими факторами на цьому шляху. Не менш важливим фактором є природна інтеграція двох областей ІІ, які раніше розвивалися майже паралельно і незалежно: knowledge-engineering і комп'ютерна лінгвістика. Вказану інтеграцію можна розглядати як закономірний еволюційний розвиток різних наукових теорій, закладених ще з давніх часів і відомих як логіка і онтологія.

Однією з важливих гілок сучасного розвитку ІС є онтолого-керовані інформаційні системи. Побудова останніх тісно пов'язана з розробкою теоретичних основ і методології проектування, що включають формальний підхід, фундаментальні принципи і механізми, узагальнену архітектуру і структуру системи, формальну модель і методологію проектування онтології предметної області (в тому числі і онтологій предметних дисциплін), формальну модель представлення знань, узагальнені алгоритми процедур обробки знань та ін. У свою чергу, кожна з перелічених складових загальної методології проектування являє собою складну інформаційно-алгоритмічну структуру. Наприклад, розробка онтології ПдО тісно пов'язана з концептуалізацією онтологічних категорій, розробкою та удосконаленням ієрархічних структур сутностей на всіх рівнях, побудовою формальної системи аксіом і обмежень. Комплексне вирішення зазначених завдань проектування має підвищити роль онтологічних (концептуальних) знань при вирішенні конкретних задач в прикладних областях і в навчальному процесі зокрема.

При проектуванні знання-орієнтованих інформаційних систем (і зокрема ОКІС) істотним є вибір як формально-логічного представлення знань, так і джерел придбання і поповнення знань. В даний час визнаним «де-факто», найбільш великим і загальнодоступним джерелом знань є Інтернет-простір, з його природним способом представлення інформації. Зазначений спосіб визначив появу численних інформаційних технологій обробки знань, що містяться в ПМО, в тому числі й NLP-технології.

В монографії зроблено акцент викладання процесів створення засобів та методик формування інтелектуальних інформаційно-освітніх середовищ, підтримки колективної взаємодії учні-викладачі-науковці при виконанні обдарованими учнями досліджень за програмами МАН. Створення засобів та методик забезпечення доступу обдарованих учнів до науково-інформаційних

баз провідних наукових центрів і університетів, проектування та побудову представлення предметних знань на основі обробки природномовної (текстової) інформації, так як об'єм текстової інформації, в першу чергу в мережі Інтернет, зростає стрімкими темпами, за якими не поспівають темпи удосконалення користувальницьких (архітектурних, технологічних) характеристик сучасних персональних комп'ютерів. Існуючий на сьогодні розрив між добре проробленими методами та засобами граматичної обробки природномовної інформації (ПМ-інформації) для вирішення прикладних задач у вузькоспеціалізованих предметних областях, з одного боку, і недостатністю таких стосовно вирішення комплексних задач, зв'язаних з функціональним аналізом і розумінням ПМ-інформації, її формально-логічним представленням, добуванням предметних знань з їх наступним використанням при проведенні експериментальних досліджень, затримує розвиток і застосування систем обробки знань, що реалізують (в тому числі) в повному об'ємі технологій Data & Text Mining.

# РОЗДІЛ 1. КЕРУВАННЯ ЗНАННЯМИ - НЕВІД'ЄМНА ЧАСТИНА НАВЧАЛЬНОГО ПРОЦЕСУ

## 1.1. Аналіз систем керування знаннями

Щоденний обсяг інформації, з яким зіштовхується наш сучасник, відповідає річному потоку столітньої давнини. Традиційна модель утворення переживає теперішню кризу: половину свого життя людина вивчає те, що вже відомо науці, і лише протягом короткого періоду може діяти продуктивно, тобто створювати нові знання [1].

Прагнучи надати учням, студентам, вчителям, фахівцям з освіти саму актуальну інформацію, школи, університети, інші навчальні заклади постійно модифікують навчальні програми, і все-таки підготовка талановитої молоді та випускників шкіл та вищих навчальних закладів підтверджує, що ці методики застаріли на кілька років (тому процес навчання усе більше перетворюється в активний обмін знаннями у вигляді семінарів, дискусій, конференцій) [2-7].

У нашого часу навіть особлива назва –«Інтерактивне століття» [1], для нього зараз характерні нові засоби та сфери спілкування й взаємодії людей. У новому столітті на якісно інший рівень виходять всі можливості - і інтелектуальні (одержання будь-яких даних, інформації, знань), і пов'язані із практичною діяльністю (мобільність людини, відкритість ринків і країн, воля бізнесу й конкуренції).

Накопичення суспільством величезного обсягу знань може в майбутньому привести до якісної зміни природи людини. Деякі вчені пророкують, наприклад, появу здатності до сприйняття інформації на генному рівні (що дозволить, скажемо, починати шкільне навчання не з таблиці множення, а із квантової механіки). Однак поки до такого не дійшло, нам, що живе в Інтерактивному столітті, необхідно якимось пристосуватися до сучасної дійсності, використовуючи свої звичайні можливості.

Для цього потрібно активно створювати системи керування знаннями (Knowledge Management, КМ), які здатні обробляти гігантські масиви даних, видаючи людині для ухвалення рішення лише «конденсат» інформації [2,3, 5,6,65,66,73].

У даному зв'язку цікава поява організацій нової формації, чия діяльність являє собою сплав традиційного виробництва (позначуваного як brick-n-mortar - «цегла й цемент», що вказує на його приземленість, реальність) і інтерактивного менеджменту (pure play, або «чиста гра», заснована на нематеріальних знаннях). Захват із приводу самого факту виникнення компаній «чистої гри» поступово стихає - приходить розуміння, що істина, як завжди, перебуває посередині. Симбіозну бізнес-модель позначають терміном click-n-mortar («щиглик і цемент», тобто керування реальним виробництвом за допомогою інформаційних технологій).



Рис. 1.1. Види знань: явні та неявні

На думку фахівців аналітичної компанії Gartner Group, уже через декілька років тим фірмам, які не зможуть перейти до моделі click-n-mortar, прийде стикнутися із серйозними труднощами через різку втрату конкурентоспроможності (імовірність такого розвитку подій оцінюється Gartner Group досить високо - на рівні 70%).

Підвищений інтерес до керування знаннями привів до того, що до даного процесу стали ставитися як до якогось містичного дійства, що дозволяє компанії здійснити якісний прорив і одержати значні конкурентні переваги.

Тому необхідно уважно розібратися в суті поняття «знання», а також самого процесу керування знаннями [4-8,66,68-73].

### **Знання - сила**

Знання не народжуються самі по собі, вони з'являються в результаті трансформації одних елементів інформаційного простору в інші.

Можна виділити п'ять рівнів інформаційного простору:

- **Дані** - це не піддані обробці, не осмислені відомості. Вони є сирим матеріалом для наступних перетворень, але вони існують незалежно від аналізу, або потреби в аналізі. Дані можуть бути представлені в різних формах, наприклад у вигляді електронних таблиць (MS Excel, Lotus і т.п.) тощо;

- **Інформація** - це оброблені, осмислені дані. Інформація відповідає на питання «хто?», «що?», «де?», «коли?», тобто такі дані сприймаються у

взаємозв'язку з іншими відомостями (наприклад, електронна база даних із чіткими зв'язками між вхідними в неї таблицями);

- **Знання** - використання інформації для досягнення певного результату, дає відповідь на питання «як?». Коли учень запам'ятовує якісь відомості, він здобуває пасивні знання: він одержує можливість відповісти на певні питання, однак не опановує принципами формування знань, тому що пасивні знання не дозволяють створювати нові знання на основі логічних правил. Наприклад, всі школярі зубрять таблицю множення, завдяки чому можуть сказати, що  $2 \times 2 = 4$ . Однак на питання, скільки буде  $137 \times 960$ , відповіді не піде або він виявиться невірним - рядок з добутком цих чисел не була заучена, її немає в таблиці. Для того щоб правильно відповісти на це питання, потрібні так звані аналітичні здатності, що визначають наступний рівень – розуміння;

- **Розуміння** - можливість створення нових знань на базі придбаних раніше, відповіді на питання «чому?». Знання й розуміння розрізняються так само, як завчання й вивчення. У комп'ютерних технологіях функції розуміння виконують аналітичні системи, які формують нові знання імовірнісного характеру на основі наявних даних, інформації й знань;

- **Мудрість** - оцінка розуміння. Тільки цей рівень ставиться до майбутнього (чотири попередні рівні працюють у минулому або сьогодні). Мудрість дозволяє використати інформацію всіх інших рівнів як вхідну для розуму — особливої «програми», що створює розуміння там, де його раніше не було. Відмінна риса мудрості полягає в тому, що людина, що володіє нею, постійно задає питання, на які або немає відповідей взагалі, або їх важко знайти. На нашу думку, комп'ютери ніколи не досягнуть рівня мудрості, оскільки володіння цією якістю вимагає більшого, ніж оперування функціями й алгоритмами (може бути, того, що називається душею).

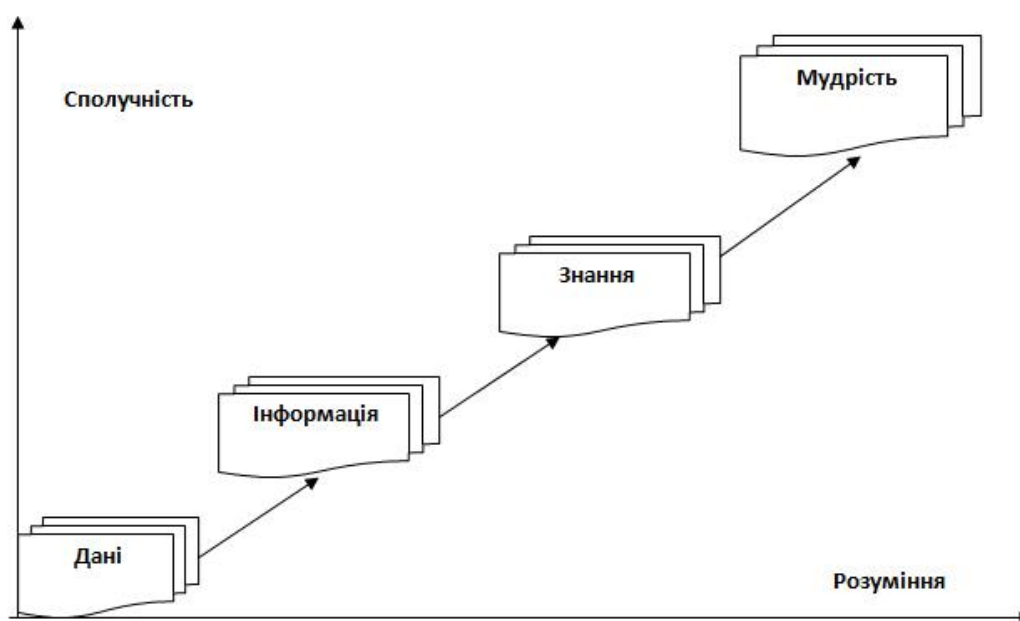


Рис. 1.2. Процес переходу з одного рівня на інший та компоненти особливого сполучного елемента - розуміння, які забезпечують кожний крок

- Дані являють собою початковий елемент. Їхнє одержання складається в констатації голого факту (наприклад, іде дощ);

- Інформація є втіленням усвідомлення зв'язків даного факту з іншими - формується судження, що вказує на причину події (температура знизилася, тому пішов дощ);

- Знання дають шаблон, що з'єднує кілька розрізнених шматків інформації й що дозволяє вивести загальні закономірності (якщо при високій вологості падає температура, то волога перестає затримуватися в атмосфері, і починається дощ);

- Мудрість — це глибоке розуміння фундаментальних принципів усього що відбувається (відбувається атмосферний цикл: іде дощ, що випали опади випаровуються, відбувається циркуляція повітря, температура змінюється, знову йде дощ).

### **Що є знання в організації?**

Для того, щоб чітко розуміти, що є знання, треба перелічити основні відмінності між інформацією та знаннями.

*Таблиця 1.1.*

### **Порівняльна характеристика інформації та знання**

<b>Інформація</b>	<b>Знання</b>
<ul style="list-style-type: none"><li>- Генерується людьми, комп'ютерами, машинами, і т.п.;</li><li>- завжди зв'язана з даними;</li><li>- перебуває скрізь;</li> <li>- може залежати від контексту, а може - ні;</li> <li>- легко сприймається й передається;</li><li>- частіше статична;</li><li>- може бути легко взаємозалежна;</li><li>- має вартість створення й підтримки;</li><li>- може використатися ким завгодно й коли завгодно</li></ul>	<ul style="list-style-type: none"><li>- генеруються тільки людьми;</li> <li>- мають відношення до даних і інформації, але не завжди з ними пов'язані;</li><li>- дефіцитні;</li><li>- завжди пов'язані з контекстом;</li> <li>- важкі;</li><li>- динамічні;</li><li>- вимагають мереж розуміння;</li><li>- дуже дорогі й ціна не фіксована;</li> <li>- мають термін використання й цільове призначення</li></ul>

Сукупність масивів інформації, знань та досвіду керівництва та фахівців, а також ефективного аналітичного інструментарію – це є знання організації.

Повинен бути організовано процес систематичного пошуку, відбору, організації, фільтрації та презентації інформації з метою поліпшити розуміння тих або інших аспектів професійної діяльності або дослідження.

Керування знаннями визначається як відбір і аналіз доступних, а також одержання необхідних знань, що дозволяють планувати заходу щодо рішення бізнесу - або навчальних проблем і контролювати виконання поставлених завдань.

Керування знаннями має наступні функції:

- допомагає фахівцям капіталізувати наявні знання й досвід і здобувати нові знання;
  - забезпечує збереження всіх накопичених знань і їхнє активне використання для організації навчання, рішення поточних проблем і складання стратегічних планів;
  - запобігає старінню об'єктів інтелектуальної власності;
  - підвищення та постійна підтримка на належному рівні професійної грамотності фахівців освітньої галузі;
  - дає користувачам підвищену гнучкість в організації навчального процесу.
- Процес керування знаннями йде по таких чотирьох основних напрямках:
- дослідження знань і їхня систематизація;
  - усвідомлення знань і визначення їхньої цінності;
  - планування й здійснення дій відповідно до результатів аналізу знань;
  - постійна капіталізація й переосмислення знань.

На перший погляд перетворити всі ці тези в життя не так вуж важко, але на справі організувати цілісну систему керування знаннями і змусити її працювати на користь компанії зовсім не просто. Для цього необхідно проробити значну попередню роботу, що складається в наступному [1-3, 7, 66]:

А) Визначити, якими коштовними знаннями володіє організація:

- Де перебувають інтелектуальні активи?
- Що в них особливо корисного?
- Що дасть їхнє використання?
- У якій формі представлені знання?
- Є чи до них колективний доступ?

Б) Проаналізувати, яку користь може принести використання знань:

- Які напрямки використання знань?
- Який очікуваний ефект від їхнього застосування?
- Які можливі перешкоди можуть зустрітися на шляху використання знань?
- Наскільки підвищиться вартість компанії в результаті використання знань?

В) Визначити, які кроки потрібно почати, щоб інтелектуальний капітал почав «працювати»:

- Як спланувати дії по застосуванню знань?
- Як запустити цей механізм?
- Як простежити за виконанням заходів?

Г) Переглянути принципи використання знань, щоб одержати максимальний ефект:

- Чи привело використання знань до бажаного результату?

- Яким образом підтримувати систему використання знань для постійного відтворення потрібного результату?

- Чи допомогло використання знань відкрити додаткові перспективи для бізнесу?

Наявні знання повинні працювати, причому важливо не тільки збільшувати обсяг і підвищувати якість цих відомостей, але й забезпечувати колективний доступ до них, формування загально корпоративного інтелектуального капіталу.

Багато організацій, в тому числі в освітньої сфері, вже накопичили істотний досвід організації, зберігання й використання знань, адже основою таких видів діяльності, як управлінське консультування (McKenzie, Boston Consulting Group), аудит і бізнес-реструктуризація (компанії «Великої п'ятірки»), розробка інформаційних технологій (Hewlett-Packard, Cisco), у першу чергу є знання. Ці бази (наприклад, в Arthur Andersen - Global Best Practices, в Hewlett-Packard - Knowledge Links) стають самостійними торговельними марками, широко відомими на ринку, і мають досить значну вартість [2, 4, 5, 66, 75].

В 90-і роки побудовою баз знань і створенням внутрішніх систем керування знаннями почали активно займатися компанії й організації інших галузей, наприклад правоохоронні органи. Сучасні комп'ютерні системи впізнавання (Face Recognition System, FRS) можуть миттєво й з високою точністю сканувати особу людини й у частки секунди зрівняти отримане зображення з базою знань центрального комп'ютера, куди стікаються дані про всіх розшукуваних правопорушників. Скануючими пристроями оснащуються зони великого скупчення людей (вокзали, аеропорти), тобто колись зручні для викриття злочинців місця стають «прозорими». У випадку збігу зображення з фотографією з бази (точність упізнавання - 90%) система миттєво передає інформацію найближчій службі правопорядку, а інше - уже справа традиційної техніки. Такі системи зараз поширюються в багатьох країнах, причому дуже швидко, тому що зараз безпеці людей приділяється дуже велике значення.

## **1.2. Принципи створення систем керування знаннями**

Для того, щоб створити ефективну систему керування знаннями, перш за все, треба їх отримати. Яким чином? [7, 15-21].

- **Усвідомлення** – це є виявлення зовнішніх ознак змін, що відбуваються;

- **Збір** потенційно значимої інформації;

- **Первинне структурування** зібраної інформації;

- **Обробка** - аналіз інформації за допомогою відповідних методів та інструментів;



- **Доступ** – це пакування та спрощення доступу до інформації;
- **Використання** інформації в процесі дослідження, прийняття та виконання рішень.

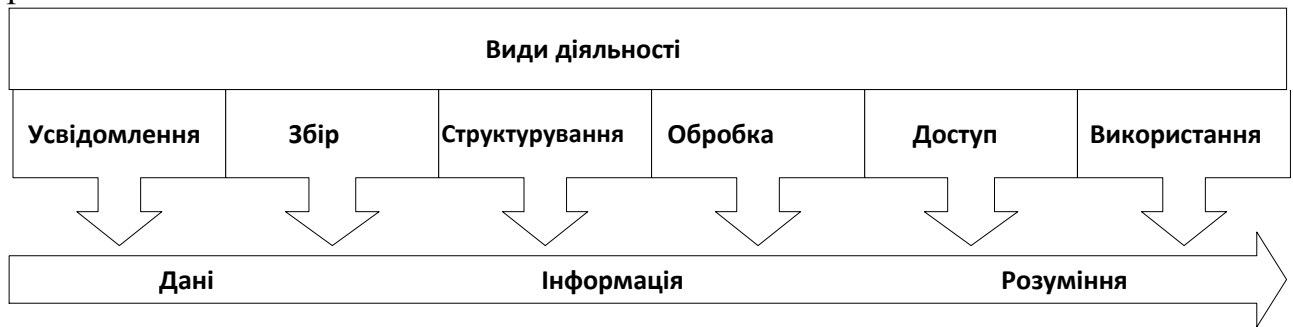


Рис. 1.3. Шляхи отримання знань

Принципи побудови баз знань і керування ними універсальні для всіх сфер діяльності. У стислому виді їх можна представити таким чином:

- **Постійний збір інформації.** До системи підключена велика кількість джерел інформації, з яких безупинно вводяться у базу нові відомості. Інформація, внесена в базу знань, ніколи (майже!) не знищується;
- **Формування знань.** Інформація міститься в базу не в довільному виді, а після ретельної обробки: виділяються спеціальні ознаки для встановлення зв'язків, що дозволяють трансформувати інформацію в знання;
- **Постійне відновлення інформації й знань.** Як відомо, у світі немає нічого вічного. База знань не може бути створена раз і назавжди, вона вимагає щоденної кропіткої праці;
- **Безперервний аналіз знань.** Даний принцип ставиться не до створення знань, а до організації роботи з ними. Вирішувати проблему потрібно на основі всіх накопичених знань.

### Складнощі в створенні систем керування знаннями

Побудова систем керування знаннями вимагає не тільки проведення аналізу, але й перегляду концепції функціонування організації - без цього сама ідея керування знаннями нежиттєздатна.

Багато баз знань, що ґрунтувалися на зазначені вище принципах і наповнені дуже коштовною інформацією, припиняли своє існування через те, що співробітники не вміли ними користуватися або робили це нерегулярно.

Головні проблеми, що виникають при створенні систем керування знаннями, можна розділити на дві основні групи — аналітичні та організаційні.

Аналітичні труднощі виникають при рішенні технічних питань, пов'язаних з необхідністю трансформувати інформацію в знання. Щоб знання мали цінність, вони повинні мати відомий ступінь абстракції й створюватися відповідно до чітких критеріїв. Наприклад, цінність винятково словесного опису накопиченого фірмою досвіду (навіть якщо представити всю історію компанії) буде невелика.

Організаційні труднощі — це налагодження роботи з базою знань на постійній основі плюс зміна щоденних рутинних процедур (переорієнтація з перманентного пошуку миттєвих рішень, у результаті якого щораз винаходиться велосипед, на дослідження бази знань, у якій, можливо, уже є відповідь). Дуже часто два працівники вирішують ту саму проблему по-різному - це значить, що як мінімум один з них робить це неефективно. Складність складається в необхідності істотних змін звичних алгоритмів функціонування організації, і саме вона найчастіше руйнує благі наміри, пов'язані зі створенням системи керування знаннями.

На жаль, багато організацій намагаються побудувати КМ-систему в «фоновому режимі», виділяючи на це ресурси й час по залишковому принципі. Крім того, виникають помилки у фундаментальних поняттях - дані плутаються з інформацією, інформація зі знанням, а багато знань попросту недооцінюються (їх відносять до категорії очевидних) [25-27, 65].

Для того, щоб проект виявився успішним, необхідно виконати наступне.

- Сформулювати мету та основний зміст проекту (включаючи коротке визначення нової бізнес-концепції), розробити план його реалізації (із вказівкою найважливіших етапів і строків їхнього виконання), визначити необхідні для цього ресурси (фінансові, матеріальні, інтелектуальні);

- Скласти інструкції із введення інформації в базу знань і провести навчання співробітників, які будуть цим займатися. Найкраще доручити таке завдання особливим працівникам – експертам - «хоронителям знань», які надалі стануть відповідати за роботу персоналу з базою знань і використання знань у виробничому процесі. На думку відомого експерта в області керування знаннями Т. Девенпорта, необхідно заснувати посаду головного директора по знаннях (Chief Knowledge Officer, СКО), що стане ключовою особою в структуруванні корпоративних знань за допомогою інформаційних технологій. Керовані ним «хранителі» повинні одержувати знання (залучити інтелектуальні можливості найбільш авторитетних співробітників компанії), групувати їх і приводити до виду, доступному для всіх службовців, а також періодично оновлювати й редагувати знання. Таких фахівців поки ніде не готують; звичайно ними стають люди, що мають досвід роботи в журналістиці;

- Сформулювати нові принципи роботи й провести навчання всього персоналу, особливо фіксуючи увагу на необхідності строго додержуватися правила первинності пошуку рішення схожого завдання в базі знань. У багатьох випадках після введення в КМ-систему основних характеристик проблеми співробітник одержить, по-перше, більше точний її опис, а по-друге, інформацію про можливі способи подолання виниклих труднощів. Якщо ж ця проблема раніше не розглядалася й система нічого не може запропонувати, то, знайшовши рішення, співробітник компанії зобов'язаний передати його «хоронителеві знань», що трансформує такі відомості в знання й уведе в базу;

- Впровадити особливі аналітичні системи, що ведуть безперервний пошук інформації і її дослідження, такі як OLAP (On-line Analytical Processing - Аналітична онлайн-обробка), MIS (Management Information System - Система управлінської інформації), BSC (Balanced Score Card - Карта збалансованої оцінки). Виходячи із заздалегідь установлених критеріїв, ці системи аналізують більші масиви інформації, що вводиться постійно, про здійснювані бізнес-процеси. Підсумкова оцінка поточної ситуації представляється керівництву фірми для прийняття рішень;

- Увести систему постійної капіталізації нових і оновлених знань, їхнього переосмислення персоналом компанії;

- Інтегрувати керування знаннями з іншими технологіями регулювання бізнесу - процесів, створивши систему, що дозволяє автоматизувати процес перетворення даних в інформацію, а інформації — у знання. Спеціалізовані модулі керування знаннями здатні працювати як окремий додаток, але найкраще його можливості розкриваються при спільному функціонуванні з іншими модулями системи.

### **1.3. Основні принципи організації дослідницької діяльності**

#### **Збір та обробка інформації**

Весь процес збору, обробки, аналізу інформації й синтезу знань являє собою ряд послідовних заходів, що повторюються від однієї мети до іншої. Цей процес одержавши назву розвідувальний цикл [4-7, 65, 66].

У його склад, як правило, входять наступні основні етапи:

- Усвідомлення - виявлення зовнішніх ознак змін, що відбуваються;
- Цілевказівка й планування;
- Збір даних і потенційно значимої інформації;
- Обробка даних (перетворення їх в інформацію);
- Структурування зібраної інформації (вибір формату й носіїв);
- Обробка - аналіз інформації за допомогою відповідних методів і інструментів;
- Доступ - упакування й спрощення доступу до інформації;
- Аналіз і синтез інформації (перетворення в знання);
- Використання інформації в процесі прийняття й виконання рішень;
- Поширення отриманих знань.

Зібрана інформація на першому етапі являє собою первинні дані, після її обробки (сортування, калібрування) вона перетворюється в інформацію, і тільки після аналізу інформації й синтезу на її основі висновків експертом, вона стає знаннями - Знаннями, що дають реальний ґрунт керівництву для прийняття рішень.

Найважливішою умовою успішної роботи експерта-аналітика - є наявність інформаційного поля досліджуваної предметної області, що повинне являти собою ряд структурованих і неструктурованих інформаційних масивів, необхідних для витягу з них необхідних даних.

Найбільш точна картина буде формуватися, якщо інформаційне поле буде містити в собі як дані, одержувані із зовнішніх джерел так і дані, одержувані із внутрішніх джерел.

Основна технологія аналітика - це встановлення причинно-наслідкових зв'язків між різного роду даними і їхнє дослідження під різними кутами зору. Побудова причинно-наслідкових ланцюжків дозволяє оброблені дані перетворити в інформацію й, зробивши висновки в предметній області, синтезувати відповідні рекомендації для керівництва підприємства [76].

Визначивши основні принципи організації роботи аналітичної групи ми можемо сформулювати вимоги до функціональності програмних засобів, які повинні забезпечити якісну роботу аналітиків.

### **Функціональність аналітичного інструментарію**

Розвиток наукових досліджень і досвідів в областях моделювання діяльності людського мозку, створення штучного інтелекту, дали розроблявачам математичні апарати для створення ПО в області нейронних технологій, інтелектуального пошуку в неструктурованій текстовій інформації (Text Mining), системи витягу даних і систем розпізнання образів (Data Mining).

Розробки в цих областях привели до створення технології керування знаннями (Knowledge Management - КМ). Це фактично підвело розроблявачів ПО до автоматизації областей людської діяльності, що важко піддаються формалізації, до якої можна віднести й процеси аналізу інформації.

Об'єктивний аналіз сучасних інформаційних технологій автоматизації бізнес-процесів організацій показує, що найбільше повно завданням розвідки відповідає КМ-технологія.

Більшість програмних засобів КМ-технології реалізують, як правило, що впливають типові функції:

- збір даних із джерел різних форматів (БД, неструктуровані джерела й т.д.);
- нагромадження й зберігання даних;
- рубрикацію архівів;
- пошук даних, у тому числі й нечіткий пошук;
- побудова звітів у різних зрізах вибірки, у тому числі й багатомірний аналіз даних;
- побудова причинно-наслідкових ланцюжків даних, що дозволяють визначати тенденції й напрямки розвитку ситуації.

Таким чином, рішення в області програмного забезпечення для автоматизації процесів аналітики й розвідки умовно можна класифікувати по ряду ознак.

По функціональності:

- повнофункціональні;
- часткові (реалізуючи одну або кілька функцій КМ-технології).

За формою подання оброблюваних даних:

- для роботи зі структурованими даними (цифрові дані);
- для роботи з неструктурованими даними (текстові, відео, графіка);
- комбіновані.

По ступені автоматизації логічних операцій:

- комплекси пошуку й збору даних (пошукові системи);
- аналітичні комплекси (утримуючі автоматизовані процедури або методи аналізу даних);
- пошуково-аналітичні комплекси.

#### **1.4. Огляд систем та пошукових платформ, які використовуються в світі для створення систем керування знаннями**

У цей час на ринку інформаційних технологій в області КМ і ВІ (Business Intelligence) представлені в основному компаніями-розроблювачами ПО закордонного походження, тому що на українському ринку такі розробки не ведуться. Це пояснюється декількома причинами:

- математичний апарат ПЗ в даній області ставитися до НИОКР в області моделювання штучного інтелекту, що вимагає залучення персоналу найвищої кваліфікації;
- низька капіталомісткість ІТ-ринку країн СНД не дозволяє виділити компаніям-розроблювачам необхідні інвестиції на НДР у даній області.
- Відповідно, про повнофункціональних КМ-системи можна говорити тільки у великих закордонних розроблювачів - це Convera, Autonomy, FAST, Hummingbird, Exalead і т.д.

Організації, здобуваючи системи класу КМ, одержують наймогутніший інформаційно-аналітичний інструмент, що істотно підсилить аналітичні підрозділи й забезпечити керівництво інформацією для прийняття рішень, що базується на всьому обсязі знань організації [2, 3, 66, 67].

Огляд технологічних рішень на ринку КМ-систем - досить складні й трудомістке завдання.

Багато закордонних КМ-систем не представлені на нашому ринку, більше того - не адаптовані для російського (а тим більше!) української мови. Це досить серйозна проблема, тому що одне з найважливіших вимог до системи керування знаннями - робота на багатьох мовах (залежно від сфер інтересів розвідки) і особливо - робота в крос-язиковому режимі (див. статтю в попередньому номері журналу) [5, 6].

Для забезпечення роботи аналітичних груп крім КМ-систем можуть також використатися інформаційно-пошукові системи.

У зв'язку із цим ми спробували представити основні речення на ринку України й СНД і основних гравців у даних групах, описати можливості, особливості й склад комплексів з урахуванням мовної підтримки для забезпечення роботи на багатьох мовах і, особливо, на українській й російській.

Аналіз виконувався по відкритих джерелах Інтернет за останніх декілька років.

## **Аналітична система Convera RetrievalWare**

Про компанію: американська компанія Convera Technologies International Ltd. ([www.convera.com](http://www.convera.com)) (раніше – Excalibur) за даними звітів Gartner Group уже багато років є світовим лідером в області розробки технологій пошуку інформації й витяги знань із текстів і мультимедійних даних будь-якої природи (текстовий, графічні, звук, фото, відеозображення у файлах практично всіх відомих форматів).

Рішення: Convera RetrievalWare (RW) – промислова аналітична платформа керування знаннями - оптимальне рішення для організації, інтеграції інформаційних масивів, пошуку й виявлення схованих неочевидних знань. Швидка й гнучка система працює поза залежністю від крапки доступу, з різними типами й сховищами даних і з різними мовами.

Застосування:

- створення єдиного інформаційного простору для забезпечення ефективної аналітичної роботи всіх груп користувачів;
- інформаційна й конкурентна розвідка;
- аналіз і створення баз за матеріалами ЗМІ;
- поширення результатів персоналізованих запитів між користувачами;
- створення спеціалізованих територіально розподілених систем;
- забезпечення інформаційної безпеки підприємства (моніторинг і маршрутизація електронної пошти, службових каталогів та ін.);
- маршрутизація й класифікація вхідної електронної пошти великих структур;
- забезпечення співробітників підприємства необмеженою інформацією з Інтернету без безпосередньої роботи в мережі (інформація із заданих вузлів Інтернету доставляється в локальну мережу підприємства й безпосередньо на робочі місця співробітників);
- ретроконверсія документів (переклад в електронний вид), їхня класифікація й ранжирування, створення й керування електронними архівами підприємств;
- забезпечення пошуку інформації на великих Інтернет - порталах, магазинах і т.д.;
- уведення, зберігання, пошук і витяг аудіо-відеоінформації;
- виконання аналіз повноти наявної інформації;
- створення інформаційних баз даних на компакт-дисках;
- багато чого іншого.

Лінгвістична підтримка забезпечується на основі:

- семантичних мереж, які представляють об'єкти реального миру як зв'язані між собою поняття з їхніми відносинами й взаємозв'язками;
- класифікаторів (таксономій і онтологій), що створюють розгалужені каталоги зрозуміти і об'єктів аналізу;

- тезаурусів – списків із системою перехресних посилань, необхідних для організації колекцій документів при їхньому відшукуванні, відображенні й зберіганні.

Розвинена система безпеки, що успадковує властивості безпеки джерел інформації дозволяє використати RW як засіб створення територіально-розподілених інформаційно-аналітичних систем практично для необмеженої кількості користувачів.

Склад комплексу:

- RW Internet Spider - спеціальний додаток для пошуку в позначених областях Інтернету й Інтранету;

- RW WebExpress - спеціальний додаток для обслуговування провайдерів, забезпечення пошуку по вмісту web-сайті й електронної торгівлі через Інтернет;

- RW CDEExpress - додаток для створення портативних баз даних на компакт-дисках, що містять пошуковий механізм RW;

- ScreeningRoom (SR) - засіб керування відео архівом, що дозволяє поряд з візуальним пошуком виділяти з відеозображень текст, що відповідає субтитрам або телетексту й перетворювати в текст супровідний аудіодорожку, існує можливість створювати анотацію до відеосюжетів storyboard;

Мовна підтримка: Український, Російська, English, French, German, Spanish, Arabic, Dutch, Italian, Japanese, Portuguese, Chinese, Finnish, Korean, Romanian, Croatian, Greek, Lithuanian, Slovak, Czech, Hungarian, Norwegian, Swedish, Danish, Icelandic, Polish, Turkish, усього більше 50 мов

Кросс-мовність: Так.

### **Рішення від Hummingbird**

Про компанію: Канадська компанія Hummingbird, Inc ([www.hummingbird.com](http://www.hummingbird.com)) Компанія Hummingbird - один з лідерів в області розробки корпоративних рішень для надання розширеного доступу до бізнесу-інформації й ресурсів підприємства.

Рішення: повнофункціональний програмний комплекс для керування інформаційними ресурсами підприємства Enterprise Information Management Systems, (EIMS).

EIMS містить у собі лінійку програмних продуктів, що реалізують ті або інші функції Км-технології. До складу комплексу входить ряд програмних модулів.

Застосування:

- розподіл і оперативна доставка корпоративної інформації;

- обмін даними й аналітичними додатками між користувачами;

- обробка й складання звітів;

- забезпечення доступу до всієї бізнес інформації й ресурсам підприємства;

- керування документами й знаннями;

- професійна обробка запитів і складання звітів;

- багато чого іншого.

Склад комплексу:

- Hummingbird KM – Керування знаннями – надає користувачам і проектним групам можливість інтегрувати інформаційні репозитории для пошуку інформації й аналізу документації.

- Hummingbird BI – Бізнес-аналітика – дозволяє користувачеві формулювати питання про дані й поєднувати результати у звітах; виконувати багатомірний (OLAP) аналіз даних і одержувати аналітичну інформацію з різних категорій.

- Hummingbird DM – Рішення по документообігу – оптимизує процеси, пов'язані з одержанням, обміном, колективним використанням і забезпеченням захисту корпоративних інформаційних ресурсів, виражених у вигляді електронних документів.

- Hummingbird RM – Керування записами – створює середовище для керування повним життєвим циклом всіх інформаційних активів корпорації.

- Hummingbird Collaboration – Середовище колективної роботи для системи керування корпоративною інформацією.

- Hummingbird Portal – Портал створений для керування документами й корпоративною інформацією, інтелектуальними активами, доступом до головного сховища й роботою в ятірному середовищі, керування взаємодією тощо.

Мовна підтримка: Російські, китайські, японські, корейський, основні європейські мови.

Кросс-мовність: Немає.

### **Рішення від NeurOK**

Про компанію: NeurOK ([www.neurok.ru](http://www.neurok.ru)) - російсько-американська група високотехнологічних компаній, що поставляє готові рішення й виконує замовлені проекти по створенню інформаційно-пошукових систем керування знаннями на основі платформи NeurOK Semantic Suite. Більше 50 чоловік працюють в офісах компанії в McLean (США), Москві й Снежинске.

Рішення: NeurOK Semantic Suite – це ефективне рішення завдань відстеження, організації й керування інформаційними потоками й масивами. Оригінальна технологія розпізнавання тематичного контексту дозволяє автоматизувати багато рутинних процесів обробки інформації.

Застосування:

- автоматична рубрикація інформації й анотування каталогів;
- автоматичне створення каталогів;
- автоматичний моніторинг відновлень і новин;
- структурування, аналіз і узагальнення даних;
- прогнозування й оптимізація процесів;
- розпізнавання образів;
- інтеграція з базами даних;
- створення інтелектуальних пошукових систем;



- пошук документів по подобі.
  - Склад комплексу:
    - Semantic Teacher – спеціалізований редактор каталогів;
    - Semantic Miner – компонента для автоматичного створення тематичного каталогу;
    - Semantic Explorer – система візуалізації змісту текстових колекцій і навігації в них. Ієрархія тематичних категорій полегшує роботі з текстовими масивами, будучи своєрідним змістом баз даних.
    - Semantic Annotator – модуль виявляє й відзначає в документі найбільш значимі терміни й фрази, у максимальному ступені що відбивають зміст документа.
    - Semantic Scanner – програмний робот із широким вибором налаштувань автоматично відслідковує відновлення в джерелах інформації, будь то інтернет-сторінки або новинні стрічки інформаційних агентств.
- Мовна підтримка: Російська, можливі адаптація системи для інших мов.  
Кросс-мовність: Немає.

### **Рішення від «Інформбюро»**

Про компанію: російська компанія «Інформбюро» ([www.informburo.net](http://www.informburo.net)) надає широкий спектр інформаційно-аналітичних і маркетингових послуг, а також рішень в області інформаційних технологій для різних суб'єктів ринку й структур.

Рішення: програмний комплекс Intellectum.BIS™ призначений для забезпечення автоматизованого керування інформаційними ресурсами підприємства. В основному, продукт орієнтований на аналітичні підрозділи підприємств і організацій різних галузей і видів діяльності.

Застосування:

- збір даних із джерел різних форматів (БД, неструктурованих джерел, web-ресурсів і т.д.);
- нагромадження й зберігання даних, їхня каталогізація;
- пошук даних, у тому числі й нечіткий пошук;
- рубрикація архівів;
- побудова звітів у різних зрізах вибірки;
- побудова причинно-наслідкових ланцюжків даних, що дозволяють визначати тенденції й напрямки розвитку ситуації.

Склад комплексу:

- сервер комплексу – забезпечує обробку процедур у рамках групи користувачів (за замовчуванням розрахований на 5 користувачів);
- клієнтське місце – інтерфейс роботи користувача з комплексом, реалізований на технології «web-клієнт»;
- модуль адміністрування – реалізує функції адміністрування комплексом по розмежуванню прав і можливостей клієнтських місць, прийому й обробці даних, що надходять, доставці інформації до вилучених клієнтських місць;

- модуль прийому інформації – реалізує функції прийому інформації від різних джерел, розпізнавання формату документа й джерела надходження інформації, розбиття інформації на статті (якщо це текстові дані), уведення інформації в сховище даних;

- поштовий робот – реалізує функцію взаємодії комплексу з поштовою системою по прийому інформації від різних джерел інформації;

- інтернет-робот WEBHunter™ – реалізує функцію збору інформації із заданих web-ресурсів із заданою періодичністю.

Мовна підтримка: Російська, інших даних немає.

Кросс-мовність: Немає.

### **Рішення від «Галактики»**

Про компанію: Російська корпорація «Галактика» ([www.galaktika.ru](http://www.galaktika.ru)) - один із провідних російських розроблювачів комплексних рішень в області автоматизації керування виробничо-господарською й фінансовою діяльністю підприємства.

Рішення: Galaktika-Zoom – автоматизована система пошуку й аналітичної обробки інформації, що поєднує функції корпоративного сховища інформації, видобутку даних і аналітичної обробки більших масивів інформації.

Застосування:

- пошук і формування інформаційних масивів по конкретних аспектах досліджуваної проблематики, здійснюваний по ключових словах з урахуванням їх морфології;

- аналіз об’єктивних значеннєвих зв’язків відібраних даних,

- одержання «образа» проблеми - моментальної багатомірної фотографії в інформаційному потоці у формі ранжируваного списку значимих слів, уживаних разом з темою проблеми;

- порівняння декількох станів проблеми, виявлення закономірностей і тенденцій (або частковостей і випадків) динаміки розвитку досліджуваної проблеми.

Мовна підтримка: Росіянин, інших даних немає.

Кросс-мовність: Немає.

### **Рішення від «Гарант-Парк-Інтернет»**

Про компанію: Підрозділ російської компанії «Гарант-Парк-Інтернет» RCO Research Group ([www.rco.ru](http://www.rco.ru)) займається апробацією й впровадженням нових підходів до побудови компонентів інформаційно-пошукових систем.

Рішення: Торговельна марка RCO™ поєднує продукти й рішення, які призначені для впровадження в бази даних і інформаційно-пошукові системи й дозволяють задіяти широкий арсенал лінгвістичних і аналітичних засобів для рішення прикладних завдань, що вимагають комп’ютерної обробки документів природною мовою.

Застосування: створення інформаційно-пошукових і аналітичних систем, що працюють з електронними документами російською мовою.

Склад комплексу:

- RCO for Oracle- продукт, що розширює можливості OracleText при роботі с базами даних, що містять документи російською мовою;

- RCO for BackOffice – продукт, що розширює можливості Microsoft BackOffice (MS SharePoint Portal, MS Indexing Service, MS Exchange Server і MS SQL Server) при роботі з документами російською мовою, забезпечуючи пошук з обліком всіх граматичних форм слів на основі морфологічного аналізу;

- RC WEB – пошукова система, що володіє можливостями як контекстного, так і реляційного пошуку. Russian Context дозволяє шукати документи з урахуванням морфології російської й англійської мов, використовуючи SQL-подібну мову запитів і комбінуючи пошукові обмеження на контекст із обмеженнями на задані атрибути документів. Продукт працює в середовищі Windows;

- RCO Morphology - продукт дозволяє включити російську морфологію в системи інформаційного пошуку. Поставляється у вигляді динамічної бібліотеки (dll) для Windows;

- RCO Thesaurus Search - продукт дозволяє включити тезаурус російської мови в інформаційно-пошукові системи для підвищення повноти пошуку;

- RCO Semantic Network - продукт призначений для розроблювачів інформаційно-пошукових і аналітичних систем і дозволяє виявити ключові поняття документа, у тому числі найменування персон і організацій, з асоціативними зв'язками між ними на основі граматичного й статистичного аналізу тексту, а також одержати кілька видів рефератів документа;

- RCO Pattern Extractor - продукт призначений для аналізу тексту й розпізнавання в ньому різних об'єктів відповідно до зразків, заданими формальною мовою;

- RCO TopTree -продукт призначений для автоматичної класифікації й побудови ієрархічних рубрикаторів по заданій безлічі довільних об'єктів, характерних них набору атрибутів.

- RCO КАОТ - інформаційно-аналітична система для роботи в локальній мережі на базі MS Windows і MS Internet Information Server, що реалізує комплекс функцій інтелектуального аналізу й пошуку текстової інформації з підтримкою Web-інтерфейсу користувача;

- RCO Fact Extractor – це персональний додаток для Windows, що призначено для аналітичної обробки тексту російською мовою й виявлення фактів різного типу, пов'язаних із заданими об'єктами – персонами й організаціями.

Мовна підтримка: Російська, англійська.

Кросс-мовність: Немає.

### **Рішення від «Интегрум-Техно»**

Про компанію: Російське інформаційне агентство "Интегрум" ([www.integrum.ru](http://www.integrum.ru)) пропонує широкий спектр онлайн послуг по інформаційному забезпеченню бізнесу: доступ до електронних архівів, професійні інструменти моніторингу й аналізу інформації, готові інформаційні продукти.

Рішення: в основі лежить технологія з використанням інформаційно-пошукової системи «Артефакт», призначеної для нагромадження й зберігання більших обсягів текстової й графічної інформації з метою виконання ефективного пошуку по всій колекції накопичених баз даних.

Застосування:

- нагромадження й зберігання більших обсягів текстової й графічної інформації з метою виконання ефективного пошуку по всій колекції накопичених баз даних.

Лінгвістичний апарат заснувань на морфологічному словнику А. А. Зализняка.

Розмір окремої бази даних обмежується лише операційним середовищем користувача й апаратним забезпеченням. Реалізовано використання багатобазового пошуку (необмеженого числа баз даних одночасно).

При збільшенні числа баз даних існує можливість їхнього об'єднання в групу. Для кожної бази даних можна створити короткий опис, що полегшує роботі з нею кінцевого користувача.

Склад комплексу:

- інтерфейс веб-додаток;
- модуль універсального інтерфейсу пошуку;
- диспетчер;
- планувальник;
- пошуковий модуль;
- модуль роботи з базами даних і індексом.

Мовна підтримка: Російська, англійська, інші європейські мови.

Кросс-мовність: Немає.

### **Рішення від Exalead [67]**

Про компанію: Французька компанія, створена в 2000 році.

Засновники - Francois Bourdoncle и Patrice Bertin - новатори в технології пошукових систем, до 1996 року – співпрацівники дослідницького центру Digital Equipment.

1996-1997: Участь в проекті AltaVista,

1998-1999: Створення дослідницької лабораторії при ENSMP (Париж),

Вересень 2000 року: засновано компанію Exalead,

2001: об'єднання з французькою промисловою групою Qualis.

Web-сайт : [www.exalead.com](http://www.exalead.com)

Застосування:

- створення єдиного інформаційного простору для забезпечення ефективної аналітичної роботи всіх груп користувачів;
- інформаційна й конкурентна розвідка;
- аналіз і створення баз за матеріалами ЗМІ;
- поширення результатів персоналізованих запитів між користувачами;
- створення спеціалізованих територіально розподілених систем;
- забезпечення інформаційної безпеки підприємства (моніторинг і маршрутизація електронної пошти, службових каталогів та ін.);
- маршрутизація й класифікація вхідної електронної пошти великих структур;
- забезпечення співробітників підприємства необмеженою інформацією з Інтернету без безпосередньої роботи в мережі (інформація із заданих вузлів Інтернету доставляється в локальну мережу підприємства й безпосередньо на робочі місця співробітників);
- ретроконверсія документів (переклад в електронний вид), їхня класифікація й ранжирування, створення й керування електронними архівами підприємств;
- забезпечення пошуку інформації на великих Інтернет - порталах, магазинах і т.д.;
- уведення, зберігання, пошук і витяг аудіо-відеоінформації;
- виконання аналіз повноти наявної інформації;
- багато чого іншого.

Стислий опис аналітичної платформи:

Корпоративна високопродуктивна інформаційно-пошукова платформа (система), через яку клієнти без додаткових налаштувань і використання інших систем отримують доступ до інформації, що може бути розташована на робочих станціях, серверах корпоративної мережі, базах даних або в мережі Інтернет.

- Універсальність - система є універсальною та легко адаптується до предметних галузей клієнтів, які є описом прикметних областей або видів діяльності людини, або бути описом приватної точці зору дослідника;

- Актуальність словників - система створення таким чином, що постійно забезпечується постійна актуальність словників системи. Різні словники, такі як морфологічні, фразеологічні, тезауруси ведуться системою автоматично та поповнюються в режимі реального часу. Для різних мов або незнайомих слів системою використовується алгоритм статистичної морфологічної лематизації;

- Гнучкість в створенні лінгвістичних ресурсів – в системі існує можливість вручну створювати і налаштовувати словники системи (онтології, тезауруси, синоніми, стоп-слова тощо). Ці лінгвістичні ресурси гнучко підключаються до системи (з урахуванням потреб аналітичної або дослідницької роботи (до кожної тематичної бібліотеки може бути підключений свій набір лінгвістичних ресурсів, що забезпечує звуження зони пошуку ));

- Мови роботи - система забезпечує роботу з документами (текстами, html-сторінками, базами даних – всього більш 350 форматів) на усіх європейських мовах, арабській, китайській, в тому числі на українській та російській мовах;

- Семантичний аналіз тексту - первинна лінгвістична обробка тексту виконується за допомогою поверхневого семантичного аналізу з виділенням базового переліку семантичних відношень. Перелік семантичних відношень має ієрархічну структуру з можливістю локального уточнення та доповнення. Семантичні відносини типу людина, географія, організація система має в базовій конфігурації, крім того в системі передбачено динамічний тезаурус, якій в режимі роботи системи постійно оновлюється. Семантичні відносини – категорії додатково створюються и конфігуруються в залежності від задач, які мають бути вирішені системою під конкретну дослідницьку задачу. Особливості семантичного аналізу, якій виконує система, та його результатів:

- Автоматичне виділення однослівних термінів предметної галузі та синтез багатослівних термінів без обмежень на кількість зв'язаних слів у терміні. Багатослівні терміни виводяться в лематизованій формі, система використовує технологію, яка використовує як морфологічну лематизацію так і статистичні лінгвістики;

- Автоматична побудова ієрархії термінів заданої глибини. Відображення відповідних речень тексту при виборі терміну чи групи термінів;

- Автоматичне створення списків пов'язаних слів – груп термінів, які найбільш характерні для документа чи групи документів та пов'язані між собою за змістом тексту;

- Створення тематичних конспектів (перелік речень, в яких присутні терміни семантичної мережі) документів як за темами, визначеними користувачем, так і за автоматично виділеними темами з можливістю автоматичного розширення заданої теми за рахунок зв'язаних тем, які можуть автоматично визначатись при аналізі документа чи задаватись користувачем заздалегідь у вигляді фрагменту онтології або тезауруса предметної області;

- Відбір визначень понять з текстових документів, формування узагальненого визначення для кожного поняття, представлення узагальненого визначення у вигляді логікової формули з використанням кон'юнкції, диз'юнкції та заперечення;

- Візуалізація автоматично побудованої мережі понять у веб-браузерах, за рахунок чого досягається кросплатформеність представлення результатів;

- Редагування мережі понять: додавання, вилучення, перейменування об'єктів та зв'язків між ними, зв'язування об'єктів з зовнішніми ресурсами, визначення ступеня зв'язку між об'єктами, перегляд фрагментів мережі заданої глибини та шляхів між окремими об'єктами;

- Використання мережі понять для створення тематичного конспекту.

В системі передбачено широке коло засобів розробки – гнучкий інструментарій для подальшого розширення функціоналу системи та подальшого її розвитку та інтеграції з іншими інформаційними системами. Система повинна має можливість підключення через інтерфейси засобів розробки до інших інформаційних систем з метою розширення їх можливостей, підтримує технологію XML для обміну даними з іншими системами.

Система дуже проста в інсталяції та налаштуванні - підключення та налаштування системи типу Plug&Play. Система має конвертори більшості поширених форматів файлів (більше ніж 350 форматів), а також модулі взаємодії з більшістю систем зберігання і управління даними, які мають суттєво спростити і прискорити інтеграцію (файлові системи, бази даних, мережа Інтернет, системи документообігу, поштові системи тощо).

Система працює з Java™ і XML, пошукова платформа підтримує більшість промислових стандартів, у тому числі технологію Java і XML (Extensible Markup Language).

Висока продуктивність системи – застосовуються інноваційні передові алгоритми, які оптимізовані для роботи на процесорах з 64-бітовою архітектурою і забезпечують швидкість обробки та пошуку документів.

Система забезпечує обробку даних в режимі реального часу - індексування та обробка структурованих і неструктурованих даних одночасно з обробкою пошукових запитів для забезпечення адекватнішою й актуальнішою інформацією. Пошукова платформа орієнтована на обробку даних, розташованих в найбільш популярних і розвинутих сховищах, таких як Web-сервери, СУБД, сервери електронної пошти, системи документообігу, файлові системи і т.д.

Інноваційні статистичні лінгвістичні алгоритми. Знаходження найбільш релевантної інформації забезпечується за рахунок застосування інноваційних статистичних лінгвістичних алгоритмів, які дозволяють виключити застосування словників і втручання людини в процес пошуку. Система є універсальною і легко адаптується до предметних областей дослідження. Крім цього в системі наявні функції обробки текстів природною мовою, морфологічної обробки, динамічної категоризації, перевірки орфографії, підбору близьких по написанню термінів, пошук з технологією «нечіткий пошук», пошук слів близьких по фонетиці, по інтервалах дати і т.д.

Система гнучка з точці зору реалізації рішення - як для корпоративної системи (робота в корпоративній мережі з усіма джерелами інформації), так є рішення для веб-порталу.

Повноцінна інтеграція і налаштування як через стандартні програмні інтерфейси (Java, XML), так і за допомогою засобів розробки та налаштування.

Постійна актуальність словників системи забезпечується самою системою. Таким чином словники, такі як морфологічні, фразеологічні та

тезауруси ведуться системою автоматично і поповнюються в режимі реального часу. Архітектура системи та простий інтерфейс дозволяють вручну створювати і налаштувати словники системи (онтології, тезауруси, синоніми, стоп-слова тощо). Ці лінгвістичні ресурси гнучко підключаються до системи (з урахуванням потреб аналітичної або дослідницької роботи (до кожної тематичної бібліотеки може бути підключений свій набір лінгвістичних ресурсів, що забезпечує звуження зони пошуку));

Гнучка і масштабована архітектура. Пошукові індекси знаходяться на жорстких дисках сервера, і при роботі з ними система використовує обчислювальні ресурси того ж сервера. При дуже великих об'ємах інформації (великий індекс) система пропонує використовувати багато серверну архітектуру. Така архітектура повноцінно масштабує платформу в трьох вимірах: обсяг дисків, об'єм оперативної пам'яті, потужність і кількість процесорів. Продуктивність системи не буде знижуватися, навіть якщо обсяг оперативної пам'яті становить 5% від обсягу індексів. Крім цього може бути передбачена можливість розподілу індексу на серверній фермі, що складається або з дрібних (PC або Linux комп'ютери), або з великих (промислові сервера або мейнфрейми) елементів.

Роздільне індексування полів - оновлення окремих полів документа в індексі не вимагає його повної переіндексації.

Підтримка транзакцій - процес індексування підтримує прості транзакції (атомарні послідовності операцій).

Системою забезпечується безперервний режим роботи і відновлення після аварій. У процесах індексування застосовані технології, такі як протоколювання, режим роботи 7x24, відмовостійкий режим обробки транзакцій. У користувачів ніколи не буде виникнути необхідність у повному перестроюванні індексів.

Підтримка XML. Для управління процесом індексації використовується мова обробки документів на основі XML. Поля XML можуть використовуватися разом з текстовими полями (embedded XML).

Інтелектуальні технології в пошуку - в системі застосовуються сучасні інноваційні методи обробки текстової інформації, незалежно від застосовуваного словника і мови, наприклад: статистичні лінгвістичні алгоритми. Вони застосовуються на всіх етапах обробки даних, це реально забезпечує оптимізацію результатів пошуку і, відповідно, задовольняє пошукові потреби користувача з точці зору швидкості та якості отримання результатів пошуку.

Зручна та швидка настройка і конфігурація - система дуже просто конфігурується через зручний графічний інтерфейс з інтуїтивно-зрозумілою системою меню і майстрів. Адміністратор системи має просто визначити оброблювані джерела даних, що використовуються для потреб користувача навігації структури категорій, налаштувати систему ранжирування і представлення результатів пошуку. Крім цього в систему дуже просто підключаються та конфігуруються всі необхідні лінгвістичні ресурси, які



мають бути використані при роботі системи. Все це не вимагає додаткового програмування.

Автоматичне визначення мови документа - застосування статистичної лінгвістики робить систему принципово багатомовною. Визначення мови документа проводиться на етапі його завантаження в систему, отримана інформація надалі використовується для того, щоб користувачі мали можливість обрати мову, якою мають бути знайдені документи. Один з варіантів - алгоритм статистичної морфологічної лематизації. Статистичний алгоритм лематизації дозволяє системі обробляти тексти на природних мовах без застосування словників і правил, які вимагають постійних доробок. В результатах пошуку система надає перелік мов з вказаною кількістю документів, які знайдені на той чи іншій мові.

Кількість мов системи. Система забезпечує роботу на всіх найбільш використовуваних мовах світу, на всіх європейських мовах, у тому числі російській та українській (всього може бути більше 50 мов). При необхідності існує можливість швидко розробити морфологічний модуль, якщо такий не включений у систему, або є задача забезпечити більш ефективний морфологічний аналіз документів.

Система забезпечує крос-мовний пошук інформації. Це реалізується за рахунок створення відповідних крос-мовних словників синонімів або категорій. Ця можливість ніяк не залежить від кількості мов, встановлених у системі.

Створення словників, тезаурусів, категорій та інших лінгвістичних ресурсів є найважливішим з точки зору як пошуку, так і навчальної та дослідницької діяльності. Система дозволяє це створити дуже швидко і просто.

Простий та зручний вибір варіантів пошуку (формування пошукового запиту та отримання результатів):

- Логічний пошук - користувачі мають можливість застосовувати стандартні логічні оператори кон'юнкції (AND), диз'юнкції (OR) і заперечення (NOT);

- Пошук словосполучень - користувачам доступний пошук точних послідовностей логічних виразів;

- Пошук з урахуванням близькості розташування - користувачі мають можливість використання оператора NEAR - для визначення максимальної відстані в тексті між зазначеними в запиті виразами;

- Додаткові пошукові оператори - користувачі мають можливість визначення додаткових пошукових операторів, що описують особливості ранжирування результатів пошуку;

- «Згортка» по сайту-джерела - система дозволяє «згорнути» результатів пошуку по цьому джерелу, при цьому з кожного сайту відбирається по одному найбільш релевантному документу;

- «Згортка» за подобою – система передбачає виявлення та видалення з результатів пошуку подібних (тих, що дублюються) документів;

- Налаштування ранжирування - вагові коефіцієнти пошукових термінів повинні налаштовуватися за допомогою потужної спеціалізованої мови ранжирування;

- Ранжирування по «купчастості» - цей метод ранжирування забезпечує ранжирування таким чином, щоб більший ранг отримували ті документи, в яких пошукові терміни зустрічаються на меншій відстані один від одного. Бажане налаштування максимальної відстані між сусідніми термінами;

- Пошук зв'язаних термінів з термінами по яких виконувався пошук;

- Пошук по полях документа - користувачам доступний пошук по полях (атрибутів) документів;

- Для спрощення пошуку у разі коли користувачеві необхідно кожен раз створювати один і той же пошуковий запит в системі передбачено визначити ці терміни, а до них підключати необхідний набір категорій;

- Пошук чисел і дат - пошук числових даних використовується для полів документів, що містять кількісні характеристики та дати. При такому пошуку допустимі оператори порівняння, такі як EQ (=), NE (<>), LE (<=), LT (<), GE (>=), GT(>);

- Пошук по інтервалам та проміжкам дати також передбачений системою;

- Сортування - список знайдених документів кожен раз є відсортований (при необхідності) по будь-якому полю, в тому числі містить числа, дату і час;

- Пошук за подібним написанням – користувачі мають можливість визначити, які з термінів, схожі на зазначені в пошуковому запиті, повинні бути знайдені. Мінімальна ступінь подібності повинна бути визначена для кожного терміна із запиту;

- Пошук за ідентичною вимовою - системою забезпечується пошук документів з термінами, які фонетично ідентичні зазначеним у запиті. Допустиме розходження між шуканими термінами і термінами в запиті може бути визначено додатково.

- Пошук подібній вимові – є можливість пошуку подібних за вимовою термінів, з можливістю настройки ступеня подібності;

- Морфологічна лематизація – в системі передбачена можливість пошуку за морфологічними основами (лема) термінів із запиту, тобто в системі повинна бути функція морфологічної лематизації;

- Пошук з макропідстановки – користувачі мають можливість використовувати в пошуковому запиті спеціальні символи макропідстановок;

- Нечіткі макропідстановки («нечіткий пошук») – в системі передбачена можливість застосовувати при пошуку нечіткі макропідстановки, які повинні бути визначені приблизно з налаштуванням мінімально допустимого ступеня подібності;

- Усікання пошукових термінів – в системі передбачено можливе спрощене застосування макропідстановок у формі пошуку без урахування префіксів і закінчень термінів запиту;

- Автоматична ідентифікація пов'язаних слів - при пошуку користувачі мають можливість ввести пошуковий запит природною мовою, при цьому система автоматично виділяє в ньому групи іменників, які і будуть далі шукатися. Можливі помилки виділення груп іменників усуваються користувачем при налаштуванні запиту;

- Відображення резюме пошуку і виділення релевантних термінів - при відображенні списку знайдених документів система автоматично формувати резюме («саммері») знайдених документів, в якому виділяються терміни з пошукового запиту і терміни, пов'язані з ними. Запит може бути уточнений шляхом додавання або виключення термінів з резюме, в тому числі і їх розширень макропідстановки, морфологічними формами або подібності з написання та вимови;

- Попередній перегляд документа - при відображенні результатів пошуку в окремому сегменті інтерфейсу користувачі мають змогу відобразити знайдені документи, в тому числі з виділеними релевантними термінами;

- Уточнення запиту по пов'язаних словах – користувачі мають можливість уточнити запит зміною термінів, що входять до нього, або знайдених системою пов'язаних слів, які відображаються в резюме результатів пошуку;

- Уточнення запиту за категоріями - у пошуковому інтерфейсі і в резюме міститися і виявлятися у знайдених документах категорії (сутності та тематичні категорії), які далі використовуватися для уточнення запиту. Для кожної категорії відображається та кількість документів, в яких вона згадується або (при великій кількості - % частка від загальної кількості документів);

- Різноманітні «зрізи» відображення резюме - система може автоматично формувати резюме результатів пошуку по декількох структурованих «зрізах», наприклад таким як дата документа, автор, джерело, мова і т.д. Користувач повинен має можливість визначати додаткові «зрізи» резюме;

- З результатами пошуку системою проводиться динамічна категоризація на основі статистично формованих пов'язаних слів, в які входять найбільш характерні групи іменників зі знайдених документів;

- Одною з особливостей системи є інтуїтивно-зрозумілий та простий в використанні інтерфейс користувача і проста (природна) форма складання запитів і навігації в пошуковому інтерфейсі.

Перевірка правопису в пошуковому запиті. У системі є вбудована функція перевірки правопису, що базується на контенті оброблених документів. Функція працює на різних мовах і використати всі словники системи, у тому числі інформацію про розпізнані власні імена і спеціальні терміни, які згадуються в оброблюваних документах.

Автоматичне створення списків пов'язаних слів - пов'язані слова представляють собою групи термінів, які найбільш характерні для знайдених документів і автоматично виявлені статистичними лінгвістичними алгоритмами. Це необхідно для деталізації або уточнення пошукового запиту. Складання списку пов'язаних слів проводиться автоматично - шляхом аналізу всіх оброблених документів, а не тільки знайдених в результаті запиту. Крім того система передбачає додаткове створення пошуку в переліку відібраних документів.

Системою підтримуються Інтернет протоколи HTTP і HTTPS, призначені для поетапного сканування Web-серверів, у тому числі з необхідною авторизацією в режимах basic, digest і NTLM. Також підтримується авторизація через cookie або через пароль входу на сайт.

Системою підтримується мережевий протокол WEBDAV, який використовується для індексування серверів WEBDAV, у тому числі Microsoft IIS з авторизацією в режимах basic, digest і NTLM.

Системою підтримується мова XML, необхідна для індексування джерел відповідних специфікації XML. Якщо для джерела визначено тип XML-документів (Document Type Definition), управління ними може здійснюватися безпосередньо платформою (витяг метаданих, категоризація, узгодження структури і т.д.).

Системою підтримуються файлові системи UNIX та Windows. Система має можливість індексування файлових систем Windows і UNIX з урахуванням ієрархії директорій і даних ACL (Access Control List). Зміна документа у файловій системі приводиться до відповідних змін в пошуковому індексі. Для файлових систем Windows індексування здійснюється у режимі реального часу.

Системою підтримуються мережеві директорії LDAP і Microsoft Active Directory для індексування файлів в мережевих директоріях (SMBFS, CIFS, NFS), у тому числі за участю системи розмежування доступу на базі Microsoft Active Directory чи сумісної з LDAP (v2 або v3).

Системою підтримується робота з поштовими системами, такими як:

- IMAP - для індексування повідомлень електронної пошти і прикріплених до них файлів, що зберігаються на IMAP-серверах;
- Lotus Notes - для індексування повідомлень електронної пошти і прикріплених до них файлів у папках Lotus Notes;
- Microsoft Exchange - для індексування повідомлень електронної пошти і прикріплених до них файлів з папок Microsoft Exchange;
- NTTP - індексування в режимі реального часу новинних груп Usenet.
- Системою підтримується робота з базами даних, такими як:
  - Documentum - для індексування документів і метаданих з баз даних Documentum;
  - Lotus Notes - для прямого підключення до бази даних Lotus Notes;
  - ODBC - для перетворення результатів певного адміністратором SQL-запиту до бази даних в набір часових XML-документів, які в подальшому

повинні індексуватися системою разом з метаданими. Крім того підтримуються СУБД DB2, Informix, SQL Server, Sybase, Oracle, Text (CSV), MySQL і більше.

В системі передбачена можливість розробки додаткових конекторів до джерел даних, якщо такий не передбачений системою (наявність засобів розробки).

В системі передбачена можливість підтримки форматів (всього більше 350) HTML, WML, XML, Text, MIME, Microsoft Office, Adobe PDF, JPEG, PNG, MP3, GIF, OGG, ZIP, RAR, WordPerfect, RTF, MacroMedia Flash і інші можливі. При необхідності є можливість додавати або видаляти необхідні формати для забезпечення роботи системи.

Робота з кодуванням – в системі підтримуються кодування: будь-які, в тому числі Юнікод (UTF), всі кодові сторінки Windows, національні кодування, в тому числі російська, арабська, китайська, корейська і японська.

Крос-мовність: так.

Аналіз спектра програмних засобів, представлених на ринку інформаційних технологій, демонструє нам досить широкий вибір можливостей використання засобів для автоматизації аналітичної діяльності. Основними факторами, що визначають вибір того або іншого програмного продукту, є:

- мети й завдання аналітичного підрозділу;
- обсяги й склад оброблюваної інформації;
- розміри підприємства;
- вартість програмних рішень, що задовольняють перерахованим вище вимогам і наявний бюджет.

Як ми вже відзначали вище, найбільшими можливостями по зборі інформації й витягу з її знань будуть володіти повнофункціональні програмні комплекси, тому що смороду забезпечать аналітику цілісну інформаційну картину, поповнювану як зовнішніми джерелами, так і внутрішньою інформацією підприємства.

При виборі комплексу необхідно взяти до уваги можливість подальшого нарощування його потужності й функціональності, якість його технічного супроводу й перспективи в майбутньому (стійкість компанії-розроблювача). Правильний облік зазначених факторів дасть гарантії підприємству, що серйозні інвестиції на автоматизацію аналітичної діяльності не будуть викинуті на вітер, а виллються в конкурентну переваги компанії, забезпечать ріст бізнесу, знизять ризики господарювання.

## **1.5. Місце систем керування знаннями в освітніх проектах**

Знання не народжуються самі по собі, вони з'являються в результаті трансформації одних елементів інформаційного простору в інші.

На даному етапі розвитку інформаційного суспільства можна виділити п'ять рівнів інформаційного простору [3, 7, 8, 33-42, 75]:

- Дані - це не осмислені відомості, які к тому ж не були структуровані. Вони є сирим матеріалом для наступних перетворень, але вони існують незалежно від аналізу, або потреби в аналізі. Дані можуть бути представлені в різних формах та форматах (наприклад .doc., .txt. ppt, MS Excel, Lotus і т.п.) тощо;

- Інформація – це оброблені та осмислені дані. Інформація відповідає на питання «хто?», «що?», «де?», «коли?» - тобто такі дані сприймаються у взаємозв'язку з іншими відомостями (наприклад, електронна база даних із чіткими зв'язками між вхідними в неї таблицями);

- Знання - використання інформації для досягнення певного результату, вони дають відповідь на питання «як?», «що робити?». Коли учень запам'ятовує якісь відомості, він здобуває пасивні знання: він одержує можливість відповісти на певні питання, однак не опановує принципами формування знань, тому що пасивні знання не дозволяють створювати нові знання на основі логічних правил. Однак на складні питання швидкої відповіді не буде або відповідь виявиться невірною – це не може бути завченими, цих даних немає в таблицях. Для того щоб правильно відповісти на складні питання, потрібні так звані аналітичні здатності, що визначають наступний рівень – розуміння;

- Розуміння – це можливість створення нових знань на базі придбаних раніше, відповіді на питання «чому?». Знання й розуміння розрізняються так само, як завчання й вивчення. У комп'ютерних технологіях функції розуміння виконують аналітичні системи, які формують нові знання імовірнісного характеру на основі наявних даних, інформації й знань;

- Мудрість є оцінкою розуміння. Тільки цей рівень ставиться до майбутнього (чотири попередні рівні працюють у минулому або сьогодні). Мудрість дозволяє використати інформацію всіх інших рівнів як вхідну для розуму — особливої «програми», що створює розуміння там, де його раніше не було. Відмінна риса мудрості полягає в тому, що людина, що володіє нею, постійно задає питання, на які або немає відповідей взагалі, або їх важко знайти.

На думку фахівців комп'ютери ніколи не досягнуть рівня мудрості, оскільки володіння цією якістю вимагає більшого, ніж оперування функціями й алгоритмами (може бути, того, що називається душею).

Все це є різні ступені володіння інформацією, яка є доступною і може знаходитися в різних якостях та у різних джерелах.

Якщо говорити про знання, то треба визначити що є знання організації. Знання будь якої організації складаються з трьох загальних компонентів:

- інформаційний простір, якій складається з комерційної, навчальної, законодавчої, поштової, дослідницької, аналітичної та іншої інформації;

- знання фахівців всіх рівнів – експертів, методистів, тренерів, вчителів та інших користувачів;

- інструментарій, якій дозволяє оперувати інформацією, зберігати та розповсюджувати знання фахівців.

Найважливішим в сучасні дні є розуміння та організація процесів видобування та керування знаннями.

Головною метою створення та впровадження систем керування знаннями є створення технологічних рішень підтримки доступу та розповсюдження сучасних знань серед молоді для забезпечення інноваційного розвитку навчального процесу сучасної освіти, а також впровадження сучасних інформаційно-комунікаційних технологій з використанням переваг електронних освітніх ресурсів.

Щоденний обсяг інформації, з яким зіштовхується сьогодні мабуть кожна людина, відповідає річному інформаційному потоку столітньої давнини.

Традиційна модель утворення переживає теперішню кризу: половину свого життя людина вивчає те, що вже відомо науці, і лише протягом короткого періоду може діяти продуктивно для того щоб створювати нові знання.

Прагнучи надати учням, студентам, вчителям, фахівцям з освіти саму актуальну інформацію, школи, університети, інші навчальні заклади постійно модифікують навчальні програми, і все-таки підготовка талановитої молоді та випускників шкіл та вищих навчальних закладів підтверджує, що ці методики застаріли на кілька років (тому процес навчання усе більше перетворюється в активний обмін знаннями у вигляді семінарів, дискусій, конференцій тощо).

У нашого часу навіть особлива назва - інтерактивне століття, для нього зараз характерні нові засоби та сфери спілкування й взаємодії людей. У новому столітті на якісно інший рівень виходять всі можливості - і інтелектуальні (одержання будь-яких даних, інформації, знань), і пов'язані із практичною діяльністю (мобільність людини, відкритість ринків і країн, воля бізнесу й конкуренції).

У працях відомого вченого-теоретика Пітера Дракера [74], що займається технологіями керування, знання визначаються як ключовий ресурс світової економіки. На його думку, такі традиційні фактори виробництва, як земля, праця й капітал, тепер стають скоріше його обмежниками, а от знання перетворюються в основний двигун. Щоб йти в ногу зі швидко мінливим ринком, потрібно усвідомлювати зміст зрушень, що відбуваються, і миттєво відбивати їх у виробничому процесі. Звичайно, і раніше виготовленню товару передувала фаза його розробки відповідно до ретельно проаналізованих ринкових вимог. Сьогодні ж до цієї попередньої фази додаються інтегровані у виробничий процес дії, що забезпечують оперативне введення мікрозмін, тобто доведення товару по гарячих слідах ринкових змін.

Інший важливий аспект знань як фактора виробництва (в тому числі і в освітньої сфері) - застосування отриманого досвіду.

В сучасних умовах вже немає часу на постійний «винахід велосипеда», тому вкрай важливим стає використання раніше знайдених рішень, а самі знання й досвід здобувають більшу цінність, чим матеріальні активи.

Більше того, цінність знань росте настільки швидко, що вже сьогодні вона становить основну частину загальної ринкової вартості багатьох компаній, особливо в галузях високих технологій.

Фактично це означає: для будь-якої організації важливіше всього не те, якими будинками, устаткуванням і сировиною вона розташовує, а те, який її потенціал створення перспективної конкурентоспроможної продукції. І цей потенціал тим вище, чим більше база знань компанії.

У даному зв'язку цікава поява організацій нової формації, чия діяльність являє собою сплав традиційного виробництва (позначуваного як brick-n-mortar - «цегла й цемент», що вказує на його приземленість, реальність) і інтерактивного менеджменту (pure play, або «чиста гра», заснована на нематеріальних знаннях).

Захват із приводу самого факту виникнення компаній «чистої гри» поступово стихає - приходить розуміння, що істина, як завжди, перебуває посередині. Симбіозну бізнес-модель позначають терміном click-n-mortar («щиглик і цемент», тобто керування реальним виробництвом за допомогою інформаційних технологій). Це особливо відноситься також до сфери освіти та наукової діяльності.

На думку фахівців аналітичної компанії Gartner Group, уже через декілька років тим фірмам, які не зможуть перейти до моделі click-n-mortar, прийде стикнутися із серйозними труднощами через різку втрату конкурентоспроможності (імовірність такого розвитку подій оцінюється Gartner Group досить високо - на рівні 70 %).

Зараз в США стрімко зростаючої ролі знань привело до надання істотних податкових пільг тим компаніям, які щорічно витрачають на навчання персоналу більше двох відсотків фонду заробітної плати, а також розвивають технології, що формують так зване «поле знань». У його рамках співробітники можуть ділитися досвідом, знаходити й обговорювати рішення виробничих, управлінських, маркетингових і інших проблем.

Підвищений інтерес до керування знаннями привів до того, що до даного процесу стали ставиться як до якогось містичного дійства, що дозволяє компанії здійснити якісний прорив і одержати значні конкурентні переваги.

На самому ділі зараз дуже багато прикладів створення реально працюючих систем управління знаннями, але не дуже багато таких проектів в освітньої сфері.

Головна мета національних проектів такого роду – створення могутнього інформаційно-аналітичного (комерційного) ресурсу по всіх областях знань, на багатьох мовах світу, для необмеженої кількості користувачів.

Основною метою таких проектів є:

- створення єдиного загальнонаціонального освітнього інформаційно-бібліотечного ресурсу по всіх областях знань;



- забезпечити зручний доступ для вивчення та дослідження до загальноосвітнього ресурсу, в тому числі до унікальних книг, рукописів, документів, що є національним надбанням України;
- створення першого національного ресурсу знань в сферах законодавства, права, економіки, науки, мистецтва та суспільних наук;
- на основі створення загальнонаціонального інформаційно-бібліотечного порталу (Web-бібліотеки) надати широкій спільноті необмежений доступ до інформаційних ресурсів бібліотеки, у тому числі і мультимедійних;
- на основі системи управління знаннями надати всім користувачам можливості більш глибокого та ефективного вивчення області дослідження, пошуку знань і аналізу закономірностей;
- інтеграція з Державними архівами, іншими бібліотеками та бібліотеками навчальних закладів для розширення інформаційної бази проекту;
- створення розподіленої загальнонаціональної інтегрованої інформаційно-бібліотечної системи в межах держави

Основними задачами є:

- істотно розширити інформаційну базу за рахунок переводу всіх документів в електронний вигляд;
- створити могутній і продуктивний загальнонаціональний інформаційно-бібліотечний портал;
- забезпечити повний і швидкий доступ до інформації широкому загалу користувачів та читачів (як в читальних залах так і назовні);
- надати користувачам аналітичний інструмент для ефективної навчальної та дослідницької роботи в різних областях знань;
- розробити гнучку комерційну політику в організації роботи із цими інформаційними ресурсами;
- забезпечити можливість створення та розповсюдження інформаційних баз на компакт-дисках по різних областях знань;
- забезпечити багатомовну підтримку при роботі з документами;
- розширити зв'язки та співпрацю з іншими бібліотечними та архівними установами для інтеграції і збільшення бази знань.

Основні користувачі таких систем:

- аналітики і експерти, фахівці і службовці;
- співробітники аналітичних і дослідницьких центрів і інститутів;
- журналісти ЗМІ;
- співробітники і фахівці наукових і учбових закладів;
- фахівці культури і мистецтва;
- учні та батьки;
- інші групи населення, які залучені до проекту.

Інформаційна база системи доцільно створювати на базі електронних бібліотек України, бібліотек навчальних закладів, фонди яких можуть складати мільйони документів, періодичних видань і матеріалів,

авторефератів і дисертацій, колекцій рідкісних книг, різних документів по всіх областях знань на багатьох мовах світу.

Для досягнення головної мети по створенню «Національної електронної бібліотеки» доцільно охопити найширше коло можливих областей знань і сфер діяльності людини:

- освітні ресурси дошкільних навчальних закладів, шкіл, гімназій та вузів;

- законодавство;
- право;
- економіка;
- бізнес;
- фінанси;
- політика;
- суспільні і точні науки;
- військові науки;
- мистецтво;
- історія;
- українська і світова література;
- багато що інше.

Інформаційний простір формується з наступних видів джерел:

- освітні ресурси дошкільних навчальних закладів, шкіл, гімназій та вузів;

- книги по різних областях знань;
- законодавчі, нормативні і правові акти;
- монографії;
- листи;
- раритетні, рідкісні і унікальні книги та документи;
- матеріали конференцій, з'їздів тощо;
- наукові праці, дисертації та дослідження;
- архівні матеріали;
- художня література;
- матеріали ЗМІ;
- друкарські документи різного роду;
- фотографії;
- репродукції картин і інших витворів образотворчого мистецтва;
- аудіо записи;
- відео зображення і сюжети;
- існуючі електронні бази і архіви.

Які ж задачі повинна вирішувати така система. Фактично система повинна забезпечити якісний та швидкий доступ до будь-якого інформаційного ресурсу організації. На даному етапі в галузі управління бібліотечними та освітніми ресурсами існують чіткі задачі, які системи керування знаннями вирішують:

- консолідація та інтеграція всієї інформації, в якій зацікавлена організація, яка реально існує та до котрої є доступ (зони інтранет та Інтернет);

- створення єдиного інформаційного простору для забезпечення ефективної навчальної, наукової, методичної та аналітичної роботи всіх груп користувачів;

- автоматичне розподілення неструктурованої інформації по ієрархічних онтологіях та таксономіях, завчасно створених на основі наукових знань та життєвого досвіду за напрямками і темами аналізу та дослідження;

- інформаційна й конкурентна розвідка по всім напрямкам діяльності;

- аналіз і створення баз знань за матеріалами ЗМІ, бібліотек та власних інформаційних ресурсів;

- поширення результатів пошуку інформації (обмін знаннями) між користувачами;

- створення спеціалізованих територіально розподілених інформаційно-аналітичних систем за напрямками діяльності, цілями та задачами;

- забезпечення інформаційної безпеки та контроль доступу до інформації;

- забезпечення користувачів необмеженою інформацією з Інтернету без безпосередньої роботи в мережі (інформація із заданих вузлів Інтернету доставляється в локальну мережу підприємства й безпосередньо на робочі місця співробітників);

- організація ретроконверсії документів (переклад в електронний вид), з поточною класифікацією та ранжируванням;

- створення і керування електронними інформаційними архівами;

- забезпечення пошуку інформації на бібліотечних та освітніх Інтернет-порталах та інших Інтернет ресурсах;

- при потребі - уведення, зберігання, пошук і витяг аудіо-відео- та іншої мультимедіа інформації;

- виконання аналіз повноти наявної інформації в інформаційних ресурсах;

- створення і управління електронними архівами текстових, фото, відео та аудіо документів із автоматичною класифікацією і ранжируванням;

- пошук і витягання інформації по смислах та непрямих ознаках;

- ведення досліджень по різних темах (політика, економіка, наука, освіта, мистецтво тощо);

- аналіз та порівняння документів за змістом;

- багато чого іншого.

На нашому мовному ринку пошукових систем, які б відповідали вимогам систем керування знаннями, реально працюючих систем практично немає.

Одна з таких систем, яка присутня на українському ринку – це інформаційно-аналітична платформа французької системи Exalead.

Сучасна корпоративна високопродуктивна інформаційно-пошукова платформа Exalead (на базі якій створюється система керування знаннями), через яку клієнти без додаткових налаштувань і використання інших систем отримують доступ до інформації може бути розташована на робочих станціях, серверах корпоративної мережі, базах даних або в мережі Інтернет.

Загальні особливості та рішення, на яких базується робота системи:

- Універсальність - система є універсальною та легко адаптується до предметних галузей клієнтів, які є описом прикметних областей або видів діяльності людини, або бути описом приватної точки зору дослідника;

- Актуальність словників - система створення таким чином, що постійно забезпечується постійна актуальність словників системи. Різні словники, такі як морфологічні, фразеологічні, тезауруси ведуться системою автоматично та поповнюються в режимі реального часу. Для різних мов або незнайомих слів системою використовується алгоритм статистичної морфологічної лематизації;

- Гнучкість в створенні лінгвістичних ресурсів – в системі існує можливість вручну створювати і налаштувати словники системи (онтології, тезауруси, синоніми, стоп-слова тощо). Ці лінгвістичні ресурси гнучко підключаються до системи (з урахуванням потреб аналітичної або дослідницької роботи (до кожної тематичної бібліотеки може бути підключений свій набір лінгвістичних ресурсів, що забезпечує звуження зони пошуку ));

- Мови з якими працює система - система забезпечує роботу з документами (текстами, html-сторінками, базами даних – всього більш 350 форматів) на усіх європейських мовах, арабській, китайській, в тому числі на українській та російській мовах;

- Семантичний аналіз тексту - первинна лінгвістична обробка тексту виконується за допомогою поверхневого семантичного аналізу з виділенням базового переліку семантичних відношень. Перелік семантичних відношень має ієрархічну структуру з можливістю локального уточнення та доповнення. Семантичні відносини типу людина, географія, організація система має в базовій конфігурації, крім того в системі передбачено динамічний тезаурус, якій в режимі роботи системи постійно оновлюється. Семантичні відносини – категорії додатково створюються и конфігуруються в залежності від задач, які мають бути вирішені системою під конкретну дослідницьку задачу.

Використання системи дозволить підняти на якісно новий рівень проходження учнями загальноосвітніх шкіл поглиблених курсів вивчення окремих навчальних предметів: від загальнопізнавальних етапів до виконання наукових досліджень під керівництвом педагога та науковця-експерта у певній галузі знань.

Принциповим у системі є широке, мабуть максимально можливе, використання лінгвістичних ресурсів мов, які необхідні для роботи з інформацією (мови як документів так и користувачів):

- Для створення відповідних семантичних мереж пропонується використати сучасні методи категоризації розподілених тематичних інформаційних ресурсів та клас систем інтелектуального аналізу природно-мовних текстів;

- Категорії (це можуть бути онтології та таксономії), які розробляються для інформаційного середовища, можуть описувати будь-яку необхідну предметну область. Вони дозволяють відображати модель предметної області у вигляді орієнтованого графу. Опис категорії створюється спеціалістом в предметній області (експертом) за допомогою спеціальних програмних засобів. Це суттєво спрощує її створення і дозволяє одночасно при створенні категорії формувати списки термінів-об'єктів, які зв'язуються відповідними відношеннями. Таки описи об'єднуються у відповідні вихідні файлові структури де у спеціальному порядку описуються тематичні фрагменти предметної області;

- Системи інтелектуального аналізу природно-мовних текстів, як засіб автоматичного аналізу текстових документів, дозволяють визначати тематику тексту, формувати тематичні конспекти, здійснювати узагальнення понять, що визначені в текстах, представляти узагальнені поняття у вигляді логікових формул, автоматично візуалізувати побудовану мережу понять. Використання вказаного класу систем прискорює процеси адаптування до предметних областей, тематика яких використовується у освітньої сфері;

- Створювана система повинна забезпечити постійну актуалізацію словників за тематикою предметних областей, включаючи такі як морфологічні, фразеологічні, тезауруси, Вони повинні вестись автоматично та поповнюватись у режимі реального часу. Для незнайомих слів повинен використовуватися алгоритм статистичної морфологічної лематизації. В той же час повинна існувати можливість ручного створення та налаштування словників системи;

- Первинна лінгвістична обробка тексту повинна виконуватись за допомогою поверхневого семантичного аналізу з виділенням базового переліку семантичних відношень. Перелік семантичних відношень повинен мати ієрархічну структуру з можливістю локального уточнення та доповнення.

Система забезпечує роботу з документами більш ніж в 350 форматах, які можуть зберігатися в локальній мережі або Інтернет, в базах даних, поштових системах і системах документообігу, а також в електронних архівах.

Система має можливість працювати із документами більш ніж на 50 мовах, та забезпечити крос-мовну роботу.

Система управління знаннями може бути і повинна бути інтегрована до інформаційних та аналітичних систем, що вже використовуються в освітньої галузі України.

Інноваційна значущість таких проектів для України є найважливішим кроком по створенню інформаційного суспільства нашої держави. Тому що інформаційне суспільство це є сучасним етапом розвитку цивілізації із

домінуючій роллю знань та технологій та суттєвим впливом інформаційно-комунаційних технологій на всі сфери людської діяльності та суспільства в цілому. Особливими рисами інформаційного суспільства є:

- збільшення ролі інформації, знань та інформаційних технологій у житті суспільства;

- зростання кількості людей, які зайняті інформаційними технологіями, комунікаціями та виробництвом інформаційних продуктів та послуг, зростання їх долі в валовому внутрішньому продукті;

- зростаюча інформатизація суспільства з використанням телефонії, радіо, телебачення, мережі Інтернет, а також традиційних та електронних ЗМІ;

- створення глобального інформаційного простору, якій забезпечив би:

- а) ефективно інформаційну взаємодію людей;

- б) їх доступ до світових інформаційних ресурсів;

- в) задоволення їх потреб в інформаційних продуктах та послугах.

Проект по створенню Національної системи електронного освітнього інформаційно-бібліотечного ресурсу може бути першим проектом такого масштабу на пост-радянському просторі. Особливого значення він набуває зараз, в період становлення Української державності в умовах демократизації внутрішнього життя і формування світової громадської думки.

Користувачами системи можуть стати мільйони людей з усього світу. Їм стануть доступні багато книг, творів, унікальних документів з законодавства, права, історії України і мистецтва, національні надбання України, в тому числі освітні ресурси.

## РОЗДІЛ 2. КОНЦЕПТУАЛЬНІ АСПЕКТИ ПОБУДОВИ МОДЕЛЕЙ ОНТОЛОГІЧНОГО ОПИСУ ОБ'ЄКТІВ І ПРОЦЕСІВ

### 2.1. Дослідження предметно-орієнтованих знань

#### Концептуальний аналіз предметних знань

##### *Онто-гносеологічний аналіз предметних знань*

Онто-гносеологічний аналіз знань взагалі, і предметних знань зокрема, припускає використання в єдності дослідницьких методів онтології, гносеології та логіки. Перш ніж перейти до детального розгляду науково-практичних задач аналізу знань, звернемося до його філософських аспектів, а точніше до тієї сукупності спільних наукових дисциплін, підходи, методи та принципи яких безпосередньо впливають на формування і становлення зазначеного вище аналізу. Такими дисциплінами є *Онтологія*<sup>1</sup>, *гносеологія* (або *теорія пізнання*), *загальна логіка* і *теорія відображення* [1]<sup>2</sup>.

Зміст зазначених дисциплін, в тій чи іншій мірі, характеризує тріаду {Суб'єкт (*S*) – Об'єкт (*O*) – Відношення (*R*)}. Під *суб'єктом* розуміється щось таке, що діє і пізнає (в якомусь роді синонім до терміну *Спостерігач*). У більш широкому сенсі під суб'єктом можна розуміти систему будь-якого рівня, яка впливає та відображає. Під *об'єктом* розуміється система, яка відображається та піддається впливу, щось пізнаване і змінюване суб'єктом. З точки зору системи  $\Lambda = \langle O, S, R \rangle$  відомі наступні (нестрогі) визначення згаданих вище наукових дисциплін [1].

**Онтологія** – це вчення про загальні властивості і відношення об'єктивного світу, теорія загального, яка розглядає об'єкт поза його відношень до суб'єкта.

**Логіка** (загальна) – це вчення про загальні структури пізнання суб'єкта, що розглядаються як такі, поза їх відношенням до об'єкту.

**Теорія відображення** вивчає таке відношення, яке відображає об'єкт до суб'єкта. Пізнання – окремий випадок і вища форма відображення, а тому **теорія пізнання** – є результат специфікації загальної теорії відображення.

Роль зазначених вище наукових дисциплін розглянемо на прикладі, де в якості об'єкта виступає наукова теорія.

*Логіка* розглядає структуру цієї теорії незалежно від того, що вона відображає, які об'єктивні основи цієї структури. З позицій загальної *теорії відображення* знадобиться охарактеризувати те загальне ставлення *R* між суб'єктом *S* і об'єктом *O*, яке виражається в теорії. Синтез теорії відображення і логіки для цілей теорії пізнання дає загальну **методологію**. Вона «знає» і об'єктивну спрямованість, і логічну структуру пізнавальних

<sup>1</sup>Цей термін написаний з великої літери, тому що в данному контексті він застосовний до онтології як філософської дисципліни, щоб відрізнити від комп'ютерної онтології.

<sup>2</sup>Сказане можна рівною мірою віднести і до двох інших аспектів дослідження – методологічного і логічного[1].

операцій, але абстрагується від соціальних і психологічних умов, їх генезису та функціонування.

Загальна взаємодія перелічених наукових дисциплін і системи  $\Lambda$  представлена на рисунку 2.1.

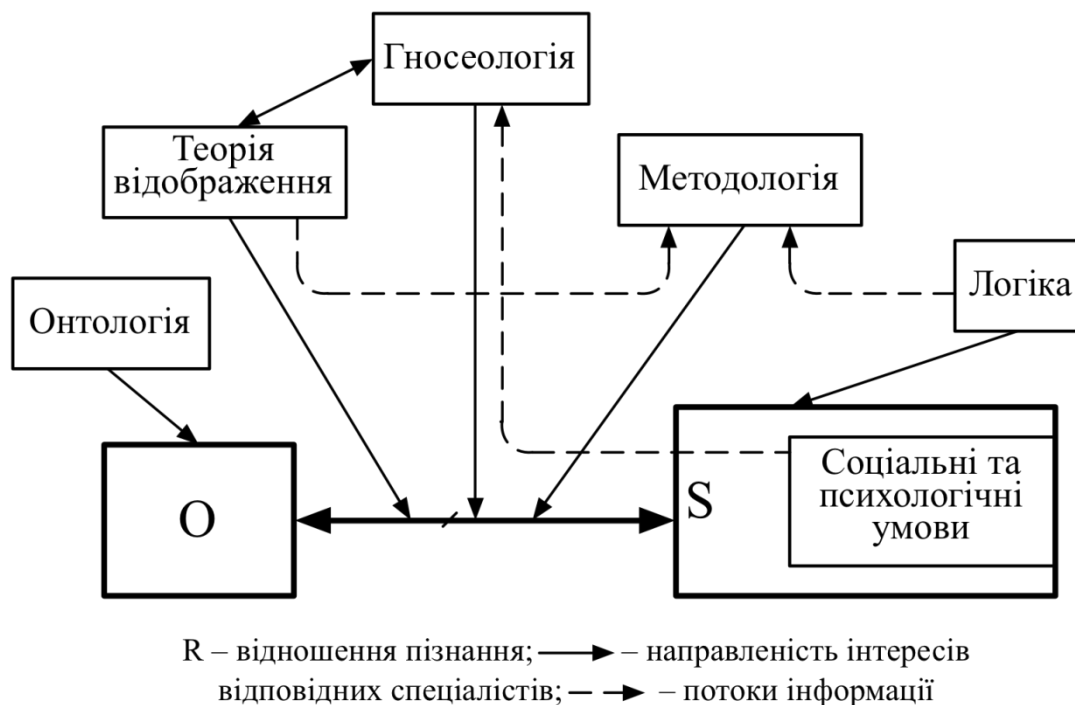


Рис. 2.1. Загальна структура і взаємозв'язки наукових теорій, що вивчають систему «об'єкт  $\leftrightarrow$  суб'єкт»

Таким чином, у системі  $\Lambda$  різні фахівці, що займають різні позиції, вивчають різні предмети.

Звичайно, описана модель взаємодії наукових дисциплін і системи  $\Lambda$  є наближеною. В ній не розглянуто декілька уточнюючих моментів, зокрема:

- 1) межі загальної логіки;
- 2) процес впливу суб'єкта на об'єкт;
- 3) у загальному випадку суб'єкт і об'єкт можуть мінятися місцями;
- 4) як застосування математичних методів у методології породжує нову дисципліну – **технологію** пізнання, де методологія переходить від загальних принципів аналізу до процесу синтезу (проектування).

Якщо система  $\Lambda$  розглядається на абстрактному рівні, то її аналізом повинна займатися онтологія як філософська дисципліна. Якщо в якості об'єкта вибрана деяка ПдО (що представляє найбільший практичний інтерес), то його розглядом повинна займатися «комп'ютерна онтологія предметної області» (О ПдО). Аналогічним чином можна перейти до онтології домену прикладних областей, що інтегрує множину взаємопов'язаних предметних областей, і, нарешті, – до онтологій процесів і задач [2–14].



## *Загальносистемний аспект дослідження «піраміди» знань*

Особливість створення сучасних інформаційних технологій полягає в прагненні до максимальної інтеграції результатів двох областей штучного інтелекту, які колись розвивалися паралельно і незалежно: knowledge-engineering і комп'ютерної лінгвістики (когнітивної семантики). Можна сказати, що це прагнення відображає в загальному випадку природну схему взаємодії людини з навколишнім світом: *вхідна інформація* → «свідомість – розуміння – знання». В останній не вказані проміжні (відчуття, рефлексія та ін.) процедури когнітивного циклу, щоб виділити «конструктивну» тріаду. Свідомість тут грає роль персоніфікованого інструменту, який виробляє сукупність предметно, ситуативно або причинно-зв'язаних сутностей, які утворюють «свідому» картину світу. Вироблення зазначених сутностей, побудова і використання на їх основі картини світу реалізується як результат сприйняття і перетворення вхідної інформації у когнітивному циклі, ядром якого є *механізм розуміння*. Свідомість людини має мовний статус, і це виражено поняттям «мовна свідомість», а свідомо картина світу при цьому представлена «мовною» картиною світу, яка є не тільки системою знань, а й деяким буфером, який пов'язує загальні знання з професійними. Для розуміння природної мови (ПМ) в прагматичному плані важливо виділити відповідний механізм, який характеризується як «... перетворення подання на природній мові в логічні висловлювання» [10,15-17].

*Архітектуру МКС* в загальному вигляді можна представити як ієрархічну мережу, структура якої, спираючись на базові категорії пізнання і сфери буття (простір, час, рух, якість, кількість, відношення, окреме) або більш місткі (матерія, свідомість, практика), містить три базових взаємопов'язаних рівня:

- категоріальний (концептуальний);
- понятійного тлумачення (слів і словосполучень);
- граматичного розбору.

Перший орієнтований на формалізацію побудови і правил виведення на семантичній (концептуальній) мові понять. Другий може бути аналогізовано з сучасними тлумачними словниками (в електронному виконанні), але володіючий замкнутою системою зв'язків між лексемами в розвинених лексичних структурах і сильною взаємодією з верхнім (концептуальним) рівнем, при якому останній забезпечує понятійне тлумачення лексем і зняття багатозначності. Рівень граматичного розбору забезпечує інформацію про морфологічні (рід, число, відмінок, вид, час і т. д.), синтаксичні (атрибутивні, кількісні, обставинні, службові тощо) і семантичні (об'єктні, ситуаційні, ознакові, актантно-ролеві, лексико-функціональні) характеристики лексем та їх семантико-синтаксичних конструкцій. Технологічною основою МКС можуть бути моделі реляційних баз даних, що забезпечують функції взаємодії у вигляді «ПМ-SQL» запитів в якості природно-мовного інтерфейсу [11].

Таким чином, МКС можна представити у вигляді віртуального індексованого інформаційного поля таблиць лексико-семантичних відношень (ціле-частина, рід-вид, множина-елемент, причина-наслідок, об'єкт-функція тощо) та їх компонентів (стовпчиків і рядків) з наборами семантичних характеристик лексем і складніших лексичних конструкцій. Основою верхнього рівня МКС є відношення типу: рід-вид, ціле-частина. Найбільш великими таксономічними категоріями нижніх рівнів є семантико-частиномовні характеристики відношень.

Знання в конкретній ( $i$ -й) предметній області ( $\text{ПдО}_i$ ) є  $i$ -е розширення загальної картини світу, так що повну картину світу (або піраміду знань) можна представити у вигляді, показаному на рисунку 2.2, тут  $i$  проходить через позиції філософської, біологічної, хімічної, фізичної і т. д. часткових картин світу (КС) або, що те ж саме, наукових знань у кожній з названих предметних областей. При цьому кожна з них ( $\text{КМ}_i = \text{ПдО}_i$ ) будується за тими ж законами системного підходу (цілісності, ієрархії, класифікації і структурування), що і МКС. Застосовуючи методи системних переходів [11] можна процедуру, що реалізує перетворення за формулою Брукса ( $\text{K}(\text{S}) + \Delta\text{I} = \text{K}(\text{S} + \Delta\text{S})$ , де  $\text{K}(\text{S})$  – початкові знання,  $\Delta\text{I}$  – нова порція інформації,  $\text{K}(\text{S} + \Delta\text{S})$  – вихідні знання), сформулювати в термінах *узагальнення сенсу*, тобто ( $i \rightarrow j$ ) –переходів, з чого випливає, що нові, невідомі раніше знання, по суті, являють собою певне узагальнення (інтерпретацію) вже відомих знань у категоріях вищих рівнів ієрархії їх подання. Насправді процес розвитку знань більш складний і глибокий. По-перше, на відміну від процедури узагальнення в структурі МКС, наукове узагальнення відбувається в умовах, коли  $j$ -й рівень ієрархії не визначений. По-друге, отримання нових знань про навколишній реальний світ часто пов'язано зі зміною уявлень (запереченням) певних старих знань. Цей процес можна схематично проілюструвати (рисунок 2.3).

Тут  $j = 1, 2, \dots$  – рівні узагальнення знань: від елементарних (аксіоматичних) знань до повних; стрілки вказують напрям розвитку знань:  $\rightarrow$  – в бік розширення елементарних знань,  $\uparrow$  – в бік узагальнення існуючих знань;  $d\text{I}_j, d\text{I}_i$  – збільшення знань по відповідним осям координат. Зона, обмежена перетином двох кривих, – це область знань, інваріантних щодо нових і старих теорій. Зони  $\text{S}_c$  і  $\text{S}_n$  – це області взаємовиключних розділів теорій. Ця інтерпретація процесу розвитку знань може бути використана як в рамках однієї предметної області, так і для опису повної картини світу. В останньому випадку нетривіальними представляються міждисциплінарні функції МКС в плані креативної взаємодії  $\text{ПдО}_i$  і процесів конвергенції їх теорій і технологій (рисунок 2.2).

Мова йде, зокрема, про побудову формальної мови категоріального рівня, що забезпечує ефективне узагальнення, перенесення і впровадження прийомів, методів і теорій з однієї  $\text{ПдО}$  в іншу, що, мабуть, підвищить ефективність процесу міждисциплінарної взаємодії. Як впливає розглянутого вище аналізу взаємодії спільних наукових дисциплін, саме

теорія пізнання визначає основні принципи та напрямки дослідницького процесу в системі  $\Lambda$ . Разом з тим важливо відзначити радикальну зміну і самої системи наукового пізнання.

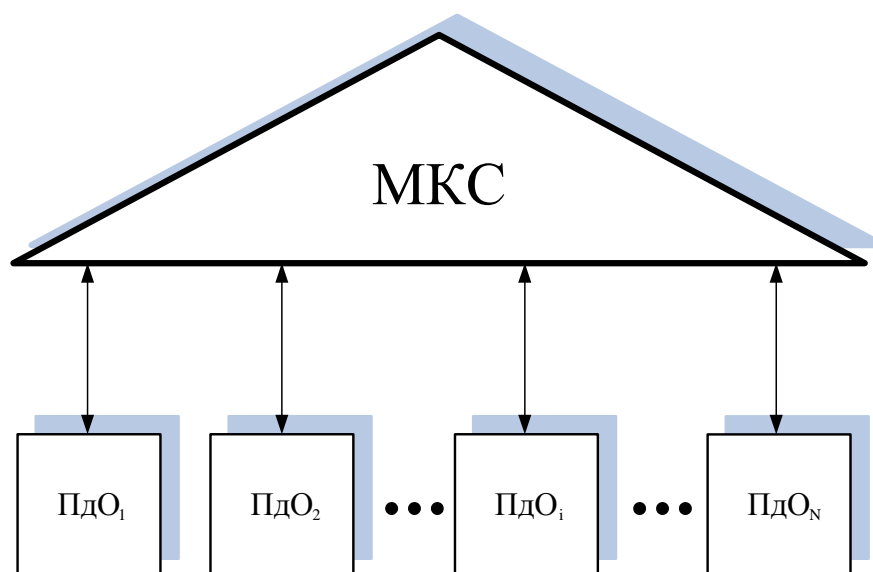


Рис. 2.2. «Піраміда» знань – повна картина світу

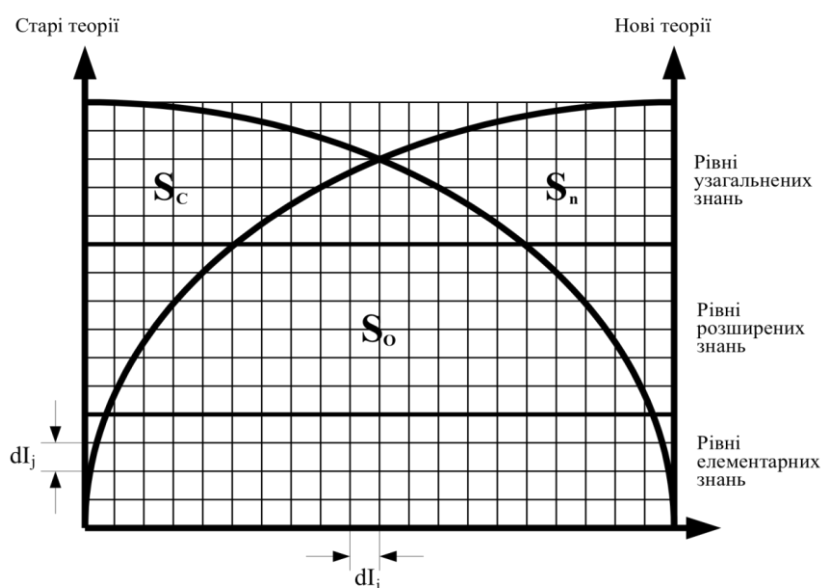


Рис. 2.3. Розвиток знань

Розмиваються чіткі межі між практичною та пізнавальною діяльністю. В системі наукового знання відбуваються інтенсивні процеси їх диференціації та інтеграції, розвиваються комплексні міждисциплінарні і трансдисциплінарні дослідження, нові способи і методи пізнання, методологічні установки, з'являються нові елементи картини світу, виділяються нові, більш складні типи об'єктів пізнання, які характеризуються історизмом, універсальністю, складністю організації, які

раніше не піддавалися теоретичному (математичному) моделюванню. Зауважимо, що проблеми і закономірності зазначеної взаємодії досліджуються в рамках нового міждисциплінарного напрямку – *синергетики*, яке, взагалі кажучи, можна розглядати як сучасний етап розвитку кібернетики і системних досліджень.

**Синергетика** (від грецького *συν* – «спільно» і *εργος* – «діючий») – міждисциплінарний напрямок наукових досліджень, задачею якого є вивчення природних явищ і процесів на основі принципів самоорганізації систем (що складаються з підсистем) [12–15]. Синергетика спочатку позиціонувалася як міждисциплінарний підхід, як принципи, що керують процесами самоорганізації, одні й ті ж безвідносно до природи систем. Нас же цікавлять конкретні науково-практичні завдання, в їх конкретній постановці – розробка знання-орієнтованих систем, що розвиваються, відкритих баз знань базових доменів прикладних областей та їх поповнення, розробка ефективних механізмів отримання знань, їх виявлення і формального подання.

## **2.2. Концептуальна парадигма роботи із знаннями**

Свого часу В.М. Глушков в якості одного з основних напрямків розвитку кібернетики виділив *пошук нових принципів побудови ЕОМ* [15, 46]. В рамках цього напрямку, в результаті «еволюції людського знання, що нагадує сходження по спіралі або сходами з великими сходинками» була сформована і розвивається *концептуальна парадигма роботи із знаннями*. Це поняття є складним, спирається на два поняття, що визначаються термінами: «концептуальна парадигма» і «робота із знаннями».

"Парадигма [46] часто виглядає як проста, елегантна і правдоподібна концептуалізація даних, на підставі якої можливо пояснення більшої частини спостережуваних фактів. Вона висловлює деяке переконання, критерій цінності, технічний прийом, що розділяється і вживається членами певного суспільства. Парадигма настільки ж істотна для науки, як спостереження і експеримент.

Після того як парадигма сформована і починається поступальний розвиток науки, більшість відомих фактів, як правило, знаходять пояснення. Проте в процесі досліджень і спостережень неминуче виявляються такі феномени, які не знаходять свого пояснення. Їх стає все більше і більше, що призводить до необхідності зміни існуючої парадигми. Так назріває нова революція і підйом на новий щабель розвитку знань.

Таким чином, парадигми несуть в собі не тільки пізнавальний, але і стверджуючий характер і тому істотно впливають на процес розвитку науки, прискорюючи або сповільнюючи його. ... Коли парадигму приймає більша частина наукового співтовариства, вона стає нормативною точкою зору і інструментом – потужним каталізатором наукового прогресу».

Термін «робота із знаннями» за належністю відноситься до галузі інженерії знань, а за належністю до онтологічної категорії – до поняття-процесу. Останній загальноприйнято визначати екстенціонально, через перерахування назв підпроцесів. До таких підпроцесів роботи із знаннями, зазвичай, відносять *вилучення та добування знань, представлення знань і маніпулювання знаннями*. В область аспектів роботи із знаннями, крім зазначених підпроцесів, включають також методи і засоби. В онтологічному інжинірингу в цю область включають онтологію процесів, онтологію задач і онтологію об'єктів [47–52, 82-99].

**1. Вилучення знань** з різних джерел, у тому числі з природномовних об'єктів (ПМО), включає два основні розділи: аналіз вихідної інформації, формалізацію якісних знань й інтеграцію знань. Перший розділ пов'язаний зі створенням методів, що дозволяють переходити від знань, виражених, в тому числі в природномовній формі, до їх аналогів, придатних для введення в пам'ять знання-орієнтованої інформаційної системи (ЗОІС). Другий розділ пов'язаний з інтеграцією знань, добутих від різних джерел, в деяку взаємопов'язану і несуперечливу систему знань про предметну область. Без вирішення цієї проблеми навряд чи можливо створити повне уявлення про ПдО.

**2. Добування знань.** Знань, що містяться в джерелах інформації, відчужених від фахівця, як правило, недостатньо. Значну частину професійного досвіду ці фахівці не можуть висловити словесно (професійне вміння або інтуїція). Тому для того, щоб добути такі знання, потрібні спеціальні прийоми і методи. Вони використовуються в інструментальних системах добування знань, створення яких – одне із завдань інженерії знань. Отримані від експертів знання потрібно оцінити з точки зору їх відповідності раніше накопичених знань і формалізувати для введення в пам'ять системи. Крім того, знання, отримані від різних експертів, необхідно узгодити між собою, тому що нерідкі випадки, коли ці знання виявлялися зовні несумісними і навіть суперечливими. Розглянутий підпроцес включає такі розділи як *організація роботи з експертами, оцінка і формалізація знань і узгодження знань*.

**3. Представлення знань.** Цей процес передбачає розробку формальної наукової теорії, що включає побудову *моделі знань, системи подання знань і бази знань* (БЗ). В інженерії знань системи подання знань включають сукупність процедур, необхідних для запису знань, вилучення їх з пам'яті і підтримки збереження знань в робочому стані. Системи подання знань оформлюються як БЗ, є природним розвитком баз даних (БД). Представлення знань – це угода про те, як і в якій формальній теорії описувати реальний світ. У природничих і технічних науках прийнятий наступний традиційний спосіб представлення знань. Природною мовою вводяться основні поняття і відношення між ними. Але при цьому використовуються раніше визначені поняття і відношення, зміст яких вже відомий. Далі встановлюється відповідність між характеристиками (найчастіше кількісними) понять і

підходящої математичної моделі. Основна мета подання знань – будувати математичні моделі реального світу і його частин.

*Системою подання знань* називають сукупність засобів, що дозволяють:

- описувати знання про предметну область за допомогою мови представлення знань;
- організувати зберігання знань у системі (накопичення, аналіз, структурне узагальнення та організація знань);
- виводити нові знання з існуючих і об'єднувати їх;
- знаходити необхідні знання;
- поновлювати знання;
- здійснювати інтерфейс між системою і користувачем.

**4. Маніпулювання знаннями.** До цього процесу відносяться такі розділи як *поповнення знань, класифікація знань, узагальнення знань, виводи на знаннях* (резольюційні методи, квазіаксіоматичні системи і системи правдоподібного виводу), *міркування за допомогою знань, пояснення на знаннях, вирішення прикладних задач ПДО.*

Нові знання, які надходять у БЗ, повинні разом з тими відомостями, які вже були раніше записані в неї, сформулювати розширення знань, які надійшли. Серед цих процедур особливе місце займають псевдофізичні логіки (часу, простору, дій та ін.), які, спираючись на закони зовнішнього світу, поповнюють БЗ інформацією, що надійшла. Знання ЗОІС утворюють впорядковані структури, що полегшує пошук необхідних знань і підтримку дієздатності БЗ. Для цього використовуються різноманітні класифікаційні процедури. Типи класифікацій можуть бути різними: родо-видові, «частина-ціле», ситуативні (коли в одну множину об'єднуються знання, релевантні деякій типовій ситуації). Зазначені процедури будуються на основі *теорії класифікацій*. У процесі класифікації часто відбувається абстрагування від окремих елементів описів (фрагментів знань про об'єкти або явища), з'являються узагальнені знання, що призводить, врешті-решт, до абстрактних знань, для яких немає прямого прообразу в зовнішньому світі.

Вивід на знаннях залежить від моделі, яка використовується для їх представлення. Якщо в основі представлення використовуються логікові системи або продукції, то вивід на знаннях стає близьким до стандартного логічного виведення. Це ж відбувається при поданні знань у каузальній формі. У всіх цих випадках в ЗОІС використовуються методи виведення, що спираються на ідеї методу резолюцій.

Можливість появи в пам'яті ЗОІС нових фактів і відомостей приводить до того, що починає порушуватися принцип монотонності, що лежить в основі функціонування всіх систем, що вивчаються традиційною математичною логікою. Немонотонність виведення у відкритих системах викликає немалі труднощі. В останні роки прихильники логічних методів у штучному інтелекті (ШІ) роблять спроби побудувати нові логічні системи, в рамках яких можна було б забезпечити немонотонне виведення. По суті,

системи, за допомогою яких представляються знання про предметні області, не є строго аксіоматичними, як класичні логічні числення. Тому й системи, які виникають при таких умовах, слід називати *квазіаксіоматичними*. У таких системах цілком можлива зміна вихідних аксіом в процесі виведення. І, нарешті, ще одна особливість виведення на знаннях – неповнота відомостей про предметні області та процеси, що в ній протікають, неточність вхідної інформації. А це означає, що виводи в ЗОІС носять не абсолютно достовірний характер, як у традиційних логічних системах, а наближений, правдоподібний характер. Такі виводи вимагають розвиненого апарату обчислення оцінок правдоподібності і методів оперування ними.

Оскільки ЗОІС приймають рішення, спираючись на знання, які можуть бути невідомі користувачеві, що вирішує своє завдання за допомогою ЗОІС, то він може засумніватися в правильності отриманого рішення. ЗОІС повинна володіти засобами, які можуть сформулювати користувачеві необхідні пояснення. Пояснення можуть стосуватися процесу отримання рішень, підстав, які були для цього використані, способів відсікання альтернативних варіантів та ін. Все це вимагає розвиненої *теорії пояснення*.

На основі проведеного вище аналізу на рисунку 2.4 синтезовано узагальнену схему введення і обробки вхідних даних, семантичної переробки інформації і роботи із знаннями. В ній також відображені зв'язки між відповідними процесами, послідовності яких складають інформаційні технології та обробки даних, текст-процесингу, роботи із знаннями і логіко-інформаційного підходу [18, 19, 100-114, 118-130], що становлять основу методології проектування знання-орієнтованих інформаційних систем з обробкою ПМО і формалізованих знань.

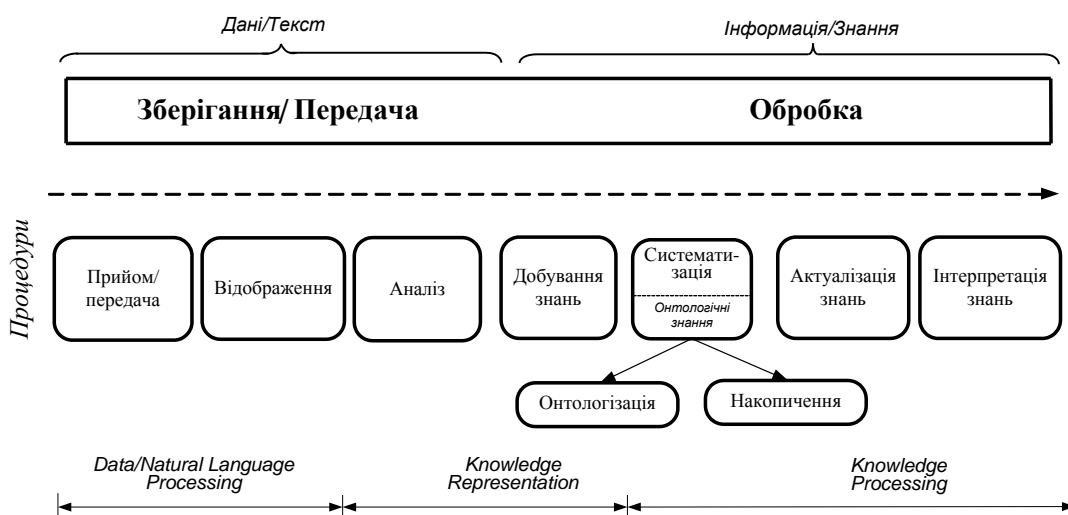


Рис. 2.4. Схема роботи із знаннями

Концептуальна парадигма роботи зі знаннями утвердилася у зв'язку і в міру розвитку концепції інтелектуальних інформаційних систем (ІС), згідно з якою головною інформаційною одиницею комп'ютерної обробки стало «знання». Незалежно і в рамках різних областей предметних знань

розвиваються інформаційні технології: автоматичної обробки природної мови – в комп'ютерній лінгвістиці, представлення знань – в математичній логіці. Почали з'являтися наукові розробки, що пропонують конструктивні підходи до інтеграції зазначених технологій.

### **Системно-сутнісний аналіз знань**

Розглядаючи знання як щось об'єктивне, феноменальне, що відображає *вихідну, перетворюючу і кінцеву* складові процесу пізнання («знання за допомогою знань перетворюється на нове знання»), зауважимо, що сукупність наукових дисциплін, які мають безпосереднє відношення до процесів мислення, розуміння, усвідомлення та здобуття нових знань, не може обходитися без вироблення своїх власних понять, певним способом фіксуючих властивості і закономірності її об'єктів. З таких основоположних понять, як *смысл, знання, (знакова) система, текст, об'єкт, ставлення, предмет, мова, структура, зв'язок*, розглянемо тільки перші два поняття, які виражають собою найбільшу проблемність досліджень в області побудови систем обробки знань, в тому числі і ЗОІС, що оперують ПМО. Але, спершу розглянемо визначення категорії «поняття», що є «цеглинкою» для більш складних форм мислення – судження і висновку.

Поняття зазвичай визначають як одну з основних форм мислення; цим підкреслюється важлива роль його в пізнанні. А саме мислення може розглядатися як процес оперування поняттями; саме завдяки поняттям мислення набуває характеру узагальненого відображення дійсності. Виділення предметів – одна з основних функцій поняття в процесі пізнання. Виділені в поняттях класи предметів – це опорні пункти в пізнанні, навколо яких концентруються всі знання [20].

**Поняття** є цілісна сукупність суджень, в яких щось стверджується про відмітні ознаки досліджуваної сутності, ядром якої є судження (або твердження) про найбільш загальні і в той же час істотні ознаки цієї сутності.

Кожне поняття характеризується об'ємом і змістом. Об'єм і зміст поняття – дві взаємопов'язані сторони поняття. **Об'єм** – клас узагальнених в понятті предметів, **зміст** – сукупність (зазвичай істотних) ознак, за якими вироблено узагальнення і виділення предметів в даному понятті. Об'єм поняття є визначальним при формуванні ієрархічної структури відповідного онтографа (ОГ), а зміст – при аксіоматизації його (ОГ) вершин.

Відомо, що теорії смислів в завершеному вигляді не існує, а тому і загальноприйнятого визначення поняття «смысл» також немає, як немає однозначного розуміння і з точки зору його формального подання і перетворення в комп'ютерній системі. В [10, 11, 15, 20–22] цей термін вводиться з гносеологічних позицій, через його властивості та атрибути, взаємодія з іншими об'єктами. Ось найбільш поширені визначення з точки зору проектування комп'ютерних систем.

*Смысл* – це є те, що робить знакові системи текстом [23].

*Смысл* – деякий безперервний, невербальний конструкт, *осмыслення* – інтерпретація в індивідуальній концептуальній системі [11].



*Смисл* – це є мережа значень у певних позиціях і оперативний алгоритм для вирішення проблем [15].

І далі [23] зазначається, що «... природа смислу може бути розкрита тільки через одночасний аналіз *семантичної тріади*: «*сенс–текст–мова*». Текстове розкриття змісту відбувається через знакові системи, які ми сприймаємо як мови. Таким чином, кожен елемент зазначеної вище тріади розкривається через два інших. Включаючи в тріаду мову, ми вносимо представлення про те, що сама тріада стає можливою, тільки коли є СПОСТЕРІГАЧ – носій свідомості, що сприймає тексти і оцінює смисли. Тріада стає синонімом свідомості», скажемо точніше – мовної свідомості.

#### *Визначення 1.1*

1. **Розуміння** – це фіксація смислу природномовного об'єкта.
2. **Фіксація смислу** природномовного об'єкта – це проектування ПМО на мовну картину світу.
3. **Смисл** – це індекс проєкції ПМО на МКС.

Відомо також досить багато визначень поняття «знання», які за способом подання можна розділити на дві основні групи – загального характеру і адаптовані стосовно їх комп'ютерній обробці.

З першої групи можна виділити наступні, найбільш характерні визначення.

*Знання* – цілісна і систематизована сукупність наукових понять про закономірності природи, суспільства і мислення, накопичена людством в процесі активної перетворюючої виробничої діяльності і спрямована на подальше пізнання і зміну об'єктивного світу [21].

*Знання* є сукупність понять і представлень про об'єктивну дійсність, їх внутрішньо взаємозалежних систем (суджень, положень, концепцій, теорій і т. д.), що виробляються суспільством у процесі пізнання і перетворення світу. Зародившись в епоху первісного духовно-фізичного синкретизму трудової діяльності людини і продукується в подальшому у спеціалізованій (науковій) діяльності суспільства, в першу чергу для досягнення практичних цілей, знання все ширше охоплює об'єкт, все глибше проникає в його таємниці, тобто розвивається як в екстенсивному, так і в інтенсивному плані [24].

Ці приклади визначень досить коректні, дають загальне уявлення про те, що таке знання і пізнання як когнітивні процеси, але мало корисні для практичних додатків і проектування відповідних комп'ютерних систем.

З другої групи можна виділити наступні визначення, які визнаються багатьма дослідниками.

*Знання* – це інформація, структурована (формалізована) таким чином, що над нею можна робити логічні операції (насамперед логічного висновку) [25].

*Знання* – це складно організована інформаційна структура, збережена в пам'яті інтелектуальних інформаційних систем і включає в себе відомості про об'єкти та відношення предметної області, процеси взаємодії об'єктів в часі і в просторі, правилах провадження логічного виводу [3].

Визначення другої групи вже дають деяке узагальнене уявлення про систему, засновану на знаннях, але все ж цього недостатньо, щоб побудувати зазначену систему. Для цього необхідно також, принаймні, знати:

Чим знання відрізняється від подібних йому понять – даних та інформації:

- якими загальними властивостями володіють знання;
- які існують джерела знань;
- які існують способи представлення та обробки знань у КС;
- які існують механізми виявлення нових знань.

Тому в задачах інформатики можна (і більш зручно) визначити категорію знання опосередковано – через його властивості та методи обробки. Зокрема, стосовно до задачі отримання знань з тексту коректним буде.

*Визначення 1.2* Під терміном «знання» будемо розуміти сукупність фактів, добутих із заданого тексту і представлених у вигляді формул (можливо, деякого обмеженого фрагмента) мови предикатів першого порядку або іншої підходящої логічної системи, яка відображає смисл, закладений у вихідному тексті (передбачається існування моделі, на якій здобуте знання можна перевірити на істинність або хибність).

*Властивості знань.* Відомо [3, 4], що знання характеризуються рядом властивостей, що відрізняють їх від традиційних моделей даних. До таких властивостей можна віднести.

*Внутрішня інтерпретація.* При зберіганні знань у пам'яті ПС поряд з традиційними елементами даних зберігаються інформаційні структури, що дозволяють інтерпретувати зміст відповідних чарунок пам'яті.

*Структурованість.* Знання складаються з окремих інформаційних одиниць, між якими можна встановити класифікуючі відношення: рід – вид, клас – елемент, тип – підтип, частина – ціле і т. п.

*Зв'язність.* Між інформаційними одиницями передбачаються зв'язки різного типу: причина – наслідок, одночасно, бути поруч та ін. Дані зв'язки визначають семантику і прагматику предметної області.

*Семантична метрика.* На множині інформаційних одиниць, що зберігаються в пам'яті, вводяться деякі шкали, що дозволяють оцінити їх семантичну близькість. Це дозволяє знаходити в інформаційній базі знання, близькі до вже знайдених.

*Активність.* За допомогою даної властивості підкреслюється принципова відмінність знань від даних. Виконання тих чи інших дій в ПС ініціюється станом бази знань. При цьому передбачається, що поява нових фактів і зв'язків може активізувати систему.

Зазначимо, що ця властивість є фундаментальною для поняття «знання». Виходячи з нього введено поняття онтолого-керованої архітектури КС, так як поняття «онтологічна база знань» є видовим по відношенню до поняття «база знань».

Крім того, властивість актуальності знань може породжувати процес актуалізації, який має визначальне значення у ланцюжку переходів «сховище\_даних → затребувана\_інформація → знання». Однак для нас більш актуальною (з практичної точки зору) є проблема не первинності компонент ланцюжка, а їх приналежність. Відомі такі описи приналежності зазначених компонент, визнані багатьма дослідниками.

*Визначення 1.3 Дані* – це зафіксовані об’єкти або явища матеріального чи абстрактного світу.

*Визначення 1.4 Інформація* – це поняття категорії верхнього рівня, що характеризує об’єкти та явища матеріального і абстрактного світу, представлені у вигляді символів, якими оперує людина чи комп’ютер.

На будь-якому етапі перетворення вимога актуалізації відповідної компоненти є обов’язковою (але не достатньою) умовою переходу у вищу ступінь абстрактності. І якщо для першого переходу («сховище\_даних → затребувана\_інформація») механізми актуалізації досить добре вивчені і опрацьовані, то для другого переходу «затребувана\_інформація → знання») зазначені механізми слабо вивчені і реалізовані в розрізних експериментальних розробках.

Проблема здобуття знань з тексту представляється не тільки не тривіальною, але і вельми складною, незважаючи на безсумнівні досягнення Computer science в цій галузі. Як справедливо відмічено в [25], «Текст є знакова конструкція і часто містить знання. Але текст є не знання, а тільки його джерело. Знання з тексту ще треба отримати. Людині або КС».

Виконаний аналіз дозволяє синтезувати деяку узагальнену схему «еволюціонування» знань, представлену на рисунку 2.5. Зауважимо при цьому, що в мові даного звіту словоформи термінів «знання», «свідомість» (в російськомовному варіанті «сознание») і «пізнання» мають спільний корінь, а самі терміни також утворюють тріаду, компоненти якої фіксують об’єкт, суб’єкт і процес пізнання.

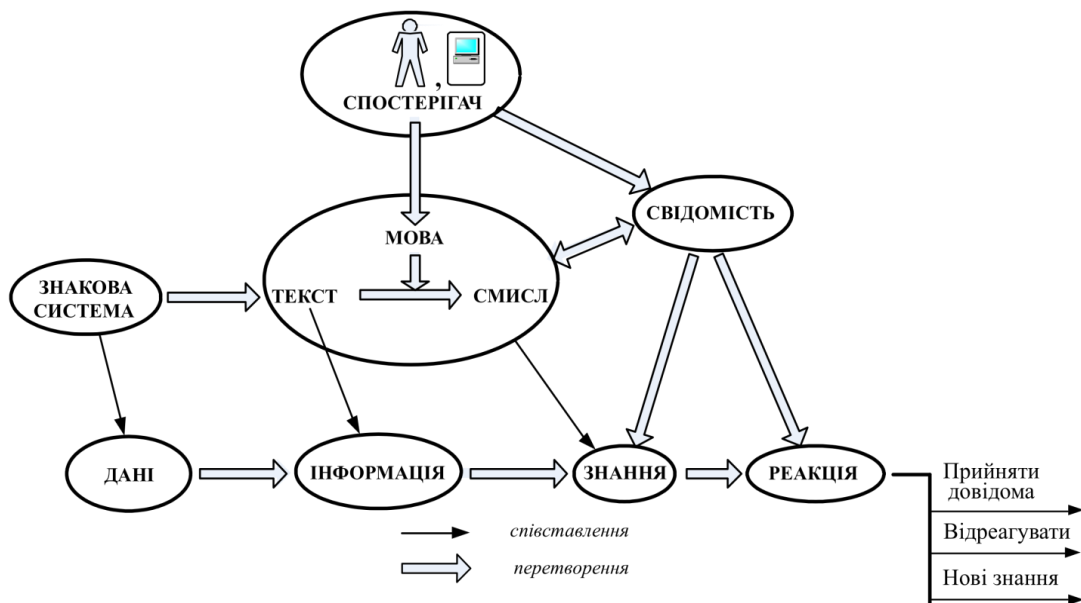


Рис. 2.5. Узагальнена схема «еволюціонування» знань

Входом є деяка знакова система, яку можна інтерпретувати як сукупність даних. Далі відбувається «осмислення» вхідних даних, для чого обов'язкова присутність *Спостерігача* (людина, комп'ютер), як носія вищих форм мислення – мовної свідомості. І якщо у нього є «внутрішній мовний інтерпретатор, налаштований на дану знакову систему», то дані перетворюються в інформацію. Далі в роботу включається "смісловий інтерпретатор" (якщо він є у *Спостерігача*), який перетворює інформацію у знання або відбувається усвідомлення інформації, яка надійшла. На заключному етапі обробки вхідного повідомлення відбувається природна реакція людини: взяти до відома, відреагувати або поповнити базу знань новою порцією знань [10].

#### *Джерела знань та схема комп'ютерної обробки знань.*

Виділяють [3–5, 11, 16, 22, 27–30, 37–42] два основних джерела знань – це експерти, фахівці у своїй предметній області і лінгвістичний корпус текстів (або множина ПМО). Якщо для першого джерела методи придбання знань досить добре вивчені і опрацьовані, а також відомі відповідні промислові експертні системи, то для другого – розроблені лише окремі методи, не пов'язані в єдину інтегровану технологію, а відповідні інформаційні системи носять експериментальний характер і не досконалі. Актуальне удосконалення відомих і розроблення нових, більш ефективних формальних підходів і моделей обробки, починаючи від пошуку релевантної інформації, її аналізу, усунення різного роду неоднозначностей, формально-логічного, онтолого-семантичного, інформаційно-кодового представлень, переходу до формалізованого опису знань, розробці процедур роботи із знаннями і відповідних їм алгоритмів, закінчуючи отриманням конкретних результатів користувачем. У більш загальному вигляді наведена вище послідовність етапів обробки знань, що містяться в природномовних об'єктах, представляється ланцюжком технологій Natural Language Processing (NLP) → Knowledge Representation (KR) → Knowledge Processing (KP).

Загальна схема комп'ютерної обробки знань, що містяться в ПМО, представлена на рисунку 2.6 [43]. Тут у загальному вигляді відображена інформаційна модель обробки знань, починаючи з пошуку на всьому інформаційному просторі відповідних текстових даних, які відповідають запиту користувача (можливо на ПМ) і подальшого їх перетворення (в загальному випадку) в прості, ситуаційні і нові (формальні) знання.

Під *простими знаннями* розуміється інформація, що надійшла про деякі сутності реального світу, якій відповідає реакція людини – «взяти до відома».

Під *ситуаційними знаннями* розуміється опис деякої ситуації, яку необхідно розпізнати і відповідним чином «відреагувати».

Під *новими знаннями* розуміється розпізнавання збільшення  $\Delta S$  і поповнення ними БЗ заданої ПдО в деякому формалізованому вигляді.

По суті, схема на рисунку 2.6 представляє собою концептуальну модель засобів, методів і технологій обробки знань, що містяться в ПМО.

Добування і обробка знань з ПМО є одним з розділів Data mining (Knowledge acquisition, Knowledge extraction, Knowledge processing) і визнано перспективним міждисциплінарним напрямком досліджень.

*Аксиоматизація наукових знань* визначається наступними структурними елементами: аксіоми або схеми аксіом (як положення, що в даній системі не доводяться); вихідні (початкові) поняття (явно не можуть бути визначені в системі); правила виводу та побудови системи знань; визначення як правила введення в теорію нових термінів, абстрактних об'єктів; теореми як висловлювання, що виводяться; завдання, наслідки, положення, що виводяться з теорії, леми, логіко-методологічні принципи теорії, що аксіоматизується. Всі вищезазначені аспекти складають базову основу формального наукового дослідження. Поряд з ними науковий підхід необхідно передбачає ще одну важливу компоненту – інтерпретацію. Зрозуміло, що науковий смисл мають лише ті абстракції, які завідомо можуть бути включені до чого-небудь і які можна зіставити істині, особливо в тих випадках, коли це практично важливо й можливо. Механізмом такого зіставлення істини, пошуку логічних можливостей її встановлення і виступає інтерпретація як логічний прийом щодо встановлення значення термінологічних виразів теорій, знакової (формальної) системи [41].

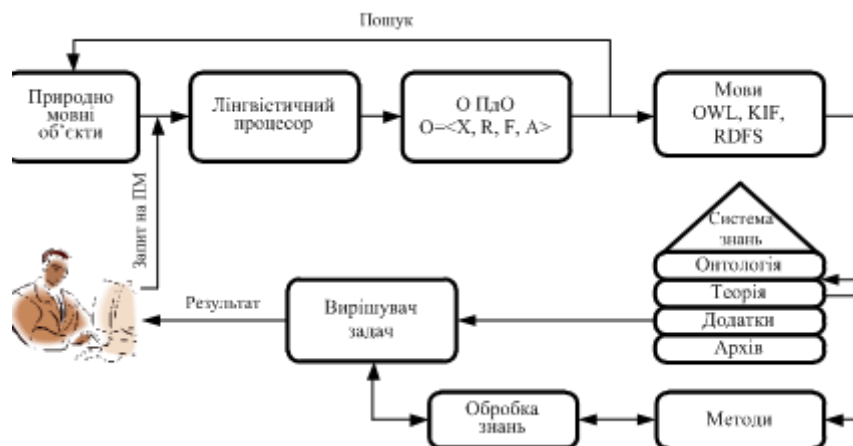


Рис.2.6. Комп'ютерна обробка знань, які містяться в ПМО

Кожному реченню, логічно правильно побудованому, в інтерпретації надається деяке значення, наприклад, речення може бути істинним або хибним. Питання ж про те, яка з цих можливостей реалізована, не вирішується інтерпретацією. Тому одна і та ж формальна система передбачає безліч інтерпретацій і тільки одну істину. Інтерпретація гарантує несуперечність теорії, але не забезпечує змістовної істинності даної теорії.

Інтерпретація має свої гносеологічні передумови. Суть їх полягає в тому, що в уявному процесі об'єктивна дійсність не просто пасивно

відбивається, але певним чином реконструюється, перебудовується. В пізнанні відбувається безперервний процес утворення абстракцій і їх виключення, переклад одного рівня абстракцій на інший. Інтерпретація як раз і виступає *логічним засобом переходу* від одного рівня знань до іншого, засобом встановлення логічних зв'язків між цими рівнями [41, 44].

Відомо досить багато моделей подання знань та відповідних їм методів обробки в КС [3, 5, 17, 25, 42, 45-52, 62-66]. Їх вибір для конкретного додатка істотно залежить від характеру знань ПдО, наявності засобів автоматизації побудови БЗ, обсягу останньої, а головне – набору реалізованих відповідною системою функцій.

Слід зазначити, що існуючі механізми вилучення нових знань оперують тільки знаннями силогістичної типу. Перехід до формування нових знань більш високого рівня абстракції методами ШІ, якщо взяти викладену в [23] ймовірнісну модель людського мислення, відповідає рівню передсвідомості *Спостерігача*. На цьому рівні знання організовані в складну онтологічну структуру з повним набором зв'язків між концептами. Зі сказаного випливає, що в даний час автоматична побудова БЗ для широких предметних областей представляє *складну наукову проблему*. При цьому ззовні (по відношенню до КС) фіксуються рівні розташування базових понять ПдО і базові відношення між ними. Таку структуру, створену вручну інженером по знаннях на основі тезауруса ПдО, можна представити як *початкову онтологію ПдО*.

### **2.3. Проблеми, методи та засоби обробки природномовних об'єктів**

*Аналіз проблеми.* Теорія і практика створення і використання систем, заснованих на знаннях (або знання-орієнтованих інформаційних систем), – найбільш актуальний та такий, що інтенсивно розвивається, напрямок Computer Science, використання результатів якого підвищить ефективність створення інструментальних засобів, прикладних систем і застосування комп'ютерів. На жаль, в даний час навіть на світовому рівні не отримано відчутних та загальноновизнаних результатів у широких прикладних областях знань, хоча можна відзначити наявні вдалі рішення в цьому напрямку для вузькоспеціалізованих програм. Складність зазначеної проблеми визначається, зокрема, складністю побудови, організації та використання великих баз формалізованих знань, а також залучення цілого ряду наукових теорій (логіки, комп'ютерної та психологічної лінгвістики, нейрон-кібернетики, теорії семантичних мереж тощо), які, цілком очевидно, повинні сприяти вирішенню проблеми вилучення, формального представлення, обробки і системної інтеграції знань і скласти концептуально-методологічну основу *теорії міждисциплінарних наукових досліджень*.

Як випливає зі сказаного, вихідною інформацією для більшості наукових досліджень є знакові системи, представлені у вигляді ПМО. В загальному випадку такими ПМО можуть бути великі бази неструктурованих

даних, що зберігаються в корпоративній пам'яті, різного роду електронні бібліотеки, колекції документів, текстова складова простого і семантичного Web, монографії, наукові статті, науково-технічні та ділові документи тощо. Таке уточнення звужує коло наукових дисциплін і підходів, методи яких залучаються до аналізу і розуміння ПМО, вилучення і формалізації з останніх знань, їх структурування та обробки. Серед інших можна виділити *лінгвістику*, яка досліджує, в тому числі лексичний і структурно-граматичний аспекти ПМ, і *логіку*, яка розглядає мову в одному обмеженому аспекті, лише в тій мірі, в якій вона є засобом *фіксації, переробки і передачі знань*. Розглядаючи мовний опис, логіка в першу чергу цікавиться відношенням його до елементів об'єктів, що позначаються і тим, як за допомогою певних зв'язків з цих елементів утворюються складні знакові системи, які виражають справжні знання про об'єктивний світ. Вирушаючи від аналізу ПМ, логіка розглядає в якості пріоритетних особливі мови – штучні мови науки, що виникають на базі природничих, але відрізняються рядом важливих рис. Для дослідження мов (природних або штучних) логіка використовує інструментарій, реалізований у вигляді формальних знакових систем, що дозволяє виявити закони побудови та функціонування, тобто утворення та перетворення систем знань. Питання про виділення об'єктивних засобів фіксації знань і виведенні одних «одиниць» знань з інших за певними, об'єктивно значущими правилами, є одним з пріоритетних у сучасних дослідженнях з логіки [10, 53-59].

Іншим важливим напрямком в рамках обраної галузі дослідження є створення ЗОІС, функціонування яких ґрунтується на автоматизації процесу формалізації змісту ПМТ з подальшою обробкою формалізованого подання цього змісту логіко-семантичними методами в інтересах вирішення конкретних завдань інформаційної або інформаційно-аналітичної діяльності фахівців деякої предметної області. Основною компонентою таких ІС є підсистема *лінгвістичної обробки ПМТ*, завданням якої є розпізнавання і добування знань, що містяться в ПМТ, і їх формалізація – з одного боку, і синтез по формалізованому представленню цих знань їх описів на природній мові – з іншого [47].

Таким чином, особливістю лінгвістичної обробки в таких системах є підпорядкованість всіх її етапів формування елементів формалізованого представлення знань. Однак її реалізація неможлива без моделювання в повному обсязі *процесу розуміння текстової інформації*. Складність створення таких систем обумовлена значною мірою відсутністю єдиної теорії розуміння ПМТ людиною і, як наслідок, неможливістю формалізації цього процесу стосовно комп'ютерних систем. Тому при розробці ПМ-систем часто накладаються різні «спонтанні» обмеження, наслідки яких не до кінця усвідомлюються розробниками. Ці обмеження приводять до того, що чимало людино-машинних систем, на розробку яких витрачаються величезні кошти, при промисловому використанні не задовольняють вимогам рішення кінцевих прикладних задач, пов'язаних з аналізом ПМ-інформації [16].

Однією з важливих гілок сучасного розвитку ІС є онтолого-керовані інформаційні системи. Побудова останніх тісно пов'язана з розробкою теоретичних основ і методології проектування, що включають формальний підхід, фундаментальні принципи і механізми, узагальнену архітектуру і структуру системи, формальну модель і методологію проектування онтології предметної області, формальну модель представлення знань, узагальнені алгоритми процедур обробки знань та ін. У свою чергу, кожна з перелічених складових загальної методології проектування являє собою складну інформаційно-алгоритмічну структуру. Наприклад, розробка онтології ПдО тісно пов'язана з концептуалізацією онтологічних категорій, розробкою та удосконаленням ієрархічних структур сутностей на всіх рівнях, побудовою формальної системи аксіом і обмежень. Комплексне вирішення зазначених завдань проектування має підвищити *роль онтологічних (концептуальних) знань* при вирішенні конкретних задач в прикладних областях.

При проектуванні знання-орієнтованих інформаційних систем (і зокрема ОКІС) істотним є вибір як формально-логічного представлення знань, так і джерел придбання і поповнення знань. В даний час визнаним «де-факто», найбільш великим і загальнодоступним джерелом знань є Інтернет-простір, з його природним способом представлення інформації. Зазначений спосіб визначив появу численних інформаційних технологій обробки знань, що містяться в ПМО, в тому числі й NLP-технології.

При створенні ОКІС слід виділити три аспекти досліджень – онтологічний, логічний і методологічний. Ці аспекти мають свої, в загальному випадку фіксовані об'єкти досліджень, відповідні процесу пізнання або розробці деякої ІС. В зв'язку з ними всі об'єкти можуть бути поділені на *три групи*: система сутностей (або об'єктів реального світу), система знань і система обробки сутностей у відповідності з цією системою знань. Перша група є предметом онтологічного дослідження, друга група – логічного дослідження і третя – методологічного. При цьому під *методологією* будемо розуміти сукупність прийомів, методів і механізмів, їх взаємодію, що застосовуються в процесі досліджень.

Велика частина досліджень відноситься до обробки ПМО або лінгвістичного корпусу текстів (ЛКТ) із заданої предметної області. Загальноприйнятого визначення для поняття «ПМО» невідомо, а для поняття «Лінгвістичний корпус текстів» змістовний огляд загальноприйнятих визначень виконаний в [48, 64]. Лінгвістичним корпусом текстів є будь-яка колекція текстів з деякої тематики (з деякої предметно-проблемної області), які є доступними в електронній формі, а також результати їх спільного аналізу. Там же наведено методіку формування довільного ЛКТ.

Слід уточнити смисли вживаних у звіті деяких термінів, які не завжди збігаються з їх вживанням в інших публікаціях. Ці терміни відносяться до вхідної інформації для ЗОІС і згруповані в категорію «Природномовна знакова система». Нижче її в онтологічній ієрархії знаходяться два терміни-синоніми – множина природномовних об'єктів і лінгвістичний корпус



текстів. Нижче цих термінів також знаходяться два терміна-синоніма – природномовний текст і природномовна інформація. І на самому нижньому рівні знаходяться два терміна-синоніма – природномовне висловлювання і природномовне речення. Останні терміни є синонімами, оскільки в звіті розглядаються тільки стилі текстів «Науково-технічний» і «Ділова проза». З цього випливає, що всі речення таких текстів є розповідними і можуть розглядатися як висловлювання.

## 2.4. Формалізовані підходи добування та представлення знань

Проблема отримання знань з ПМО, їх представлення та обробка відносяться до області ШІ. Цією надзвичайно важливою проблемою займалися багато видатних учених (Аристотель, Лейбниц, Тарський та ін.). У загальній проблемі аналізу та розуміння ПМО, у всьому різноманітті робіт [62–66], методів, методологій та термінології явно проглядаються наступні завдання:

- синтактико-семантичний аналіз ПМО;
- аналіз ПМО з метою вибору підходящого формально-логічного представлення в деякій підсистемі логіки та подання в пам'яті комп'ютера і подальшої ефективної обробки;
- розробка методів добування знань з ПМО.

Розглянемо ці завдання більш детально.

Проблема аналізу ПМО природним чином пов'язана з операційною і денотаційною семантикою мови. У системах штучного інтелекту і, зокрема, в системах аналізу ПМО переважає операційна семантика, оскільки очевидно переважання інтересу до засобів представлення знань (як?), а не до їх змісту (що?), яке підлягає формалізації [67].

Системи (мови) подання знань визначають конструкції, які підтримуються комп'ютером і забезпечують ефективність організації процедур, і як наслідок – максимальну уніфікацію мовних засобів.

На інформаційно-логічному рівні переслідується зовсім інша мета – забезпечення відображення змісту об'єктів реального світу в мовних конструкціях, що узгоджуються зі сприйняттям, властивим людині. Для користувача такої системи (схеми) комп'ютерне подання незручне, оскільки він спілкується (або, принаймні, повинен) з системою на природній мові.

Проте, для системного аналітика, що використовує мову формалізації для побудови моделі предметної області і створює для цього орієнтовані на цю ПДО словники понятійних одиниць і на їх основі – інтелектуальні системи, наповнені конкретними знаннями, комп'ютерне представлення грає ключову роль. Аналітику доводиться мати справу з конкретною множиною термінів та ПМ-конструкцій, саме їм давати тлумачення в термінах мови формалізації, встановлювати між ними інформаційно значимі зв'язки. Тому, в разі розбіжності мови формалізації зі складом мислення фахівця, реалізація системи обробки даних у конкретному смислового матеріалі може стати занадто складною або взагалі нерозв'язною проблемою.

Найбільш прийнятним на сьогоднішній день є мови логічного типу, в яких поняття «зв'язок за змістом» формалізується з максимальною повнотою. Одним з головних питань при цьому є питання про рівень абстрактності зафіксованих категорій логічної мови. Якщо нас цікавить лише загальна схема логічного слідування, то досить розглядати лише числення предикатів першого порядку з правилом резолюцій в якості правила виводу, не конкретизуючи нічого, крім списку аксіом і логічних констант. Однак при більш багатих змістом ПдО повинно бути вирішене питання про типи термових констант і вигляді нелогічних аксіом (власних аксіом), що використовуються при формалізації знань про емпіричні об'єкти.

Викладене свідчить про актуальність проблеми розробки мови логічного типу для представлення знань, що містяться в тексті. Вочевидь, що ця мова має бути розширюваною у бік здатності більш повного опису об'єктів, здатності робити передбачення (гіпотези), які враховують контекст [67].

## 2.5. Формальна постановка задачі добування знань з ПМТ

Перш ніж перейти до розгляду системи обробки і добування знань, що містяться в природномовних текстах (ПМТ), визначимо поняття «добування знань з ПМТ». З цією метою скористаємося поняттями, які використовуються в constraint programming [82] як однієї з форм декларативного програмування.

Нехай дана множина  $D$ , на якій визначена деяка кінцева сукупність  $n$ -арних відношень  $R$  на  $D$ , тобто  $R_i \subseteq D^n$ , де,  $R_i \in R \subseteq D$ ,  $i = 1, \dots, k$ . Мовою обмежень  $L$  на  $D$  називається деяка не пуста множина  $L \subseteq R \subseteq D$ . Проблема виконання обмежень формулюється таким чином.

Для довільної множини  $D$  і довільною мови обмежень  $L$  на  $D$  проблемою виконання обмежень CSP ( $L$ ) є рішення такої комбінаторної проблеми:

Дано: трійка  $P = (V, D, C)$ , де

–  $V$  – кінцева множина змінних;

–  $C$  – деяка множина обмежень  $\{C_1, \dots, C_q\}$ ;

– кожне обмеження  $C_i \in C$  – це пара  $(s_i, R_i)$ , де

–  $S_i$  –  $n$ -ка елементів з  $V$ , яка називається областю обмеження,

–  $R_i \in L$  –  $n$ -арне обмеження на  $D$ , яка називається відношенням обмеження.

Знайти: функцію  $\varphi: V \rightarrow D$  таку, що,  $\forall (s, R) \in C$  де  $s = (v_1, v_2, \dots, v_n)$ ,  $n$ -ка  $(\varphi(v_1), \varphi(v_2), \dots, \varphi(v_n)) \in R$  або переконатися в тому, що такої функції не існує.

Множина  $D$  в цьому випадку називається областю проблеми, а множина всіх рішень CSP виду  $P = (V, D, C)$  позначається  $Sol(P)$ .

Що стосується аналізу ПМТ з метою добування знань множина  $D$ , як область проблеми, інтерпретується як множина об'єктів, добутих з вхідного тексту  $T$ , яке факторизовано за деякими відношеннями еквівалентності  $R$  (це відношення названо відношенням синонімії) і в якому «закодовані»

відношення  $R_i, i=1,2,\dots,k$ . Змінні з множини змінних  $R_i, i=1,2,\dots,k$  приймають свої значення в цій факторизованій множині об'єктів, що фігурують в тексті  $T$  (це можуть бути лексико-граматичні розряди, конкретні об'єкти (люди, дати, предмети і т.д.)).

Проблемою вилучення знань з ПМТ називається проблема пошуку інтерпретації  $\varphi:V\rightarrow D$  з явною побудовою відношень  $R_i$  з множини  $L \subseteq R$ . При цьому відношення  $R_i \in L, i=1,2,\dots,k$  вилучені з тексту  $T$ , будемо називати елементарними знаннями.

При аналізі ПМТ першочерговим завданням є побудова двох фундаментальних відношень, присутніх практично у кожному ПМТ. Це відношення еквівалентності і часткового порядку, які названі загально-значущими. Перше відношення визначає класи синонімічних об'єктів, а друге відношення – ієрархію підпорядкованості класів еквівалентності. Обидва ці відношення складають основу побудови онтологій. Знання, отримані на цьому етапі, будемо називати первинними. У відношення часткового порядку може вкладатися різний семантичний зміст: це може бути відношення таксономії («належати» множині, класу, групі і т. д.), відношення партономії («складатися з»), відношення генеалогії («батько-син»), причинно-наслідкові відносини («якщо-то»), атрибутом відношення і т. д.

Як зазначено в [66], найбільш перспективними представляються ЗОІС, функціонування яких спирається на автоматизацію процесу вилучення та формалізації змісту природно мовних текстів з подальшою обробкою формалізованого подання цього змісту логіко-семантичними методами з орієнтацією на конкретну ПдО.

Вищенаведений процес вилучення знань з ПМТ представлений в достатньо загальному вигляді і потребує уточнення. Виконаємо деяку конкретизацію цього визначення стосовно до предметної області "Аналіз ПМТ". Розглянемо деякі приклади.

## **2.6. Комп'ютерні онтології ПдО**

### **Системно-онтологічний аналіз ПдО**

Під системним аналізом зазвичай розуміється вид дослідження, при якому реальний чи мислимий об'єкт розчленовується на складові частини (елементи) і потім досліджуються ці елементи і зв'язки між ними. Аналіз предметної області становить особливий вид наукової діяльності, в результаті якої будується інтерпретаційна модель предметних знань (в широкому сенсі) [87]. В процесі аналізу останні діляться на інваріантні та прагматичні знання, концептуальні складові яких представляють онтологічні знання ПдО.

Деякі ідеї з розробки методології проектування онтології ПдО беруть свій початок в літературі з об'єктно-орієнтованого підходу (ООП), що виник як технологія програмування великих програмних продуктів [88]. Проте розробка онтологій, як ієрархічної структури понять (концептів)

відрізняється від проектування об'єктів, як класів і відношень в об'єктно-орієнтованому програмуванні. Останній зосереджується головним чином на методах опису класів – програміст приймає проектні рішення, засновані на операторних властивостях класу, тоді як розробник онтології приймає ці рішення, ґрунтуючись на структурних властивостях класу. В результаті структура понять і відношення між поняттями в онтології відрізняються від структури класів об'єктів подібної ПдО в об'єктно-орієнтованій програмі [89]. Крім того, при розробці онтології внутрішній зміст поняття есплікується завжди, в той час як в об'єктно-орієнтованому програмуванні найчастіше застосовується метод інкапсуляції як спосіб обмеження доступу до внутрішнього змісту об'єкта.

Системний підхід до пізнання орієнтує аналітика на розгляд будь-якої ПдО з позицій закономірностей системного цілого та взаємодії складових його частин. Системність знань виходить з багаторівневої ієрархічної організації будь-якої сутності, тобто всі об'єкти, процеси та явища можна розглядати як множину дрібніших підмножин (ознак, деталей) і, навпаки, будь-які об'єкти можна (і треба) розглядати як елементи більш високих класів узагальнень.

90-і роки минулого століття вважаються періодом зародження парадигми комп'ютерних онтологій. Вона була сформульована як спроба згладити (і по можливості усунути) різного роду суперечності при функціонуванні та впровадженні інтелектуальних систем з використанням баз знань предметних областей. Яскравим представником таких систем на той час були експертні системи (ЕС). У розробників постійно виникало питання: «Як забезпечити їх просування до кінцевого користувача?». Були запропоновані розробки оболонок – «пустих» ЕС і ряд інших нововведень. Але вони не мали вирішального значення. Напрошувався висновок, що для ефективного функціонування ЕС на найважливішому етапі «життєвого циклу» – функціонування у кінцевого користувача при вирішенні реальних завдань – необхідно до кожної ЕС «прикласти» експерта у відповідній ПдО. Необхідність присутності експерта пояснювалася, зокрема, швидкоплинністю зміни знань у багатьох предметних областях і відповідно необхідністю оновлення бази знань ЕС в «реальному часі». У тих же ПдО, де знання мали відносну інваріантність, ЕС продовжували ефективно функціонувати.

Сказане вище і ряд інших факторів «підштовхнули» до розробки парадигми комп'ютерних онтологій, основні принципи якої були сформульовані в [7].

1. *Дохідливість, ясність* (Clarity). Терміни (і поняття) онтології повинні відображати реальну дійсність. Їх символічні позначення (знаки) повинні формуватися на основі загальноприйнятих правил в семіотиці і повинні висловлювати загальноприйняті смисли реальних об'єктів. В свою чергу, ці смисли вилучаються із загальноприйнятих визначень термінів (понять), зафіксованих в тлумачних словниках, різних глосаріях ПдО.

Судження, що входять до визначення, формалізуються на основі формального загальноприйнятого апарату у вигляді істинних логічних аксіом.

2. **Обґрунтованість, зв'язність** (Coherency). Формування початкового набору понять онтології і їх додаток має бути обґрунтованим, визначеним, в першу чергу, вимогами передбачуваної сукупності розв'язуваних завдань. Логічні аксіоми початкового набору понять повинні бути несуперечливими. Для цього повинен бути передбачений механізм логічного виведення, який, в тому числі, перевіряє на несуперечність аксіоми які додаються і виведені в онтології твердження.

3. **Розширюваність** (Extendibility). Ядром онтології є спочатку введені (спроєктовані) поняття і аксіоми, що їх описують. В онтології повинен бути передбачений механізм розширення (обмеження) спільно використовуваних словників понять без порушення цілісності системи.

4. **Мінімальний вплив кодування** (Minimal encoding bias). В онтологічній системі (ОнС) повинен бути реалізований принцип спільного використання онтологій, який передбачає: специфікацію онтології на рівні повноцінного представлення, а не символічного кодування; запис такої специфікації на загальноприйнятій і платформонезалежній мові опису онтологій можна передати для використання будь-якому програмному агенту.

5. **Мінімальні онтологічні зобов'язання** (Minimal ontological commitment). Цей принцип перекликається з принципами обґрунтованості та розширюваності/обмеження. Важливо, щоб множина понять онтології відображала концептуальну структуру ПдО, відносно стабільну протягом «життєвого циклу» ОнС. А остання надавала б можливість розширення або спеціалізації окремих гілок онтологічного графа. Відділення *концептуальних знань* від знань, виражених фактами, є стратегією побудови ОнС, а точніше – онтологічних баз знань.

Зупинимося на розгляді комп'ютерної (формальної) онтології предметної області, онтології верхнього рівня і онтології домену предметних областей. Остання (у тому числі) має важливе значення для об'єднання (інтеграції) концептуальних знань близьких предметних областей або реалізації технології системної інтеграції трансдисциплінарних наукових знань. Ми також виділяємо *початкову онтологію* ПдО, що є ініціалізуючою домінантою при реалізації технології автоматизованої побудови онтології ПдО.

Нижче наведені відомі визначення поняття онтології ПдО, починаючи з початкового визначення Т. Грубера, сформульованого в [7], з наступним його уточненням: «*Онтологія* – це формальна специфікація узгодженої концептуалізації», [8]. Іншим важливим визначенням онтології ПдО є визначення Н. Гуаріно [90]: «*Онтологія* – це формальна теорія, що обмежує можливі концептуалізації світу». На основі цих визначень різні дослідники формулюють свої, окремі визначення онтології ПдО – відповідно до їх конкретної області професійних інтересів. Тематика і практична

спрямованість наших досліджень (онтологія як засіб побудови баз знань трансдисциплінарних наукових досліджень) зумовлює таке визначення комп'ютерної онтології ПдО.

*Визначення. Комп'ютерна онтологія ПдО* – це:

- 1) ієрархічна структура скінченної множини понять, що описують задану предметну область;
- 2) структура є онтограф, вершинами якого є поняття, а дугами – семантичні відношення між ними;
- 3) поняття і відношення інтерпретуються відповідно до загальнозначущих функцій інтерпретації, взятих з електронних джерел знань заданої ПдО;
- 4) визначення понять і відношень виконується на основі аксіом і обмежень їх області дії;
- 5) формально онтограф описується на одній з мов опису онтологій;
- 6) функції інтерпретації та аксіоми описані в деякій підходящій формальній теорії.

В загальному випадку онтологію деякої ПдО формально представляють впорядкованою трійкою [3, 5, 10, 90]:

$$O = \langle X, R, F \rangle, \quad (2.1)$$

де  $X, R, F$  – кінцеві множини відповідно:  $X$  – концептів (понять, термінів) предметної області,  $R$  – відношень між ними,  $F$  – функцій інтерпретації  $X$  та/або  $R$ .

Розгляд граничних випадків множин (1.3):  $R = \emptyset$ ;  $R \neq \emptyset$ ;  $F = \emptyset$ ;  $F \neq \emptyset$  у всіх чотирьох комбінаціях значень  $R$  і  $F$  дає різні варіанти онтологічних конструкцій, починаючи від простого словника і таксономії до формальної структури концептуальної бази знань для високоінтелектуальних знання-орієнтованих систем. Детальний розгляд різних зазначених комбінацій з їх змістовною інтерпретацією виконано в [91].

За своєю функціональною повнотою і ступеню формальності розрізняють три види онтологій: проста, повна (чи строга) і множина проміжних або неповних онтологій.<sup>3</sup>

Проста онтологія – це така онтологія, в якій  $R = \emptyset$ ;  $F = \emptyset$ . Вона служить (в основному) для однозначного сприйняття науковою спільнотою понять у відповідній прикладній області.

*Строга* або *повна онтологія* ( $R \neq \emptyset$ ;  $F \neq \emptyset$ ) – це така онтологія, в якій множини концептів та концептуальних відношень максимально повні, а до функцій інтерпретації додаються аксіоми, визначення та обмеження. При цьому опис всіх компонент представлений на деякій формальній мові, яка доступна для їх інтерпретації комп'ютером. Схема формальної моделі повної онтології описується четвіркою:

---

<sup>3</sup>Класифікації онтологій за різними підставами наведені в численній літературі [4, 50, 91 та ін.].

$$O = \langle X, R, F, A(D, R_s) \rangle, \quad (2.2)$$

де  $X$  – множина концептів;

$R$  – множина концептуальних відношень між ними;

$F: X \times R$  – скінченна множина функцій інтерпретації, заданих на концептах і/або відношеннях;

$A$  – скінченна множина аксіом, які використовуються для запису завжди істинних висловлювань (визначень і обмежень);

$D$  – множина додаткових визначень понять;

$R_s$  – множина обмежень, що визначають область дії понятійних структур.

Повна (комп'ютерна) онтологія є (формальним) вираження концептуальних знань про предметну область і за своєю значимістю порівнюється з базою знань ЗОІС, а її побудова є специфічною формою творчості. Творчий процес при цьому можна представити сукупністю операцій-процедур з судженнями, твердженнями, поняттями і відношеннями між ними, а його результат – основою для побудови складової частини наукової теорії – онтологічної бази знань у заданій предметній області, описаній в декларативній формі.

Множина *проміжних* або *неповних онтологій* ( $R = \emptyset, F \neq \emptyset; R \neq \emptyset, F = \emptyset$ ) виникає, коли для кожного концепту (або їх більшої частини) додані аксіоми і визначення, представлені на ПМ.

Одним з поширених варіантів неповної онтології є структура виду  $O = \langle X, R \rangle$ , де множина  $F$  в явному вигляді відсутня ( $F = \emptyset$ ), в припущенні, що концепти  $x_i \in X$  загальновідомі (визначені за умовчужанням) або досить повно інтерпретовані відношеннями  $R$ .

Така схема класифікації за функціональною ознакою узгоджується з описом [90]: «Онтологія або концептуальна модель предметної області складається з ієрархії понять предметної області, зв'язків між ними і законів, які діють в рамках цієї моделі».

Відомо, що засоби інформатики, що проектуються відповідно до їх проблемної орієнтації, базуються на певній сукупності фундаментальних принципів, методик та алгоритмів. Парадигма комп'ютерних онтологій, що розвивається у взаємодії з методами і засобами системного аналізу, поклала початок розвитку нової гілки методів системного аналізу ПдО – системно-онтологічного аналізу (підходу) [2].

**Центральною ідеєю системно-онтологічного підходу** (СОП) є розробка онтологічних засобів підтримки рішення прикладних задач – *поліфункціональної онтологічної системи*. Така система (точніше, її концептуальна частина) описується двійкою (3.3), що включає онтологію ПдО (складається з онтології об'єктів і онтології процесів) і онтологію задач [92, 93].

$$O_{HC} = \langle O^{ПоО}(O^O, O^П), O^З \rangle \quad (2.3)$$

На рисунку 2.7 представлена схема онтологій-компонентів предметної області і *проблемного простору* (ПрП). ПрП – це модель всіх таких аспектів або компонент ПдО, з якими пов’язані (опосередковано або безпосередньо) знання, необхідні для вирішення різних задач у цій ПдО.

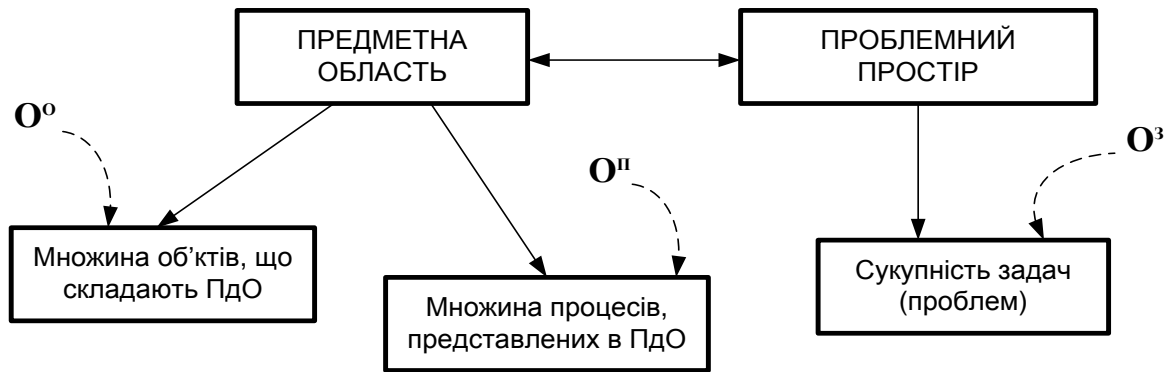


Рис. 2.7. Схема онтологій-компонентів предметної області

Будь-який ПрП складається з двох блоків: інваріантної (відносно незмінної) частини і множини змінних частин, відповідних окремим задачам. У складі інваріантної частини, наприклад в методології SMEE (Structured Methodology for Elicitation of Expertise), виділяють сім типів компонент: об'єкти, інструменти, оператори, операції, кінцеві продукти, побічні продукти і обмеження [98]. Типи компонент – суть поняття, які добре групуються в онтології об'єктів і процесів, представлених на рисунку 2.7:

- $O^o$  – онтологія множини об'єктів (понять, концептів) ПдО, яка розглядається як ієрархічна структура класів, підкласів та елементів класів;
- $O^п$  – онтологія множини процесів ПдО, яка розглядається як ієрархічна структура процесів, підпроцесів, дій та операцій;
- $O^з$  – онтологія сукупності задач (типових наборів), які можуть бути поставлені і вирішені в ПдО. Розглядається як ієрархічна структура задач, підзадач, процедур і операторів.

## 2.7. Лінгвістичні онтології

Застосування онтологічного підходу для автоматичної обробки текстів на природній мові передбачає порівняння понять онтології предметної області (до якої належить множина текстів) мовних виразів (слів і словосполучень), якими поняття можуть бути виражені в текстах. Процедура порівняння понять і мовних виразів може бути реалізована різними способами [91].

1. До понять вже розробленої онтології ПдО приписуються мовні одиниці. За складністю реалізації цей спосіб є самим простим, але йому властивий ряд очевидних недоліків:



1) імена понять досить громіздкі, неприродні, з ними важко оперувати як розробникам, так і можливим користувачам;

2) надмірна багатозначність лексичної одиниці, що виникає через те, що одне і те ж слово може відповідати занадто великій кількості понять залежно від контексту;

3) онтологічна гіпотеза Сепіра-Уорфа (те, що не описується словами, не може бути відображено в онтології).

2. Встановлення відповідності між лексичними ресурсами (наприклад, WordNet) і деякою (концептуальною) онтологією. WordNet-ресурси описують лексичні відношення між значеннями слів, представлені у вигляді окремих одиниць в ієрархічній мережі – синсетів. Відношення між лексичними одиницями в значній мірі відображають відношення об'єктів зовнішнього світу, тому такі ресурси часто розглядаються як особливий вид онтологій – лексичні або *лінгвістичні онтології*.

Головною характеристикою лінгвістичних онтологій є те, що вони пов'язані зі значеннями мовних виразів (слів, іменних груп тощо). Лінгвістичні онтології охоплюють більшість слів мови (в ідеалі – повний набір слів ПМ, структурованих у мовну картину світу) і одночасно мають онтологічну структуру, яка проявляється у відношеннях між поняттями. Тому лінгвістичні онтології можуть розглядатися як особливий вид лексичної бази даних і особливий тип онтологій.

3. Розробка єдиного ресурсу (змішаної онтології), в якому були б збалансовані обидві частини: система понять і система лексичних значень, – що досягається розумним розподілом цих одиниць в створюваному ресурсі і акуратним описом їх взаємозв'язків. Такий підхід реалізується, наприклад, в онтології Mikrokosmos.

## **Порівняльний аналіз онтологій**

Останнім часом в світі відомі декілька широко використовуваних лінгвістичних, концептуальних і змішаних онтологій, на основі яких розробляються інформаційні системи з обробкою знань (представлених, в тому числі, природномовними текстами) в різних предметних областях. Нижче наведені загальні відомості про лінгвістичної онтології WordNet (описи онтологій SUMO і Mikrokosmos представлені в Додатку Е).

**WordNet.** Лінгвістична онтологія для лексики англійської мови розроблялася в середині 1980-х років в лабораторії когнітології Принстонського університету (штат Нью Джерсі, США) під керівництвом проф. Дж.А. Міллера. Вільно поширюється в мережі Інтернет і являє собою лексикографічну базу даних, структуровану відповідно до лексико-семантичних відношень: синонімії, антонімії, гіпонімії, меронімії та морфологічних відношень [100, 101].

Найважливішим відношенням WordNet є подібність значень. В основу поняття «синонімія» покладене таке: два слова є синонімічними в контексті *C*, якщо заміна одного на інше в *C* не змінює його істинності.

Отже, синонімічні ряди можуть утворювати лише слова однієї частини мови, оскільки синоніми повинні бути взаємозамінними (слова різних синтаксичних категорій не можуть замінювати одне іншого).

Антонімія – це лексичне відношення протилежності між формами слів (слововживання), а не семантичне відношення між значеннями слів, тобто розрізняються семантичні відношення між слововживаннями та семантичні відношення між значеннями слів. Антонімія в WordNet є центральним організуючим принципом для прикметників і прислівників.

На відміну від синонімії та антонімії, які є лексичними відношеннями між формами слів, гіпонімія і гіперонімія є семантичними відношеннями між значеннями. Оскільки, зазвичай, поняття має тільки один гіперонім, воно породжує ієрархічну семантичну структуру, в якій вважається, що гіпонім належить гіпероніму. Гіпонім вбирає в себе всі риси загального поняття і додає риси, які відрізняють його від гіперонімів і від інших гіпонімів цього ж гіпероніму (тобто дотримується принцип успадкування ознак). Це центральний принцип організації іменників у системі WordNet.

Іншим типом відношень є «частина-ціле», яке в лінгвістиці відоме як меронімія/голонімія. Поняття, представлене набором  $\{x, x', \dots\}$ , є меронімом поняття, представленого набором  $\{y, y', \dots\}$ , якщо носій мови сприймає речення, сконструйоване з такої основи, як:  $Y$  має  $X$  (як частина) або ж  $X$  є частиною  $Y$ . Відношення меронімії може використовуватися для побудови ієрархії частин (з деякими обмеженнями, оскільки меронім може мати кілька голонімів).

Морфологічні відношення між формами слів є важливою частиною лексичних відношень. В системі WordNet спочатку інтерес обмежувався лише семантичними відношеннями: не було проектів, які б враховували морфологічні відношення. Проте з часом програми удосконалювалися і стало очевидним, що WordNet повинен служити будь-яким запитам, в тому числі пов'язаних з морфологічними характеристиками [102].

База даних WordNet розділена на п'ять лексико-граматичних класів: іменник, дієслово, прикметник, прислівник і функціональні слова. Частини мови організовані в ієрархію синонімічних множин (вузлів), названих *синсетами*. Кожен синсет представляє одне базове лексичне поняття і складається з множини слів і стійких словосполучень, рівнозначних у певному контексті. Синсети пов'язані між собою відношеннями різних типів.

Математичною моделлю WordNet служить граф  $V = (X, R)$ , множина вершин в якому розбита на дві непересічні підмножини:  $X = X_1 \cup X_2$ . Вершини з  $X_1$  відповідають словам і словосполученням, вершини з  $X_2$  – їх значенням (смыслам, тлумаченням). Кожне зі значень співвідноситься з однією з частин мови: іменником, дієсловом, прикметником або прислівником. Множина ребер також розбита на дві непересічні підмножини:  $R = R_1 \cup R_2$ . Ребра з  $R_1$  пов'язують слова зі значеннями, тобто елементи із  $X_1$  з елементами із  $X_2$ . Ці ребра представляють відношення, що входять у добуток  $X_1 \times X_2$ . Ребра  $R_2$ , які належать другій підмножині, пов'язують слова зі

словами і значення із значеннями, тобто представляють відношення, що входять у добутки  $X_1 \times X_1$  і  $X_2 \times X_2$  [4]. Об'єднання слів і словосполучень в синсети виражає відношення синонімії. Інші відношення (з перерахованих вище) задають типи ребер з  $R_2$ . У WordNet виділено 14 базових типів таких відношень (крім них використовуються зворотні відношення для кожного з наступних типів): антонім; «має відношення до»; дієслово, на основі якого утворено дієприкметник; слово, на основі якого утворено прислівник; дію, яка супроводжує дану дію; дієслівна група; атрибут; «дивись також»; подібність; рід-вид, вид-рід; ціле-частина, частина-ціле; «зроблений з», «служить субстанцією для»; множина-елемент, елемент-множина; мета-спосіб, спосіб-мета.

Технічно, WordNet є електронним тезаурусом, який визначає широкий клас значень слів, пов'язаних між собою семантичними показниками. Логічна структура WordNet показана на рисунку 2.8.

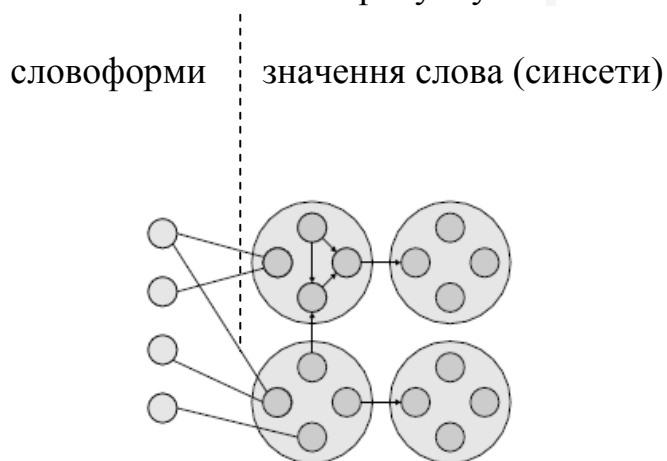


Рис. 2.8. Логічна структура WordNet

Розглянемо приклад запису синсету в базі даних.

**00047131 04 n 02 accession 0 addition 0 001 @  
09536731 n 0000 | something added to what you  
have already; «the librarian shelved the new accessions»;  
«he was a new addition to the staff»**

Перша частина запису стверджує, що число 00047131 є унікальним ідентифікатором синсету іменника {accession, addition}. Частина запису між символами «@» і «|» вказує, що цей синсет безпосередньо підпорядкований синсету з ідентифікатором 09536731. Останній синсет відповідає значенню «acquisition». І остання частина запису (після символу «|») являє собою тлумачення синсету і деякі приклади вживання слів, що входять в синсет.

Онтологія WordNet дуже популярна в наукових дослідженнях з комп'ютерної обробки природної мови та пошуку інформації в мережі (Information Retrieval). Вона постійно оновлюється. В даний час для Windows-користувачів доступна версія 2.1, для UNIX-подібних ОС – версія

3.0 і в ній існує понад 82 тис. синсетів іменників, більше 18 тис. синсетів прикметників, більше 3600 синсетів прислівників і близько 14 тис. синсетів дієслів. Всього вона включає більше 147 тис. англійських слів і фраз). Інші аспекти опису WordNet і EuroWordNet та велика кількість прикладів їх використання наведена в [91, 100, 102].

Відомі й інші (змішані та концептуальні) онтології верхнього рівня Sensus [103], Omega [104], Cys [105], DOLCE [106], PropBank [107, 108], FrameNet [109–111]. Деякі дані, що характеризують згадані онтології, наведені в таблиці 2.1. Аналіз публікацій свідчить про те, що найбільший інтерес (крім лінгвістичної онтології WordNet) для розробників представляють онтології Mikrokosmos, SUMO і Cys (онтографи перших двох представлені в розширеному варіанті в Додатку Е).

Верхній рівень онтології Cys розроблявся відповідно до загальної концепції онтології – «представити всі знання про світ». Він не має явного розмежування на загальноприйнятій групі категорій «Матеріальне і Абстрактне» (M&A), «Тривале і Подійне» (C&O) і «Незалежне, Відносне і Проміжне» (I&R&O) і має свою специфічну концептуальну структуру. Крім того, найбільш розроблена частина проекту Cys є комерційною і недоступною для загального користування. Існує і доступна частина проекту онтології (OpenCys). В якості позитивних моментів Cys можна відзначити наявність формальної мови опису CysL, бази знань і широку аксіоматизацію, що постійно розвиваються та доступні (в режимі клієнт-сервер) [105].

Онтологія Mikrokosmos має досить гнучку ієрархічну структуру з явним пріоритетом групи категорій I&R&O. Дві інші групи категорій розподілені по гілкам ієрархічного дерева і не мають чітко виражених обрисів. Онтологія має практичну спрямованість, в якій не всі гілки дерева розроблені однаково. Тому, вона може бути рекомендована для адаптації в тих предметних областях, поняття і терміни в яких найбільш широко представлені в категоріях і концептах Mikrokosmos. Інформація про онтологію доступна як на веб-порталі розробників, так і в численних публікаціях [113, 114, 119-121, 124-131].

Численні фрагменти онтології SUMO є варіантами адаптації різних ОВР в одну онтологію верхнього рівня. В ній явно виражений пріоритет групи категорій M&A з подальшим переважанням групи категорій C&O. Слід підкреслити (що також зазначено розробниками SUMO), що гілка онтологічного дерева для категорії *Матеріальне* розроблена істотно глибше, ніж гілка для категорії *Абстрактне*. До переваг SUMO можна віднести можливість трансляції опису онтології на будь-яку з основних мов подання знань, наявність онтології середнього рівня, інтегрованої з верхнім рівнем SUMO, розроблені приклади практичного застосування, а також зв'язок з WordNet.

Тому, дану онтологію можна рекомендувати для адаптації з онтологіями предметних областей, де явно переважають матеріальні концепти.

Таблиця 2.1

## Основні характеристики онтологій

№ п./п.	Назва онтології	Тип онтології	Розробник	Браузер	Кількість концептів (лексем)	Формальна аксіоматизація	Інтеграція з іншими онтологіями
1	WordNet	Лінгвістична	Princeton University	WordNet	близько 100 тис.	-	SUMO, FrameNet
2	Sensus	Змішана	ISI USC	Ontosaurus	понад 70 тис.	-	-
3	Omega	Лінгвістична	ISI USC	Omega	понад 120 тис.	-	WordNet, Mikrokosmos
4	Mikrokosmos	Лінгвістична	CLR UNMS	Mikrokosmos	понад 7 тис.	+	WordNet, Omega
5	OpenCyc	Концептуальна	Cycorp	Opencyc	понад 100 тис.	+	WordNet
6	DOLCE	Концептуальна	LAO ICST	DOLCE	понад 4 тис.	+	-
7	PropBank	змішана	University PennState	-	понад 4300	-	FrameNet VerbNet
8	FrameNet	змішана	ISI, Berkeley, CA	Frame Grapher	близько 900 фреймів	+	WordNet PropBank SUMO
9	SUMO	Концептуальна	Teknowledge Corporation, SUO WG	SUMO	понад 1000	+	FrameNet, WordNet, EMELD

*Мовно-онтологічна картина світу*

Онтологія МКС забезпечує фіксацію змісту вхідної інформації, причому інформаційні моделі, які ви користуєтеся для цього, повинні актуалізувати певні процедури інтерпретації ПМТ.

Для вирішення цього завдання комп'ютерно-орієнтована МКС повинна відповідати наступним вимогам:

- системної повноти як щодо множини лексем даної мови, так і щодо формул їх тлумачення;
- мати у своєму арсеналі інструментарій для семантико-синтаксичного аналізу вхідного тексту і, насамперед, вирішення проблеми потенційної многозначності вираження змісту вхідного тексту;
- надавати можливості використання як семіотичної основи для фіксації змісту вхідної інформації;
- виконувати функції мовно-категоріальної надбудови баз знань в конкретних предметних областях.

*Визначення.* МКС – це формалізована інформаційна (лексикографічна) система, що розвивається і забезпечує семантичне тлумачення елементарних одиниць природної мови та їх сталих конструкцій різної складності в процесі аналізу та інтеграції вхідної інформації, представленої природною мовою.

Як випливає з визначення, МКС є змішаною онтологією, лінгвістична частина якої названа мовно-онтологічною картиною світу. Остання є центральним компонентом мовно-онтологічної інформаційної системи. У МОКС включені тільки загальнонаукові, загальноповсякденні (або енциклопедичні) поняття, тобто основні лексичні одиниці Глумачного словника. При цьому множини мовних одиниць частин мови структуруються окремо, відповідно зі своїми сукупностями базових відношень (для повнозначних частин мови).

Однією з найважливіших функцій МОКС вважається ефективно формалізоване представлення синтаксичних одиниць природної мови. Причому, формалізми повинні бути такими, щоб результати представлялися в уніфікованій, операціонально орієнтованій формі, доступній та зручній для комп'ютерних програм.

Місце онтології МКС знаходиться нижче метаонтології категорій верхнього рівня, але вище онтологій прикладних областей. По суті, вона є лінгвістичною і категоріальною метаонтологією предметних знань поряд з іншими метаонтологіями загального характеру. Звідси одне з призначень МОКС – служити ланцюгом, який пов'язує лінгво-семантичні відношення між різними предметними областями. Іншим важливим призначенням є забезпечення формальної вимоги множинного наслідування в онтологічній ієрархії *категорія* (супертип) → *концепт* (тип) → *примітив*.

Розвиток знання-орієнтованих інформаційних систем однозначно пов'язаний з побудовою ієрархічних структур категорій верхнього рівня. Побудова системи категорій [115] є необхідним етапом розробки інтелектуальних систем, заснованих на знаннях, а також систем, які забезпечують автоматизацію розуміння змісту тексту. При цьому істотним є побудова мовно-онтологічної картини світу як основи інформаційної взаємодії систем різної матеріальної природи [3]. З іншого боку, ґрунтовна наукова розробка мовної картини світу та ієрархії сем [32] передбачає побудову моделі семантичних відношень у лексиці як деякого смислового каркасу, на якому базується глобальна «сферична» мережа різнотипних семантичних відношень. Смисловий каркас будується виходячи з двох аспектів: гносеологічного – категорії пізнання (буття, простір, час, рух, окреме, якість, кількість, відношення) і онтологічного – сфери буття (людина, природа, суспільство).

Існуючі схеми структуризації МОКС, по-перше – страждають певною суб'єктивністю підходу до систематизації, структуризації і класифікації понять, по-друге – неповнотою лексичного континууму і відношень між представленими лексичними одиницями (розуміється, перш за все, не їх теоретичний доробок, а комп'ютерна обробка, а точніше – комп'ютерне моделювання), по-третє – вони відокремлені від сучасних інтелектуальних інформаційних технологій та їх додатків.

В даній роботі:

- за основу була взята ієрархічна структура категорій верхнього рівня, синтезована на логіко-філософських засадах;
- повнота лексичного континууму забезпечується відтворенням у відповідних таблицях лексикографічної бази даних повної множини лексем Тлумачного словника;
- дослідження в своїй кінцевій меті орієнтоване на розробку апаратно-програмних засобів підтримки (граматичного процесора) на сучасній мікроелектронній базі.

Різниця між прикладною онтологією ПдО і загальною базою знань ПдО пов'язана з метою і повнотою онтології, яка є специфічною базою знань, що описує тільки факти, прийняті спільнотою користувачів як справжні, на підставі узгоджених значень понять заданого словника. Загальна база знань може також описувати факти і твердження, пов'язані зі специфічною кон'юнктурою або специфічним епістемологічним станом. Тому, в межах загальної бази знань ми можемо відрізнити два компоненти: онтологію (що містить стано-залежну інформацію) і «ядро» бази знань (що містить стано-залежну інформацію).

Крім того, слід уточнити інтервал істинності стано-незалежної інформації (або по іншому – концептуальних знань). Він повинен бути як мінімум рівним «життєвому» циклу ЗОІС, що розробляється.

МОКС є основним вихідним і підготовчим матеріалом для комп'ютерної обробки ПМТ і одним з головних блоків в мовно-онтологічній інформаційній системі. Його основною функцією є підтримка побудови формалізованого опису відображення вхідного ПМТ з відтворенням повної акторної складової семантики, присутній в ньому.

Таке відображення можна записати у вигляді:

$$G : C_o(EAT) \rightarrow C_A(EAT), \quad (2.4)$$

де  $C_o(EAT) = \bigcup_{i=1}^N C_{li}$ , де  $C_{li}$  – формалізований опис  $i$ -ої компоненти

об'єктної складової семантики слів повнозначних частин мови, які входять у вихідний ПМТ,  $i = \overline{1, N}$  – кількість вказаних слів;  $C_A(PMT) = F_A(C_o)$  – формалізований опис акторної складової семантики вихідного ПМТ.

Програмно-апаратна частина МОІС – суть семантичний процесор (СП). Завдання користувача в повному обсязі вирішуються при взаємодії лінгвістичного процесора (ЛП), СП і процесора інтерпретації бази (мета) знань і бази знань конкретної предметної області (побудова останньої представляє окрему самостійну задачу).

Обов'язковою умовою реалізації МОКС є її формалізована комп'ютерна інтерпретація (як програмними, так і апаратними засобами). Таку онтологію іноді називають наївною картиною світу. Знання про довкілля в ній вичерпуються системою понять, сформульованих на деякій мові на рівні здорового глузду, пов'язаних між собою максимально повною системою відношень, що відображають навколишній світ з усією множиною

його об'єктів і явищ, тобто представляє собою лінгвістичну проекцію буття, в якій зафіксовано досвід взаємодії людини з навколишньою дійсністю. МОКС – складова частина прагматичної моделі мовної свідомості, що є ключовим компонентом сучасних інтелектуальних ІС з природномовним представленням, обробкою та актуалізацією знань.

*Визначення.* МОКС – це відкрита, експліцитно задана на лексико-смісловому континуумі лексикографічна система, в якій множина понять формально обґрунтована і впорядкована в складну ієрархічну структуру за основними типами лексико-семантичних відношень.

МОКС як формальний опис загальноприйнятої лексики представляється стандартною схемою

$$O = \langle X, R, F \rangle, \quad (2.5)$$

де  $X$  – кінцева множина понять, характеристик, ролей і атрибутів (або контент-слів), виражених лексичними засобами ПМ, перш за все такими повнозначними частинами мови, як іменник, дієслово, прикметник і прислівник. У логіці вони представляються, як правило, одномісними предикатами  $P(x)$ ;

$R$  – повна система відношень, така як  $P(x,y)$  або  $P(x,y,z)$  (за твердженням Пірса, відношення з валентностями чотири і більше можна представити композицією двох - і трьохвалентних відношень);

$F$  – множина функцій інтерпретації, заданих на  $X$  і/або  $R$ .

Графічно МОКС представляється деяким гіперграфом, який є результатом склеювання ациклічних орієнтованих графів лексичних одиниць для кожної повнозначної частини мови. Службові частини мови враховуються на етапі зняття багатозначності і логічного представлення вихідного ПМТ.

### *Розвиток абстрактної моделі МОКС*

Уточнимо зміст поняття ієрархічного структурування в онтології МОКС щодо лексичного континууму природної мови. Смісл будь-якого висловлювання, як такого, виникає, коли текстові компоненти вступають в певні діючі відношення (об'єкт ↔ дія). Сфера дії визначається кількістю об'єктів, які вона охоплює.

Такий розгляд дуже спрощено описує структуру відображення і більше орієнтований на опис об'єктної складової семантики. Очевидним є те, що для повного відображення побудованих лінгвістичним процесором синтаксичних структур вхідного ПМТ, опису акторної складової тексту, необхідно враховувати, перш за все, базові повнозначні частини мови.

В узагальненій схемі взаємодії структурних компонентів МОКС за основу приймаємо припущення щодо подання в ній семантичної інтерпретації відповідних частин мови:



- об'єкти – як іменники;
- дії – як дієслова;
- атрибути об'єктів – як прикметники;
- атрибути дій – як прислівники.

Така інтерпретація добре узгоджується з відомою концепцією WordNet. Ієрархічна структура МОКС повинна являти собою деякий гіперграф, як композицію чотирьох ациклічних орієнтованих графів ієрархічних структур іменника, дієслова, прикметника і прислівника. Кожен з цих графів будується відповідно до своїх концептуальних та лінгвістичних відношень, властивих кожній із зазначених частин мови (відображення семантичних особливостей об'єднання іменника з прийменником враховується в графі для іменника).

Відповідно до наведеного вище опису на рисунку 2.9 представлена абстрактна модель мовно-онтологічної картини світу. На ньому прийняті наступні скорочення:

- З, П, І, Д, П, Ч – повнозначні частини мови, відповідно займенник, прикметник, іменник, дієслово, прислівник і числівник;
- АО, О, Д, АД – елементи синтаксичних структур, відповідно атрибут об'єкта, об'єкт, дія і атрибут дії;
- $KГ_{АО}$ ,  $KГ_{О}$ ,  $KГ_{Д}$  та  $KГ_{АД}$  – елементи семантичних структур першого ступеня, відповідно концептуальні графи атрибута об'єкта, об'єкта, дії і атрибуту дії;
- $Z^*$ ,  $Ч^*$  – певним чином впорядковані морфологічні таблиці для частин мови, відповідно займенника і числівника, у вихідному ПМТ;
- ПМТ\* – вихідний природномовний текст після інтерпретації модулем обробки семантики першого ступеня.

Щодо подання абстрактної моделі ще раз зазначимо, що модель морфологічних структур ПМ є спрощеною і не розкриває всіх повнозначних частин мови та взаємовідношень між ними. Таке спрощення є справедливим, тому що в цій моделі зроблено акцент саме на онтологічну складову.

Як було визначено, онтологічний граф не включає поняття нижнього рівня. Але формалізованого критерію визначення границі нижнього рівня не існує, і в кожному випадку повністю залежить від інтуїції та професійних поглядів дослідника.

Одним з етапів розширення абстрактної моделі є включення в лексикографічну базу даних (ЛБД) додаткових таблиць, які включають як повнозначні частини мови (дієприкметник і дієприслівник), так і службові частини (прийменник, сполучник і частка) і стійкі словосполучення. При цьому доцільно таблицю дієприслівників реалізувати як підтаблицю в таблиці «Дієслово», виділити безособову форму дієслова в окрему підтаблицю, таблицю «Іменник» розбити на три підтаблиці «Іменники\_загальні», «Іменники\_власні назви» і «Іменники\_віддієслівні», а також взяти до уваги ряд інших граматичних особливостей української (російської) мови.

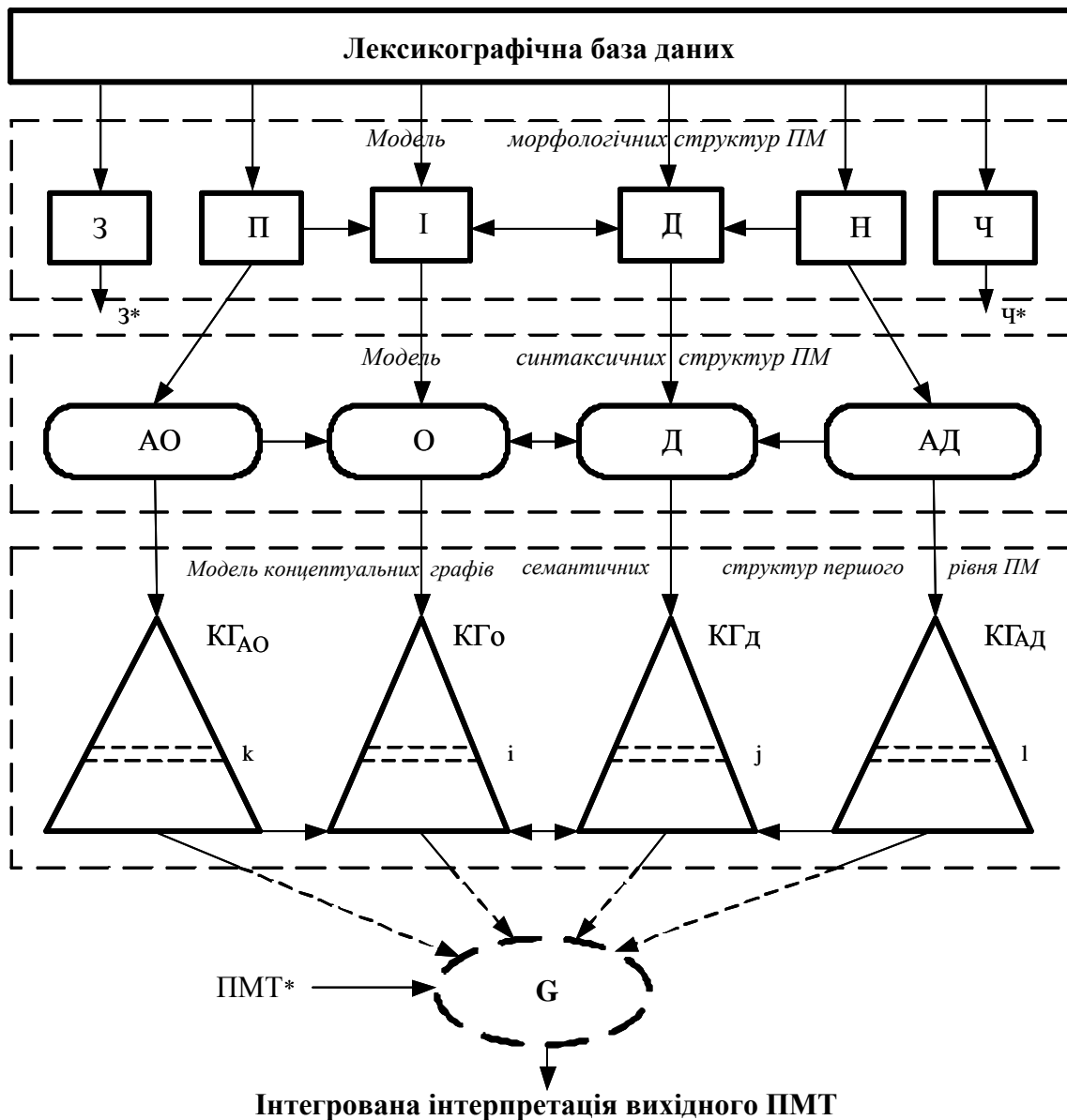


Рис. 2.9. Абстрактна модель МОКС

Наступним етапом є розширення синтаксичних структур природної мови. Додаються два блоки, пов'язаних з таблицями «Дієприкметник» і «Дієприслівник», моделі морфологічних структур: «Атрибут атрибута об'єкта» – для дієприкметника та «Атрибут атрибуту дії» – для дієприслівника. Крім того, таблиця «Прийменник» пов'язується з синтаксичною структурою «Об'єкт», а таблиці «Сполучник» і «Частка» включаються до фразової синтаксичної структури для формування більш складних конструкцій.

Модель семантичних структур першого рівня (або об'єктної семантики) сформована з тих же самих КГ для об'єкта, дії, атрибута об'єкта і атрибуту дії. При цьому в узагальненому вигляді виконується їх структурування.

Для КГ об'єкта – це іменники, які позначають загальні поняття → іменники, які позначають сутності з ознакою дії → іменники, які позначають узагальнені об'єкти → іменники, які позначають конкретні об'єкти.

Для КГ дії – дієслова, які виражають семантичні теми, структуруються на верхніх рівнях графа; інші дієслова структуруються на нижніх рівнях графа.

КГ атрибута об'єкта розбивається на три підграфа, відповідно якісних, відносних і присвійних атрибутів об'єкта.

КГ атрибута дії своїми гілками зв'язується з відповідними групами семантичних тем дієслів.

Формальне відображення представляє семантичну модель другого рівня (або акторної семантики) і включає підмножини логіки першого порядку концептуальних графів, обчислення предикатів та мови представлення знань (KIF), спроектованих на вихідний ПМТ.

Наведемо у вигляді таблиці 2.2 належним чином структуровані формальні основи застосування правил, принципів, абстракцій і відношень при розвитку метаонтології категорій верхнього рівня і МОКС (окремо для рівнів концептів і примітивів) [5, 26–28, 38, 52, 116–118], і деякі правила-рекомендації, які необхідно враховувати при класифікації концептів.

Семантичні відношення нижнього рівня (для деякого речення ПМ) визначаються на основі роботи лінгвістичного аналізатора і залежать від архіву лінгвістичних знань: словники словоформ для морфології, граматичні правила для синтаксису і концептуальні патерни (або множини слотів фреймових структур) для семантики. Вони можуть бути отримані з лінгвістичних категорій і є основними стандартними блоками для визначення більш складних відношень в концептуальних графах, базах даних або експертних системах. Наведемо деякі правила-рекомендації, яких (в тому числі) необхідно дотримуватися при побудові онтології предметної області [52].

1. *Процедура переходу між сусідніми рівнями повинна бути сумірною.* Це означає, що об'єднання об'ємів або з'єднання значень концептів нижнього рівня повинне складати концепт верхнього рівня.

2. *Процедура переходу повинна виконуватися по одній основі* або характеристика, яка вибирається як основа поділу, в ході розподілу не повинна підмінятися іншою характеристикою.

3. *Компоненти діленого концепту* повинні виключати один одного. Це означає, що їх обсяги не повинні містити спільних елементів або їх значення не повинні мати спільних частин.

4. *Процедура переходу повинна бути послідовною.* Це означає, що від родового поняття необхідно переходити до видових понять одного і того ж рівня, а від цілого до частин, частинам частин і т.д.

## Формальні основи структурування онтологій

<b>I. Метаонтологія категорії верхнього рівня</b>	
Види абстракцій	Класифікація, узагальнення, агрегація і асоціація
Принципи класифікації	Фундаментальні принципи дихотомії Аристотеля, трихотомії Пірса і решіток Лейбніца, природна класифікація
Типи класифікаційного поділу	Таксономічні, мереологічні, просторові і часові
Спадкування ознак	Множинне
<b>II. Мовно-онтологічна картина світу</b>	
<b>А. Рівень концептів</b>	
Концептно-ролеві відношення	Виробляють два репрезентативні примітиву, відповідно названі приєднанням і кореферентним зв'язком. На природну мову переводяться різними граматичними формами дієслів <i>мати</i> і <i>бути</i>
Семантичні відношення	Класифікаційні – рід-вид, ціле-частина, клас-елемент, вищерозташований-нижчерозташований, клас-підклас; Ознакові – об'єкт-атрибут, об'єкт-дія; Кількісні – мати міру
Лінгвістичні відношення	Гіпонімія, метонімія, синонімія і антонімія
<b>Б. Рівень примітивів</b>	
Предикативні відношення	Узагальнено-одиничного характеру, окремо-одиничного характеру і одинично-одиничного характеру
Семантично-рольові відношення	Класифікаційні – мати ім'я, бути еталоном; ознакові – мати значення атрибута, мати значення способу дії; кількісні – мати значні міри; порівняння – рівне, порівняльне, більше, більше або дорівнює, менше, менше або дорівнює, незрівняне; приналежності – простої приналежності; прості часові – бути одночасно, бути раніше, бути пізніше, збігтися в часі, перетинатися в часі, бути всередині в часі, починатися одночасно, закінчуватися у часі; прості просторові – збігатися в просторі, бути ліворуч, бути справа, бути попереду, бути позаду, бути зверху, бути знизу, бути навскоси, перетинатися в просторі, торкати, розміщуватися на, розміщуватися в; каузальні – бути метою, бути мотивом, причина-наслідок; інструментальні – служити для, бути засобом для, сприяти, бути інструментом, бути допоміжним засобом; інформаційні – бути відправником, бути одержувачем, бути джерелом інформації; порядкові – бути наступним, бути поточним, бути найближчим; модальні - можливості, дійсного втілення, необхідності; модифікаційні; квантифікаційні – загальності, існування, унікальності, одиничності, можливості, необхідності; кореляційні

Одним з можливих варіантів перекладу ПМ-речень у формальні структури може бути переведення у концептуальні графи [62], так як вони з самого початку розроблялися для цих цілей, що забезпечують читабельність, наочний ітераційний процес виявлення помилок перекладу і в той же час є підмножиною логіки першого порядку. Крім того, концептуальні графи мають можливість формування процедури вкладеності, яка відповідає ланцюжку синтаксичних одиниць «фраза → пропозицію → абзац → параграф → розділ → текст → том» або фрагментам дійсності «елементарна ситуація → ситуація → сценарій».

## 2.8. Системна інтеграція лінгвістичних і предметних онтологій

Розробка і комерційне використання баз знань і відповідного інструментарію в багатьох прикладних областях (Knowledge based engineering systems) на основі онтолого-керованих інформаційних систем у багатьох науково-дослідних центрах і корпораціях привели до зростання теоретичних розробок формалізованих методологій проектування онтологічних структур. При цьому сутність зазначених методологій зводилася до формального обґрунтування структурування ієрархічного дерева онтології (формалізованої побудови наборів концептів і пов'язуючі їх концептуальні відношення або категоризації) і розробці формальних мов представлення знань, які описують аксіоматизацію концептів предметної області. Наступним кроком розвитку теорії баз знань була необхідність теоретично обґрунтованого об'єднання (або системної інтеграції) вже розроблених як загальнодоступних онтологій, так і комерційних баз знань для різноманітних прикладних задач, проблем, цілих предметних областей і трансдисциплінарних знань широкого призначення.

В загальному вигляді процес, що забезпечує системну інтеграцію множини онтологій, можна описати наступною формулою

$$O = \prod_i^U O_i, i = \overline{1, N}, \quad (2.6)$$

де  $\prod_i$  – знак концептуального об'єднання. Смысл цього знака полягає в системній інтеграції вихідних онтологічних графів з урахуванням областей визначень  $O_i (i = \overline{1, N})$  і їх взаємозв'язку (взаємодії).

Узагальнена архітектура знання-орієнтованої системи, ефективно реалізує процедури обробки предметних знань, розвитку БЗ та системної інтеграції, представлена на рисунку 2.10.

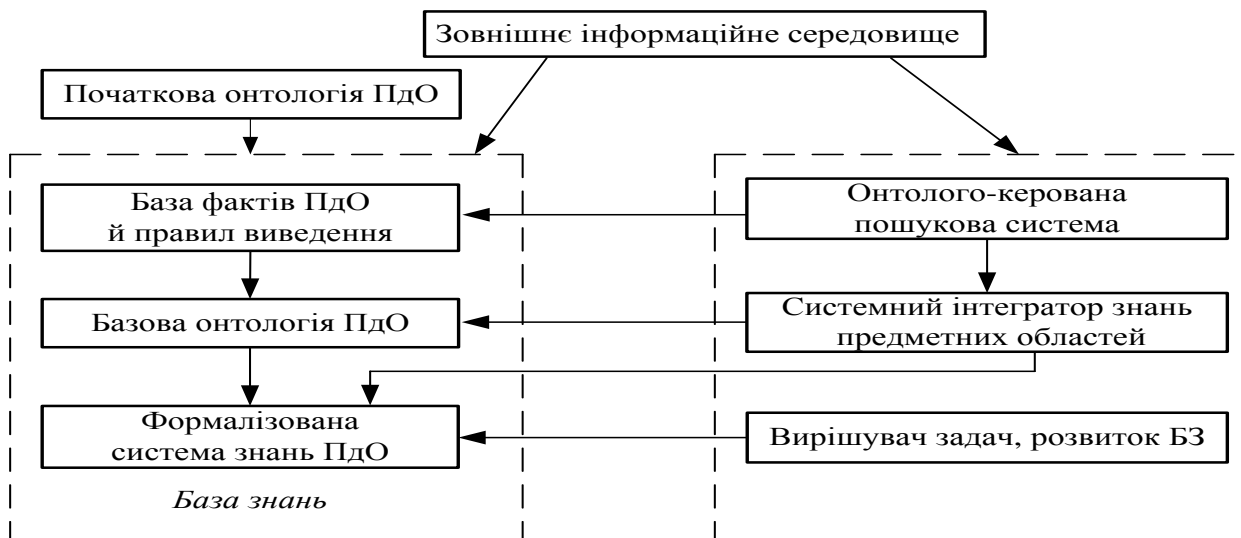


Рис. 2.10. Архітектура знання-орієнтованих систем, що розвиваються

Реалізація технологій представлення та обробки знань і процесу системної інтеграції знань передбачає врахування різних формально-методологічних вимог, критеріїв і оцінок. Наведемо основні з них.

1. Сутність методів онтологічного дослідження як розділу системного аналізу передбачає:

- побудова статичних і динамічних моделей;
- дослідження структури ПдО і взаємозв'язку компонент знань;
- дослідження поведінки системи, яка описує задану ПдО – розробка алгоритмів процедур композиції знань і діаграм станів.

2. Вплив предметної області на вид онтології.

3. Правила формування достовірних тверджень і виведень, що описують систему та її поведінку.

4. Категоріальний рівень онтології ПдО і відповідна йому система відношень.

5. Характеристики онтографа (середній коефіцієнт розгалуження  $i$ -ї вершини, кількість рівнів розгалуження, кількість типів відношень, в тому числі по підтипам).

Пізнання йде від вивчення речей до вивчення процесів (у тому числі взаємодії компонент, включаючи рівень «суб'єкт-об'єкт»), від стійко функціонуючої системи – до системи, яка змінюється і розвивається, а результатом, наслідком процесу розвитку є виникнення нової якості [15].

#### *Формалізовані аспекти інтеграції та її оцінка*

Об'єм знань  $W$  в предметних областях можна оцінити через характеристики (параметри) їх формально-онтологічних представлень. Зокрема, при представленні онтологічним графом (без урахування типів відношень і складності функцій інтерпретації) величина  $W$  може характеризуватися числом вершин ОГ. У разі простої деревовидної структури це число може бути виражене формулою

$$W = \sum_i \sum_h \sum_l O_i \cdot S_{h,l} \quad (2.7)$$

де  $O_i$  – онтограф  $i$ -ої предметної області,  $i = \overline{1, N}$ ;  $S_{h,l}$  – ступінь вершини, що дорівнює числу ребер, що виходять з неї,  $h = \overline{1, H}$  – кількість рівнів ОГ,  $l = \overline{1, L_h}$  – номер вершини на відповідному ( $h$ -му) рівні ОГ.

При рівномірній щільності розподілу ОГ, тобто при  $S_{h,l} = S$  ( $h, l = 1, 2, \dots$ ) зводиться до відомої формули суми геометричної прогресії

$$W = \sum_i O_i \frac{1-S^h}{1-S} \quad (2.8)$$

Врахування типів відношень і складність функцій інтерпретації призводить до ОГ зі зваженими вершинами і ребрами. Вираз (1.10) при цьому перетворюється в вид

$$W = \sum_i \sum_h O_i \cdot \left( \alpha_l + \sum_j \beta_{l,j} \right) \quad (2.9)$$

де  $\alpha_l$  і  $\beta_{l,j}$  – значення вагових функцій відповідних відношень і функцій інтерпретації, приписані вершинам ( $\alpha_l$ ) і ребрам ( $\beta_{l,j}$ ) ОГ. Вираз (1.11) дає повну оцінку складності ОГ, а відношення  $\omega = W^0/W$  характеризує середню щільність зваженого ОГ.

Розглянуті оцінки дозволяють порівнювати різні варіанти представлення знань про предметні області, а також відслідковувати процес еволюції наукових теорій.

Процес розвитку знань в будь-якій ПдО пов'язаний з її аналізом, концептуалізацією і побудовою формальної теорії. При цьому формалізація в загальному випадку відноситься до чотирьох основних видів представлення інформації

$$I = I(V, A, T, G), \quad (2.10)$$

тобто до вербального ( $V$ ), аналітичного ( $A$ ), табличного ( $T$ ) і графічного ( $G$ ). Між ними існує взаємно-однозначна відповідність, не завжди на практиці реалізується строго і повністю. Тому всі вони знаходять своє, цілком певне місце при описі наукової теорії. В багатьох випадках коректним виявляється їх обмеження до двох: вербального і аналітичного.

Як правило, процес розвитку теорій супроводжується перерозподілом обсягу інформації про предметну область між вербальною і формальною компонентами, тобто між природномовним описом предмета дослідження і формально-аналітичним (формульним  $\Phi$ , табличним  $T$ , графічним  $G$

представленням суті теорії). Очевидно, що формалізоване представлення є більш компактним, а головне, – більш строгим і придатним для комп'ютерної обробки.

### *Практична інтеграція онтологій*

Накопичений досвід практичного використання SUMO та інших онтологій (в першу чергу WordNet і FrameNet) показав необхідність в об'єднанні двох і більше онтологій (побудові взаємних відображень концептів) в єдину концептуальну структуру. Два види онтологій (лінгвістична і концептуальна) відображають дві сторони концептуалізації реального світу. Якщо в WordNet концептуалізація відображається в термінах природної мови, то в SUMO концепти організовані в логічну структуру. При цьому об'єднання передбачає побудову відображення між відповідними вузлами різних онтологій.

*Відображення SUMO ↔ WordNet.* Робоча група SUO WG розробила таке відображення. Воно зіставляє кожному синсету в базі даних WordNet шляхом тегування відповідного концепту в SUMO. При цьому вид відношення між синсетами WordNet і концептами SUMO можуть бути: синсет еквівалентний концепту; синсет є більш вищим класом, ніж концепт, і синсет є елементом концепту. Розроблені файли відображення дозволяють співвіднести слова природної мови в терміни SUMO, використовуючи синсети WordNet як проміжний рівень [119].

Наведемо приклад формалізованого запису для першого виду відношень між синсетами WordNet і концептами SUMO:

**00008864 03 n 03 plant 0 flora 0 plant\_life 0 027  
@ . . . | a living organism lacking the power of  
locomotion &%Plant=**

Префікс '&%' вказує на те, що термін взятий з онтології SUMO, а суфікс '=' вказує, що відношення відображення – синонімія.

Мотивацією для побудови таких відображень може бути наступне.

Відображення можуть функціонувати як індекси природної мови до концептів в онтології, як міст між цими структурованими концептами і нетекстовою лінгвістичною структурою. Розробники створили спеціальний інструмент і вбудували його в SUMO-браузер, який дозволяє користувачеві задавати терміни природної мови і бачити на екрані комп'ютера відповідні концепти онтології SUMO з відповідною формальною аксіоматизацією. Це дозволить значно спростити процес інжинірингу знань і моделювання даних. Крім того, відображення можуть служити важливим джерелом знань для застосувань NLP, зокрема при реферуванні текстів і семантичному пошуку документів в мережі. І ще, за допомогою відповідного інструментарію можна перевірити на повноту ступінь завершеності проектованої онтології.



*Відображення SUMO ↔ FrameNet.* Онтології SUMO і FrameNet є відносно сформованими засобами, але доцільно об'єднати їх сильні сторони для істотного спрощення автоматичної обробки текстів NLP. Зокрема, NLP-застосування, які використовують FrameNet, вимагають знань про можливі заповнювачі елементів кадру (FEs). Наприклад, семантичному фреймовому аналізатору необхідно знати, чи є певний фрагмент тексту (або іменована сутність) строгим заповнювачем для FE, тобто йому необхідно перевірити, чи є тип заповнювача FE сумісним з типом іменованої сутності. Тому в FrameNet представлено близько 40 семантичних типів (FT) як обмеження на заповнювачі FEs і мають відповідності з SUMO класами як FTs [120]. Фрагмент зв'язування SUMO і FrameNet представлений на рисунку 2.11.

Слід зазначити, що існують і інші проекти зв'язування лексикон-онтологій [121]. Але розглянуте зв'язування SUMO і FrameNet має ряд переваг. Справа в тому, що FrameNet, на відміну від WordNet, моделює семантичні та синтаксичні валентності, включаючи високоякісні анотації. Фреймові семантики природно представляють крослінгвістичні абстракції і нормалізацію парафраз. Що стосується SUMO, то вона значно ширше DOLCE, а на відміну від Сус є відкритою для доступу.

При інтеграції знань слід враховувати прикладні аспекти. Принципово розрізняти області:

1. взаємодії з користувачем – когнітивізація представлення, пошук інформації, створення початкової онтології і початкової системи знань;
2. розвитку системи знань на основі наявних, генерація нових знань.

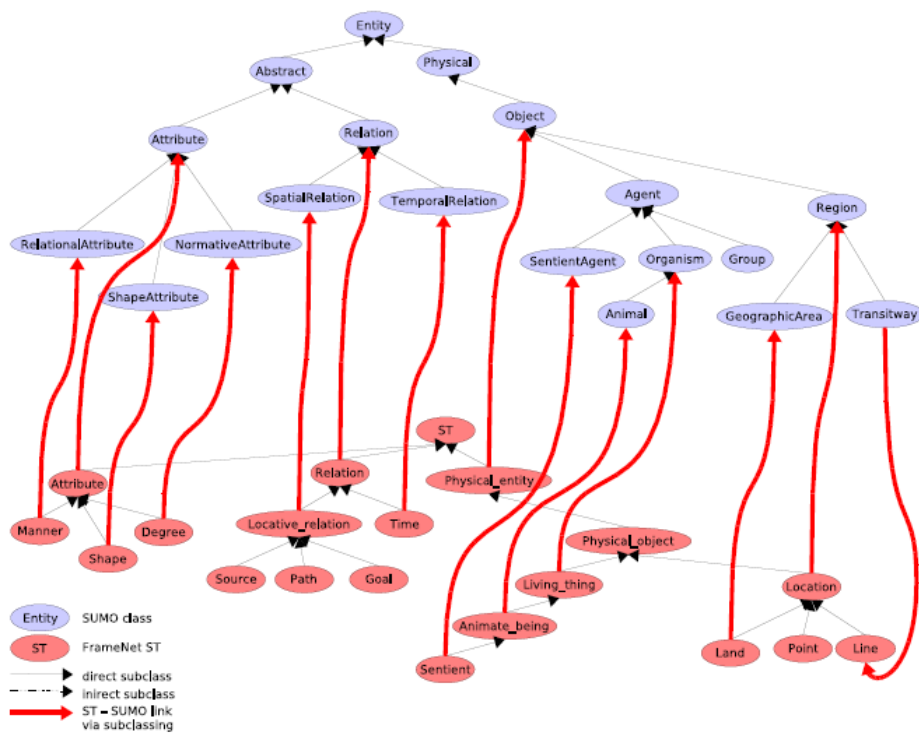


Рис. 2.11. Фрагмент зв'язування SUMO і FrameNet

## 2.9. Методологія розробки онтології ПдО

### *Загальний підхід до проектування*

Відома методологія структурного аналізу і проектування (SADT) складних систем в довільній предметній області [122]. Ця методологія породила сімейство методик (і відповідних стандартів) IDEF (Integrated DEFinition), орієнтованих на розробку моделей ПдО й акцентуючих увагу на деякому конкретному аспекті проектування [123, 124]. Зокрема:

- методика IDEF0 рекомендована для змістовного аналізу та функціонального проектування складних систем управління, в тому числі і програмного забезпечення. Опис об'єктів і процесів в методиці виконується у вигляді ієрархічної сукупності діаграм з лаконічним описом функцій. Блоки на діаграмах виражають функції, тому їх назви – дієслова або віддієслівні іменники;

- методика IDEF1x призначена для інформаційного моделювання, заснована на концепції «сутність-зв'язок». Зазвичай відповідним пунктом для розробки інформаційної моделі є IDEF0-модель;

- методика IDEF3 описує поведінкові аспекти конкретних програм, розглядає послідовність виконання та причинно-наслідкові зв'язки між ситуаціями і подіями для структурного представлення знань про ПдО. Якщо IDEF0 пов'язана з функціональними аспектами і відповідає на питання «Що робить система?», то в IDEF3 деталізуються IDEF0-функції. Ця модель відповідає на питання «Як система це робить?»;

- методика IDEF5 призначена для онтологічного аналізу ПдО, аналізу основних термінів і понять, що використовуються для опису об'єктів і процесів, меж використання, а також взаємозв'язків між ними. Служить для ефективного дослідження і документування: словника термінів, що використовується при описі характеристик об'єктів і процесів, що мають відношення до розглянутої ПдО, точних і однозначних визначень усіх термінів та класифікації логічних взаємозв'язків між ними.

Для кожної з методик сімейства IDEF розроблені етапи та стадії побудови моделі ПдО, мови і діаграми представлення результатів. Аналіз отриманих в результаті застосування методик IDEF моделей ПдО та їх описів з точки зору формалізації та комп'ютерної обробки показав, що, по суті, вони (функціональні моделі та описи) є змістовне технічне завдання на проектування ОнС, що включає онтологію об'єктів (сутностей), онтологію процесів і онтологію задач. Опис самої ОнС має дещо інші цілі. Він орієнтований на комп'ютерне представлення за допомогою стандартизованих мов предметних знань з метою широкого використання спільнотою користувачів.

Отже: 1) методики IDEF і системно-онтологічний підхід використовують одні й ті ж множини сутностей ПдО і ПрП, акцентуючи увагу на різних сукупностях характеристик і атрибутів; 2) кінцеві цілі цих

двох підходів відрізняються: для IDEF – функціонально-блокова модель, а для системно-онтологічного підходу – онтолого-сміслова модель.

Виконаний вище аналіз дозволяє стверджувати, що методики IDEF і СОП вирішують різні задачі (з певним «перекриттям» деяких функцій). Вельми очевидно це проявляється для пари «онтологія процесів ↔ поведінкова модель ПдО».

Вочевидь, етапи проектування онтологічних систем виходять з прийнятих в методології SADT фаз проектування складних систем: 1) аналіз – визначення того, що система буде робити, 2) проектування – визначення підсистем та їх взаємодію; 3) реалізація – розробка підсистем окремо, об'єднання – з'єднання підсистем в єдине ціле, 4) тестування – перевірка роботи системи; 5) установка – введення системи в дію; 6) функціонування – використання системи.

### *Попередній аналіз предметної області*

У всі методології включений етап попереднього аналізу ПдО або складання змістовного ТЗ на проектування [2, 3, 5, 122–127]. Цей етап (як і процес проектування бази знань ПдО в цілому) має складний аналітичний характер і полягає в багаторазовому абстрагуванні, в результаті якого з усього різноманіття сторін і властивостей сутностей предметної області виділяються найбільш істотні, релевантні конкретним задачам. Знання предметної області, розуміння суті відповідних в ній процесів, законів, правил і обмежень, які керують її розвитком, є необхідною умовою успішного вирішення задач, що стоять перед дослідником. Більше того, наявність таких знань є необхідною умовою постановки, формулювання цих задач, без чого неможливе саме рішення [99]. Онтологічні системи покликані зробити знання колективним надбанням широкого кола осіб, дати потужний інструмент для фіксації, придбання та обробки знань, перевірки їх на несуперечність, повноту і т. п. Крім того, складається систематизоване представлення знань про ПдО, виявляються джерела формування елементів множин і процедур, задач, що виконуються в аналізованій ПдО. Складається і документується словник термінів ПдО.

Етап попереднього аналізу ПдО включає:

- обґрунтований вибір точного (і достатнього) фрагмента ПрП, щодо якого будуть ставитися і вирішуватися задачі користувача;
- вибір методів і процедур системно-онтологічного аналізу, якими, зокрема, можуть бути абстрагування і конкретизація, композиція і декомпозиція, структурування, кластеризація і класифікація, тестування і верифікація;
- складання детального словника термінів і його розбиття на підмножини термінів-об'єктів, термінів-процесів і термінів, які іменують задачі та методи.

Якщо предметна область (і проблемний простір) представляє складну систему, то слід розглянути питання про попередній етап проектування на

основі методик IDEF, які доповнюють описані вище кроки проектування. Тим більше, що відомі інструментальні засоби аналізу ПдО і побудови баз знань (наприклад, KADS-системи, SIMER+MIR) по ряду причин не можуть бути використані [3].

Як правило, методика зводиться до алгоритму, який носить ітеративний характер. Для процесу розробки необхідно передбачити ряд «контрольних точок» для перевірки отриманих результатів на відповідність обраним критеріям. Зазначені критерії повинні мати співвідношення із заданими критеріями на проектування бази знань ПдО, оскільки створення останньої є метою для розробників. Оптимальний результат, як правило, залежить від ступеня опрацювання передбачених програм і варіантів використання онтології.

### *Онтологія об'єктів ПдО*

*Визначення 1.28.* Під **онтологією об'єктів** предметної області розуміється четвірка [2]:

$$O^o = \langle X, R, F, A(D, R_s) \rangle, \quad (2.11)$$

де  $X = \{x_1, x_2, \dots, x_i, \dots, x_n\}$ ,  $i = \overline{1, n}$ ,  $n = \text{Card } X$  – скінченна множина концептів (понять-об'єктів) заданої ПдО;

$R = \{R_1, R_2, \dots, R_k, \dots, R_m\}$ ,  $R \subseteq X_1 \times X_2 \times \dots \times X_n$ ,  $k = \overline{1, m}$ ,  $m = \text{Card } R$ , – скінченна множина семантично значущих відношень між поняттями-об'єктами ПдО. Вони визначають тип взаємозв'язку між поняттями. У загальному випадку, відношення ділять на загальнозначущі (з яких виділяють, як правило, відношення часткового порядку) і конкретні відношення заданої ПдО;

$F: X \times R$  – скінченна множина функцій інтерпретації, заданих на поняттях-об'єктах і/або відношеннях;

$A$  – скінченна множина аксіом, яка складається з множини визначень  $D_i^!$  і множини обмежень  $R_s^!$  для поняття  $X_i$ . Визначення записуються у вигляді тотожно істинних висловлювань, які можуть бути взяті, зокрема, з тлумачних словників ПдО. В них можуть бути зазначені додаткові взаємозв'язки понять  $X_i$  з поняттями  $X_j$ . В множині обмежень  $R_s^!$  можуть бути задані обмеження на інтерпретацію відповідних понять  $X_i$ .

Онтологія визначає загальноживані, семантично значимі «понятійні одиниці знань», якими оперують дослідники і розробники знання-орієнтованих інформаційних систем. Вона відокремлює «статичні» і «динамічні» компоненти знань ПдО від операціональних знань. На відміну від знань, закодованих в алгоритмах, онтологія забезпечує їх уніфіковане і багаторазове використання різними групами дослідників, на різних комп'ютерних платформах при вирішенні різних задач.

### *Побудова компонент онтології*

Найбільш істотним компонентом концептуальної моделі ПдО є множина понять заданої предметної області. Нагадаємо деякі твердження, безпосередньо пов'язані з побудовою онтографу ПдО [128].

Всі поняття (або концепти) поділяються на ряд класів (по семантичній залежності).

- В залежності від відображення виду або роду предметів – на *видові* і *родові* поняття.
- Залежно від відображення частини або цілого предметів – на *поняття-частини* і *поняття-цілі*.
- В залежності від кількості відображуваних предметів – на *одиночні* і *загальні* поняття.
- В залежності від відображення предмета або властивості, абстрагованого від предмета, – на *конкретні* поняття і *абстрактні* поняття.

Онтологія ПдО – це концептуальна модель реального світу і її поняття повинні відображати цю реальність.

Побудова множини  $X$  вважається найбільш важливим моментом при розробці онтології ПдО. Вона повинна бути обов'язково не пустою. Співвідношення між *Card X*, *Card Ri* *Card F* характеризують онтологію за функціональною ознакою.

Для добре опрацьованих предметних областей за основу множини елементів  $\{x_i\}$  може бути взято вміст відповідних тлумачних словників. В іншому випадку слід скласти повний список термінів, в якому вказати (причому перетин об'ємів і сенсу понять в такому попередньому списку не суттєве):

- чим є кожен термін – поняттям-класом предметів або конкретним поняттям;
- вказати для кожного терміна можливі суттєві відношення з іншими термінами зі списку;
- описати можливі істотні властивості понять.

Відомо, що в будь-якій предметній області існують терміни-синоніми. Для них в онтології відводиться тільки одне поняття, в аксіомах якого може бути вказаний синонімічний ряд термінів. Іншими словами, синоніми одного і того ж поняття не представляють різні класи.

Далі слід уточнити і визначити остаточний список класів-понять, імена яких будуть входити в онтологію, що розроблюється, і є вершинами онтографу. Також слід прийняти єдині *правила присвоювання* імен поняттям і їх властивостям.

Потім, можливо, слід повторити деякі фрагменти процесу аналізу ПдО (з прив'язкою до складеного списку понять), виконані на попередньому етапі. Відзначимо, що в число зазначених вище «контрольних точках» (точок входу ітерації) повинно бути включено завершення розробки будь-якого компонента онтології.

В результаті має бути отриманий повний список істотних для заданої ПдО (і передбачуваних додатків) понять і їх машинно-інтерпретовані формулювання.

Побудова множини  $R$  також заснована на результатах попереднього етапу аналізу ПдО. По суті, потрібно встановити між елементами  $x_i \in X$  семантичні  $k$ -арні відношення  $R_k \subseteq X^k$ . Іншими словами, необхідно побудувати множину ребер, що зв'язують вузли *орієнтованого графа*. В якості вузлів онтографа виступає множина понять ПдО. Вершиною (або вершинами) онтографа є родові поняття, яке не має надкласу, а найнижчий рівень представляють конкретні поняття (примітиви), тобто не мають видових понять в заданій ПдО.

На практиці множину  $R$  на початковому етапі подають деяким узагальненим відношенням «*вище-нижче*». Відомо кілька підходів для розробки ієрархії класів: процес *низхідної* розробки, процес *зростаючої* розробки і *комбінований* процес розробки. Останній найчастіше використовується розробниками, так як він є більш природнім, спочатку оперує поняттями середнього рівня, до яких найчастіше звертаються розробники. Потім ці поняття узагальнюються і обмежуються.

Наприкінці даного підетапу слід співвіднести розроблені класи та їх ієрархії з результатами попереднього аналізу ПдО. Зокрема, уточнюються залежності для конкретних пар  $(x_i, x_j)$ . В процесі співвіднесення (і побудови ієрархії) слід враховувати, що [89]:

- прямі підкласи в ієрархії повинні розташовуватися на одному рівні узагальнення;
- клас може бути підкласом декількох класів, і тоді він може наслідувати властивості від всіх цих класів;
- якщо клас має тільки один прямий підклас, то, можливо, при моделюванні допущена помилка або онтологія неповна;
- якщо у даного класу є більше дюжини (іноді говорять про число 7) підкласів, то, можливо, необхідні додаткові проміжні класи;
- в онтології число класів співвідноситься з числом передбачуваних додатків.

І слід пам'ятати, що не існує єдиної правильної ієрархії класів.

Описана побудова онтографа є спеціальним видом класифікації понять ПдО – онтологічною класифікацією.

Побудова множин  $F$  і  $A$ . Залежно від функціональної орієнтації проектованої онтології множини  $F$  і  $A$  можуть інтерпретуватися по-різному [92, 125 – 127 та ін.]:

1)  $A \equiv F$  – множина аксіом тотожна множині функцій інтерпретації. В цьому випадку встановлюються істотні зв'язки між компонентами онтології і варіантами її використання. Основним призначенням такої онтології є однозначна інтерпретація понять, що входять в онтологію, спільнотою користувачів;

2) – множина аксіом  $A \subseteq F$  не тотожна множині функцій інтерпретації. В аксіомах задаються а) базові функції (підмножина  $F$ ) або б) додаткові відношення (не є елементами множини  $R$ ) між поняттями, обмеження та умови, які аналізуються в машині виведення ОнС і використовуються в процесі вирішення задач;

3) – множина аксіом  $A \supseteq F$  не тотожна множині  $F$ . Функції інтерпретації розглядаються як спеціальний вид відношень на множини понять  $F : x_1 \times x_2 \times \dots \times x_{n-1} \Rightarrow x_n$ . В цьому випадку встановлюються істотні зв'язки між вже розробленими компонентами онтології і сукупністю задач передбачуваного додатку (додатків). Онтології з таким поданням використовуються в питально-відповідних системах, в яких результатом є одне зі значень двоелементної множини або ім'я предиката.

В кінцевому рахунку, незалежно від того, яке з цих формулювань буде прийняте, ефективність розробленої онтології буде визначатися кінцевими результатами додатків.

Крім того, з повного списку відібраних в онтологію термінів не всі представляють поняття. Існують терміни (наприклад, рольові), які відповідають властивостям певних класів-понять. Такі властивості слід прив'язати до опису самого загального класу, до якого вони належать. А підкласи цього класу будуть наслідувати вказану властивість (звичайно, якщо між ними встановлено певне відношення часткового порядку).

*Властивості понять* мають певні значення, такі як *тип* значень, *потужність* значень, *дозволені* значення (для даного класу) та інші. Наприклад, значення бувають з одиничною потужністю, потужністю без обмежень і потужністю з деяким допустимим інтервалом.

На основі побудованих множин кортежу можна синтезувати концептуальну модель ПдО, наприклад, за допомогою відомого інструментального засобу Protégé і отримати формальний опис розробленої онтології на одній з мов опису онтологій, а також графічне представлення онтографа.

### *Онтологія процесів ПдО*

Синонімами онтологій об'єктів і процесів є відповідно статична і динамічна онтології ПдО. У науково-технічній літературі, коли говорять про онтології ПдО, то мається на увазі її статична складова. Саме компоненти останньої найбільш розроблені, як в літературі з філософії, так і в конкретних описах ряду предметних областей. Поведінковий опис сутностей-процесів найчастіше виконується у вигляді графічних діаграм і природномовних описів. Розробка ж бази знань не є прямою метою зазначених методик. Тому методики розробки онтології процесів практично невідомі, хоча в деяких відомих онтологіях верхнього рівня [9, 83, 114, 129] сутність поняття «Процес» розглянута досить детально.

На рис. 2.12 представлений синтезований онтограф, який представляє схему початкової ділянки поняття «Процес», а саме тієї її частини, яка відповідає процесам в науково-технічних предметних областях (гілки онтографа «Соціальний процес», «Матеріальний процес» і їм подібні не розглядаються).

Категорія *Процес* розглядається як *Дійсність* і *Подійність*, на відміну від категорії *Об'єкт*, що характеризується як *Дійсність* і *Тривалість*[27]. В першу чергу *Процес* розглядається як залежна від часу категорія і потім поділяється за видами змін, наявністю початкових і кінцевих точок і т. д. Далі *Процес* поділяється на безперервний і дискретний. Перший з них характеризується наявністю експліцитних початкової і кінцевої точок або без явної вказівки цих точок. Другий вид процесу вказує, що зміни відбуваються дискретними кроками, названими подіями, які чергуються з періодами спокою, названими станами. Наведена схема початкового розвитку онтології процесів не відображає всіх характеристик (підстав розгалужень в онтографі) категорії «Процес», навіть для тієї її частини, що представлена на рисунку 2.13.

На рисунку 2.12 представлена загальна схема онтології процесу ПдО, в якій категорія «Процес» представлена онтографом з  $p$  рівнями і  $n_p$  підпроцесами (Пп) на кожному рівні. Передостанній ( $p-1$ ) рівень представлений множиною дій (Д), на які розбивається кожен Пп попереднього рівня. В свою чергу, кожна дія на останньому ( $p$ -му) рівні розбивається на послідовність операцій.

Зв'язки між підпроцесами для сусідніх рівнів відповідають відношенням «Ціле-частина», а в середині кожного рівня – деякою змішаною формою організації з'єднань. На рисунку 2.13 показаний окремий випадок такої організації – паралельний. Подальший розвиток (конкретизація) онтології процесів можливий, коли задана конкретна предметна область і відповідний проблемний простір, а в більш вузькому смислі – конкретні ознаки розгалужень (умови ініціювання Пп, умови закінчення Пп та обмеження) в онтографі.

### *Онтологія задач проблемного простору ПдО*

З предметно-проблемних знань в загальному випадку виділяють поняття і процеси ПдО, поняття і класи задач ПрП, а також методи вирішення останніх. Сюди ж відносять і алгоритми, що реалізують відповідні методи. Перераховані типи знань з темпоральною ознакою поділяються на статичні і динамічні, а за онтологічною схемою на:



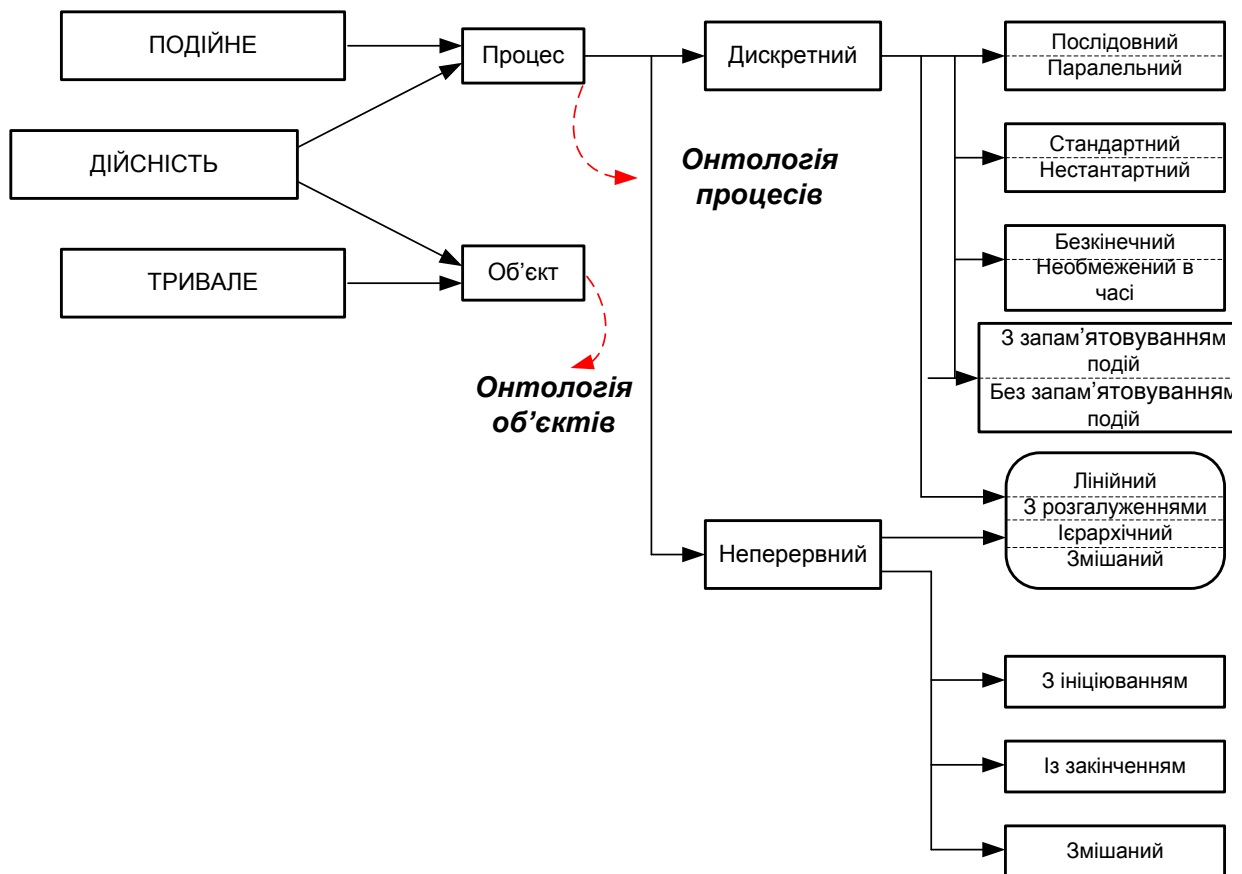


Рис. 2.12. Початкова ділянка онтології процесів

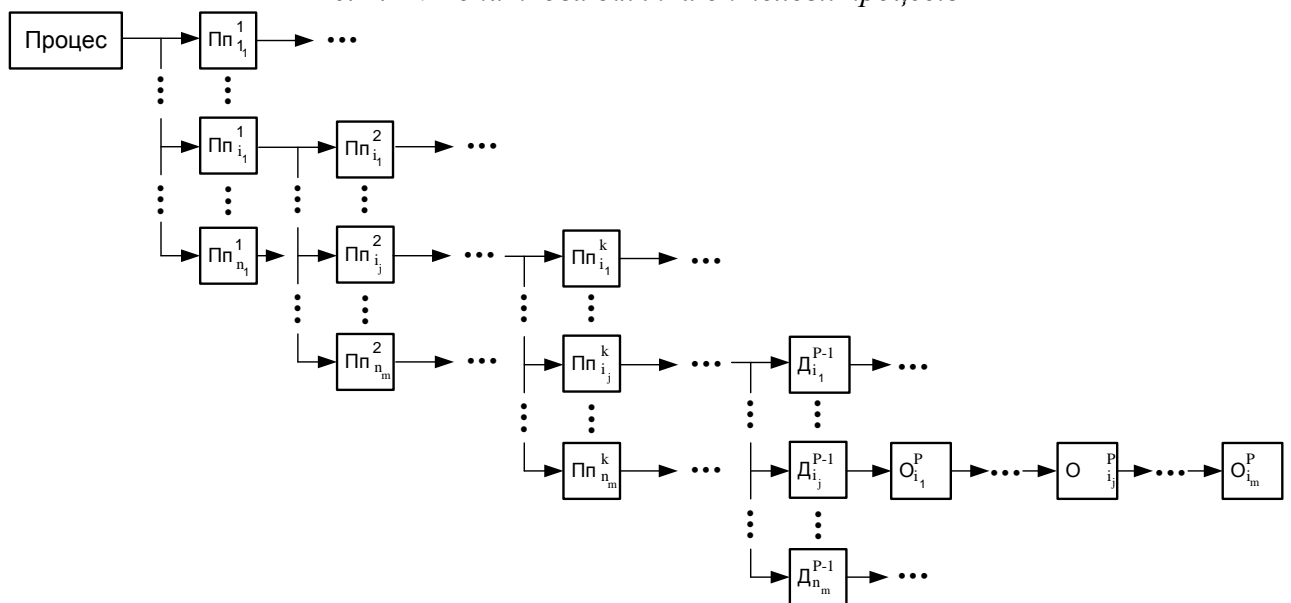


Рис. 2.13. Загальна схема онтології процесу

- онтологія об'єктів ПдО (в деяких публікаціях використовується термін «тезаурус»);
- онтологія процесів ПдО;
- онтологія додатків;
- онтологія класу задач ПрП;
- онтологія методів вирішення задач.

Відомі різні схеми конструювання (синтезу) онтологічних структур з предметно-проблемних знань [130, 131]. В даній роботі в онтологію ПдО, включені об'єкти і процеси (статистичні знання), які згруповані в окремі онтології, а в онтологію задач включені класи задач, методи їх вирішення та відповідні алгоритми. При такій схемі групування предметно-проблемних знань, останні строго розділені на статичні і динамічні. Якщо змінюється клас розв'язуваних задач, то онтології об'єктів і процесів вже готові до повторного використання, перепроектується тільки онтологія задач [92, 93].

В текстових описах (специфікаціях) цільових задач виділяються набір об'єктів і набір процесів (методів), необхідних і достатніх для виконання конкретних цільових задач. Можна виділити деяку уніфіковану (поповнювану) множину базових задач (типових фрагментів задач), на основі яких за допомогою певних логічних послідовностей конструюються більш складні задачі.

Онтологічні знання, що описують деякий ПрП, можна розділити на наступні компонентні знання: типи вхідних і вихідних даних, інструменти, оператори (людина або комп'ютерна програма) та операції (дії оператора або *Вирішувача задач*) [98].

Для реалізації підходу необхідно розробити уніфіковану мову представлення онтологічних знань та інструментальне середовище як набір спеціалізованих і універсальних базових операцій, які керують процесом вирішення. Необхідно також розробити *Вирішувач задач*, який здійснює вибір засобів і методів формування структури задачі на основі базових операцій.

Схема моделі онтології задач описується трійкою

$$O^3 = \langle OZ^{ПрП}, M, PЗ \rangle, \quad (2.12)$$

де  $OZ^{ПрП}$  – узагальнена задача проблемного простору, що складається з  $p$  задач, які, в свою чергу, складаються з  $w = \overline{1, \bar{W}}$  фрагментів кожна. Кожен фрагмент представлений процедурою, реалізованою на множині  $v = \overline{1, \bar{V}}$  операцій кожна. Крім того, задача

$$P^p = \langle D_{in}^p, R^p, C^p, D_{out}^p \rangle \quad (2.13)$$

визначається множинами вхідних даних  $D_{in}^p$ , вимог (умов, обмежень)  $R^p$ , контексту задачі  $C^p$  і вихідними даними (або метою вирішення задачі)  $D_{out}^p$ ;

$M$  – множина методів вирішення задач. Визначається як відображення

$$M^p : (D_{in}^p, R^p, C^p) \rightarrow D_{out}^p \quad (2.14)$$

компоненти якого визначені вище;

$BЗ$  – вирішувач задач.

На рисунку 2.14 представлена схема онтології задач.

Одна з переваг онтологічного підходу, зокрема, ієрархічного подання полягає в тому, що складна задача великої розмірності розбивається на послідовно розв'язувані групи задач малої розмірності. Онтологія задач в якості понять містить типи розв'язуваних задач, а відношення цієї онтології, як правило, специфікують декомпозицію задач на підзадачі.

### Алгоритм проектування онтології

На рисунку 2.14 представлена блок-схема алгоритму проектування онтології ПдО і онтології задач. Онтологія ПдО представлена онтологією об'єктів і онтологією процесів, для яких у блок-схемі передбачені окремі гілки. Передбачається, що множина функцій інтерпретації в моделях онтологій тотожна множині аксіом [2]. На рисунку прийняті наступні скорочення:

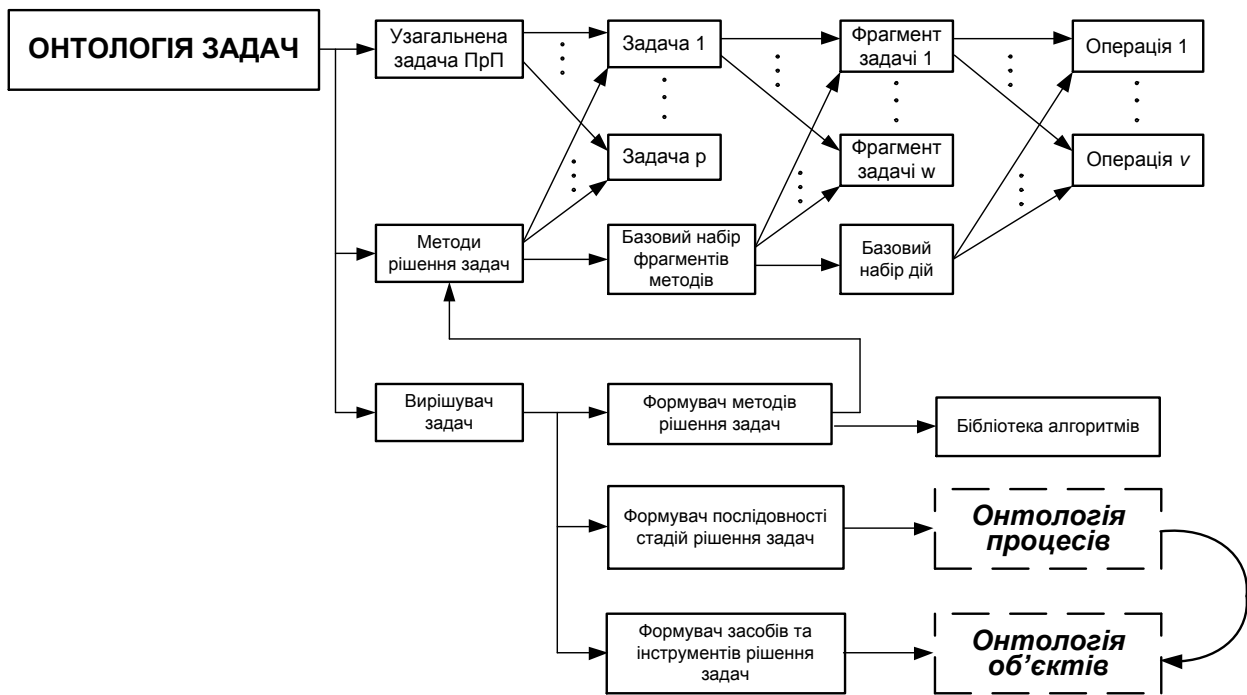


Рис. 2.14. Схема онтології задач

- ПрП – Т-О, Т-П – терміни-об'єкти і терміни-процеси;
- Т-З – списки задач, процедур і методів їх вирішення;
- $X^O, R^O, F^O$  – скінченні множини понять, відношень і функцій інтерпретації онтології об'єктів;
- $X^П, R^П, F^П$  – скінченні множини понять, відношень і функцій інтерпретації онтології процесів;
- $X^З, R^З, F^З$  – скінченні множини понять, відношень і функцій інтерпретації онтології задач;
- $OG^O, OG^П, OG^З$  – онтографи відповідно об'єктів, процесів і задач.

Нижче наведено опис алгоритму проектування онтології об'єктів, онтології процесів і онтології задач з посиланнями на пункти (номери блоків) на рисунку 2.15.

При виконанні постановки задачі на проектування онтології заданої ПдО і пов'язаного з нею ПрП систематизуються методи і засоби дослідження предметних і проблемних знань предметної області, а також виявляють джерела їх опису. Потім виконується добування та систематизація цих знань заданої ПдО і множини задач ПрП.

Алгоритм передбачає послідовну побудову онтології ПдО і онтології задач.

В результаті формуються списки Т-О, Т-П і їх властивостей з урахуванням синонімії і списки задач, процедур і методів їх вирішення. Слід відмітити, що в виконується формування проблемного простору заданої ПдО, у тому числі конкретизується інформаційний простір для вирішувача задач. Далі аналізується повнота сформованих списків і перехід до проектування онтології об'єктів, онтології процесів і онтології задач, яке складається з трьох блоків для кожної з них (О, П, З): побудова множини  $X$ ; побудова множини  $R$  (і побудова множини  $F$ ). Після кожного з них перевіряється повнота вичерпності списку відповідних термінів.

При побудові множини  $R$  формуються множини пар,  $i$ , відповідно до відношення «вище-нижче», аналізуються і встановлюються взаємозв'язки  $x_i^o R_k^o x_j^o$ ,  $x_i^n R_k^n x_j^n$  і  $x_i^z R_k^z x_j^z$  відповідно і будуються попередні онтографи для  $X^o$ ,  $X^n$  і  $X^z$ . При побудові множини  $F$  приписуються властивості для кожного  $X^o$ ,  $X^n$  і  $X^z$  зі списку властивостей Т-О, Т-П і Т-З відповідно, а також формуються функції інтерпретації  $\{f_i^o\}$  для кожного  $X^o$ ,  $\{f_i^n\}$  для кожного  $X^n$  і  $\{f_i^z\}$  для кожного  $X^z$  з визначень тлумачного словника (або з іншого відповідного джерела опису типових наборів задач і процедур і методів їх вирішення).

Далі виконується побудова кінцевих онтографів  $OG^o$  (п. 16),  $OG^n$  і  $OG^z$  за елементами множин понять  $X^o$ ,  $X^n$  і  $X^z$  відповідно; аналізуються на повноту і несуперечність онтографи  $OG^o$ ,  $OG^n$  і  $OG^z$  і їх формалізовані описи. Якщо задані критерії не виконані, то здійснюється повернення для реінжинірингу онтології об'єктів, онтології процесів або онтології задач відповідно.

На заключному етапі роботи алгоритму виконується аналіз на предмет побудови всіх онтологій  $O^o$ ,  $O^n$  і  $O^z$ . Якщо так – то END, інакше здійснюється повернення до п. 3.

Алгоритм проектування онтології ПдО і онтології задач являє собою послідовність кроків аналізу предметно-проблемних знань і синтезу онтологічних структур об'єктів, процесів і класів задач. Блоки 1-9 алгоритму описують процес аналізу, а блоки 10-17, блоки 18-25 та блоки 26-33 – процеси синтезу відповідно онтології об'єктів, онтології процесів і онтології задач. Алгоритм придатний як для ручного, так і для автоматизованого проектування зазначених онтологій.

Без сумніву зацікавленість представляє автоматизований спосіб проектування, оскільки навіть для порівняно невеликих предметних областей списки термінів-понять, термінів-процесів, термінів-відношень і термінів, що описують класи задач, можуть становити кілька тисяч записів. Складання вручну таких списків представляється проблемним. В розділі 2 описується інструментальний комплекс, програмні модулі якого реалізують, зокрема, представлений на рисунку 2.17 алгоритм, а в додатку А – приклада автоматизованого проектування онтології ПдО з кількістю термінів біля тисячі. Але, слід зазначити, існує багато додатків, в яких кількість термінів обмежена та допускають ручне проектування онтології. Методика проектування простої онтології наведена в додатку Б.



Рис. 2.15. Блок-схема алгоритму проектування онтології ПдО та онтології задач

## РОЗДІЛ 3. ЛІНГВІСТИЧНІ МОДЕЛІ ОБРОБКИ ПРИРОДНОМОВНИХ ТЕКСТІВ (ПМТ)

### 3.1. Загальний підхід до проблеми аналізу ПМТ

Ступінь розвитку та впровадження комп'ютерних технологій значною мірою визначається не стільки можливостями традиційної обчислювальної техніки (ОТ), скільки особливостями предметних областей, і успіх ефективної комп'ютеризації останніх суттєвим чином залежить від глибини дослідження та розуміння явища, що моделюється. У відповідності з цим етап узагальненого підходу до формування інформаційних технологій виключно з позицій потенційних можливостей ОТ поступово зміщується в сторону диференціації окремих напрямків ІТ, де на перший план виступають особливості (домену) конкретних ПДО. До такого напрямку слід зарахувати ІТ, які тим чи іншим чином пов'язані з мовним моделюванням поведінки людини, або системи, орієнтовані на опрацювання природномовної інформації. Причому під опрацюванням ПМО розуміється реалізація низки ІТ, кінцевою метою яких є комп'ютерне опрацювання знань як вищої форми розумової діяльності людини.

Розробка формальної теорії та методології комп'ютерного опрацювання знань складає одну з нагальних проблем у загальній теорії штучного інтелекту. Складність зазначеної проблеми визначається, зокрема, необхідністю залучення цілої низки наукових теорій (інформатики, математичної логіки, комп'ютерної та психологічної лінгвістики, нейрофізіології, нейрокібернетики, теорії семантичних мереж та ін.), які в сукупності, будучи застосованими до розв'язання проблеми формального представлення та обробки знань, склали б концептуально єдину (міждисциплінарну) формальну теорію. Складові вказаної теорії мають враховувати сутність етапів мовної та предметної обробки інформаційних об'єктів (для першого – ПМО, а для другого – вилучені з ПМО знання) і формального зв'язку між ними. Врешті необхідна така теорія, яка дозволяє розробити комплекс інформаційних технологій комп'ютерної обробки ПМО, представлення знань і комп'ютерної обробки знань.

При цьому виникає декілька проблем.

**Першою проблемою** (і однією з основних) є проблема аналізу текстової інформації, представленої на природній мові (морфологічний, синтаксичний, семантичний та логічний аналіз) з метою добування знань [1,6,7,40-46].

**Другою проблемою** є проблема проектування системи пошуку, обробки та добування знань, побудова її архітектури та розробка інструментарію для користувача [1-5, 24-27].

**Третя проблема** – це інтеграція знань із декількох предметних областей для забезпечення ефективності проведення досліджень міждисциплінарного характеру, використання вже існуючих алгоритмів, фактів, теоретичних положень та практичних рішень [1].

Третя проблема тісно зв'язана з ефективністю використання систем автоматичного пошуку доведення теорем у формальних логічних теоріях і подібними їй проблемами. Основу систем автоматичного пошуку доведення складають прувери – програми, що виконують доведення. Застосування прувера буде успішним тільки тоді, коли в його розпорядженні буде вся необхідна інформація для забезпечення процесу доведення. Наприклад, якщо пруверу необхідно довести теорему Лагранжа про ділення порядку скінченної групи на індекс її підгрупи, то пруверу недостатньо мати тільки аксіоматику теорії груп. Йому також знадобиться аксіоматика теорії ділення, а можливо і аксіоматика Пеано разом з деякими додатковими фактами з інших областей знань. Рішення цієї проблеми полегшується, якщо в розпорядженні прувера є інтегрована система, яка включає в себе необхідні відомості з інших областей знань. Тоді прувер сам знайде необхідну йому інформацію й використає її для успішного проведення доведення.

При вирішенні вказаних проблем важливою задачею є побудова природномовних лінгвістичних моделей та створення на їх основі ефективних лінгвістичних процесорів, які в сукупності з МОКС становлять мовно-онтологічну інформаційну систему.

Лінгвістичні моделі – це, по суті, фундаментальна науково-прикладна область досліджень, яка допомагає будувати системи, що опрацьовують ПМТ. Під обробкою ПМТ розуміється процес взаємодії *Система-ПМО-Користувач*, який включає в себе різні способи впливу на ПМТ, такі як аналіз, генерація, інтерпретація, трансформація, синтез та ін. Таке визначення лінгвістичних моделей, засноване на їх функціональному аспекті, є корисним з методологічної точки зору, що забезпечує проведення відповідної класифікації моделей за їх прагматичними ознаками, тобто за ціллю розробки та сферою використання. Виділимо наступні класи лінгвістичних моделей:

- 1) діалогові «запит-відповідь» чи інтерактивні моделі;
- 2) концептуально-формальні моделі;
- 3) концептуально-функціональні моделі;
- 4) когнітивні (семантико-контекстні) моделі.

Очевидно, найбільш складними є останні моделі. Саме вони забезпечують глибинне проникнення у поточний контекст і його трансформацію із збереженням смислу як всередині однієї моделі, так і між різними моделями.

Таким чином, об'єктом дослідження є природномовні тексти, представлені в електронному вигляді та вилучені з мережі Інтернет, монографій, підручників, науково-технічних документів тощо, які в сукупності складають вхідний лінгвістичний корпус текстів.

В існуючих ІС виділяють п'ять основних *рівнів розуміння* ПМТ [3].

**Перший рівень** характеризується схемою, яка показує, що будь-які відповіді на запитання система формує тільки на основі прямого змісту, що випливає з тексту. В лінгвістичному процесорі виконується морфологічний, синтаксичний і семантичний аналіз тексту та питань, що відносяться до нього. На виході ЛП отримуємо внутрішнє представлення тексту й питань, з

якими може працювати блок виведення. Використовуючи спеціальні процедури, цей блок формує відповіді. Іншими словами, вже розуміння на першому рівні потребує від ІС певних способів представлення даних і виведення на цих даних.

**Другий рівень.** На другому рівні додаються способи логічного виведення, засновані на інформації, що міститься в тексті. Це різні логіки тексту (часова, просторова, каузальна та ін.), що породжують інформацію, явно відсутню в тексті. Архітектура ІС, з допомогою якої може бути реалізований другий рівень розуміння, повинна мати додаткову базу знань, в якій зберігаються закономірності, що відносяться до часової структури подій, можливої їх просторової організації, каузальної залежності і т. д., а логічний блок – всі необхідні засоби для роботи з неklasичними логіками.

**Третій рівень.** До засобів другого рівня додаються правила поповнення тексту знаннями системи про середовище. Ці знання в ІС, як правило, носять логічний характер і фіксуються у вигляді сценаріїв чи процедур іншого типу. Архітектура ІС, в якій реалізується розуміння третього рівня, зовнішньо не відрізняється від архітектури ІС другого рівня. Але в логічному блоці повинні бути враховані засоби не тільки для чисто дедуктивного виведення, а й для виведення по сценаріям.

Три перераховані рівня розуміння повністю чи частково реалізовані практично в усіх діючих ІС.

**Четвертий рівень.** На цьому рівні відбувається зміна вмісту бази знань. Вона доповнюється фактами, відомими системі та вміщуваними в тих текстах, які введені в систему. Різні ІС відрізняються одна від іншої характером правил породження фактів із знань, опираючись на методи дедуктивного виведення та розпізнавання образів. Правила можуть бути засновані на принципах ймовірностей, розмитих виведеннях і т. д. Але в усіх випадках база знань виявляється апріорно неповною. В ІС виникають складності с пошуком відповідей на запити. Зокрема, в базах знань стає необхідним *немонотонне виведення*.

**П'ятий рівень.** На цьому рівні відбувається породження метафоричного знання. Правила породження знань метафоричного рівня, що використовується для цих цілей, представляють собою спеціальні процедури, що спираються на виведення по аналогії та асоціації. Відомі на цей час схеми виведення по аналогії використовують, як правило, діаграму Лейбніца, яка відображає тільки окремий випадок суджень по аналогії. Ще менше розроблені схеми асоціативних суджень [3].

Існують й інші інтерпретації феномену розуміння. Можливо, наприклад, оцінити рівень розуміння по спроможності системи до пояснення отриманого результату. Тут можливий не тільки рівень *пояснення*, коли система пояснює, що вона зробила, наприклад, на основі введеного в неї тексту, а й рівень *обґрунтування* (аргументації), коли система обґрунтовує свій результат, показуючи, що він не суперечить тій системі знань і даних, якими вона володіє. На відміну від пояснення обґрунтування завжди зв'язано



з сумою фактів і знань, які визначаються поточним моментом існування системи. І введений для розуміння текст в одних станах може бути сприйнятий системою як істинний, а в інших – як хибний. Існуючі ІС типу експертних систем, як правило, здатні давати пояснення і тільки частково обґрунтування [3].

Узагальнений критерій розуміння ПМТ науково-технічного профілю складається із здатності вирішувати на основі знань, що в них містяться, – прикладні задачі.

Особливості аналізу ПМО визначаються направленістю на формування структури понять, тобто, на автоматичне добування знань з текстів та їх прагматичну інтерпретацію в термінах прикладної задачі. При цьому текст розглядається як об'єкт різних рівнів аналізу: як знакова система, як граматична система і як система знань про світ (предметна область). Кожен рівень має свої особливості, свої способи вираження і, отже, допускає наявність специфічних методів обробки.

### *Методи та моделі аналізу ПМТ*

Якщо коротко характеризувати ідеї, які домінували останні десятиліття, то вони зводяться до таких положень.

Починаючи з 70-х років у дослідженнях природних і штучних (формальних) мов домінували спроби побудувати теорію, яка б охоплювала як природні, так і штучні мови.

Синтаксис викликає інтерес тільки в зв'язку з семантикою.

Метою синтаксису є характеристика синтаксичних категорій, з яких побудовані висловлювання.

Метою семантики є пояснення понять істини та логічного наслідування.

Ці ідеї привели до нових поглядів на такі поняття як граматики та значення виразів. А саме:

а) граматики – це інтерпретоване числення виразів мови (генеративна лінгвістика);

б) значення виразів обчислюються інтерпретатором (інтерпретаціонізм);

в) очевидна композиційність категорій синтаксису, семантики і прагматики (категоріальні граматики).

### **3.2. Етапи і моделі аналізу ПМТ**

Передісторія опрацювання природної мови (ОПМ) визначена багатьма факторами, із яких можна виділити два наступні: 1) спроби моделювання нейронів у вигляді логічних пристроїв; 2) зародження «інформаційної парадигми» – погляду на числа і на текст як на те, що представляє загальне поняття «інформації», яке має кількісні характеристики із усіма наслідками, що впливають, для математичної теорії інформації [4].

Власне історію ОПМ іноді представляють як таку, що складається з чотирьох етапів [4-8].

50–60-і роки. Ідея «інформації» привела до концепції «машинного перекладу» як встановлення тієї інформації, яка прихована в реченні природної мови (ПМ), та її вираження іншою – цільовою мовою. Мова інформаційного змісту названа «мовою-посередником». На цьому етапі введена ідея евристичного пошуку, а також мови програмування для вирішення задач ОПМ.

60–70-і роки. Рекомендується використовувати обмежені змістовні області для моделювання ОПМ і будувати як можна більше систем для обробки обмежених підмов, присвоївши цим системам статус «експертів» в певній предметній області, а потім об'єднувати всі ці системи в одну в рамках «великого експерта». Останній керує підключенням «малих»експертів на конкретному етапі обробки при вирішенні конкретної задачі – концепція «великого перемикачання». Здійснюється переклад ПМ на формальні мови (деяким із цих мов належать власні числення, типу числення предикатів формальної логіки; існують мови пошуку в базах даних. Проявлена зацікавленість в «когнітивній науці», названій потім когнітологією. Цей період іноді називають ерою «інженерного підходу»: основний наголос робився на вирішення конкретних (як правило, прикладних) задач, без спроб моделювати психологічну діяльність. Дослідження цієї епохи виявили цілу сукупність недоліків.

70–90-і роки. В центрі уваги знаходяться проблеми: співвіднесеності мовленнєвих актів, фокусів уваги та загальних (для ЕОМ і людини) презумпцій; використання «новаторського» потенціалу мови (зокрема, інтерпретація метафор); побудові прикладних систем, що розуміють ПМ (а не просто її обробляючих); обробки висловлювань, що описують просторові події; введення факторів «здорового глузду» в склад логічного виведення систем ОПМ, зокрема, врахування ступеня правдоподібності при інтерпретації висловлювань.

Однією з головних відмінних рис теперішнього етапу є «онтологічний підхід» до створення систем ОПМ. Він характеризується, зокрема широкомасштабним залученням до інтерпретації як мовних, так і предметних знань, зростаючою активністю біонічного напрямку досліджень при моделюванні «мовної свідомості» людини. Іншими словами, дослідження приймають явно виражений міждисциплінарний характер.

Теоретичні та практичні дослідження створили необхідний описовий апарат для комп'ютерного моделювання автоматичного аналізу текстів. Множину теоретичних підходів можна розділити на два основних напрямки: *формалізм* і *функціоналізм*. Формалізм стверджує, що мова є вродженим компонентом людського мислення, який може бути представлений у вигляді абстрактної моделі на метамові формальної граматики та не залежить від способів використання мови, а функціоналізм, навпаки, припускає, що будова мови визначається сферою його використання [9]. Дослідження в

формальній лінгвістиці можна також умовно розділити на два підходи: *побудова універсальної граматики*, справедливої для всіх існуючих мов світу, – генеративна модель, і *побудова формальної моделі*, яка найбільш повно охоплює всі множини граматичних явищ конкретної мови – модель «Смисл ⇔ Текст» [10].

В задачах автоматичної обробки ПМТ, як правило, використовуються концепції, розроблені в рамках останнього формалізму. З'єднавши дві вище зазначені парадигми формальної лінгвістики, програмні моделі є лише частковою реалізацією теоретичних досліджень.

Розглянемо вищезгадану модель формалізації семантики ПМТ «Смисл ⇔ Текст», на якій базується більшість сучасних теоретичних досліджень і практичних розробок для флективних мов [11, 12]. В основу моделі покладені такі принципові положення:

- синонімія мовних висловлювань виникає не тільки за рахунок лексичних синонімів – це можуть бути синтаксичні варіанти представлення одного значення слова, які по-різному орієнтують певну ситуацію щодо її учасників;

- комунікативна організація тексту включає: розподіл на тему і рему, виділення «старої» і «нової» інформації, виділення більш-менш важливих частин змісту;

- носіями смислу є не тільки слова, а й морфеми.

В моделі запропоновано мову семантичних множників для формалізації опису сутності і міри семантичних тотожностей лексично різних одиниць. Під множниками в даному випадку розуміються елементарні смисли, різні комбінації яких відповідають складним лексичним значенням слів. Значення слова при цьому є не простою сукупністю множників, а певним чином організованою структурою. В моделі прийняті такі рівні представлення тексту: семантичний, синтаксичний, морфологічний. Семантичне представлення включає в себе дві компоненти: семантичний граф і відомості про комунікативну організацію смислу. *Семантичний граф* – це орієнтований граф, вершини якого позначені символами сем (атомами смислу). Семі класифікуються на два типи: предикати та імена об'єктів. Логічні зв'язки та квантори трактуються як предикати. *Синтаксичний рівень* поділяється на глибинно-синтаксичний та поверхнево-синтаксичний. Перший включає в себе дерево синтаксичного підпорядкування, вузли якого – словоформи речення у вигляді абстрактних або узагальнених лексем з відповідною морфологічною інформацією (число, час, вид тощо), другий – максимально наближений до текстової форми представлення синтаксичних зв'язків у реченні. *Морфологічний рівень* також включає глибинне та поверхневе представлення.

Недоліком вищеописаної моделі є те, що вона не враховує закономірностей опису знань на знаковому рівні представлення тексту, що в свою чергу призводить до грубих помилок при семантичному аналізі тексту. Крім того, прагматичний аналіз зводиться лише до виявлення загальних знань про світ (предметну область). При цьому аналіз тексту, з точки зору

аналітика, полягає у виявленні таких семантичних ознак, які підпорядковані цільовій установці кінцевої (прагматичної) задачі.

В [12] запропоновано розвинений (порівняно з моделлю «Смисл  $\Leftrightarrow$  Текст») підхід до розробки методів розпізнавання та добування знань з ПМТ (які належать до деякої ПдО) і який ґрунтується на таких концептуальних положеннях:

- вхідний ПМТ – зв'язний текст (тобто дискурс);
- зв'язність дискурсу забезпечується графемними засобами оформлення тексту, лінгвістичними (граматичними узгодженнями, анафоричними посиланнями і т. п.) і екстралінгвістичними (наприклад, часовими, причинно-наслідковими зв'язками, які відповідають певній ПдО) засобами;

- всі ці засоби є інструментом кодування знань про світ (ПдО);
- як елементи реального чи абстрактного світу виступають його об'єкти, відображені в тексті у формі природномовних понять, відношень між ними та мовні характеристики понять і відношень;

- формування об'єктів знань про світ може бути різним і залежить від цільової установки прикладної задачі, проблемної області та логічної картини світу як носія інформації, так і того, хто її обробляє.

Особливості аналізу ПМТ визначаються спрямованістю на формування понятійної структури, тобто на автоматичне добування знань з текстів та їх прагматичну інтерпретацію в термінах прикладної задачі. При цьому текст розглядається як об'єкт різних рівнів аналізу: як знакова система, як граматична система, система знань про світ, джерело вихідних даних для вирішення прикладних задач. Кожен рівень має свої особливості, свої засоби вираження, а, отже, передбачає наявність специфічних методів обробки.

### **3.3. Загальні принципи та нормаль на постановка задачі аналізу ПМТ**

Дослідження процесів інтерпретації та розуміння мовних висловлювань, має як теоретичний, так і прикладний інтерес. Роботи в галузі автоматичного аналізу тексту та автоматичного вирішення задач, сформульованих природною мовою (ПМ) (або мовою, близькою до неї), показали актуальність побудови інтерпретаційної теорії мови. Поняття «інтерпретація» з самого початку лежало в основі загально-лінгвістичних теорій, а також в основі логічних досліджень. Інтерпретаційний підхід представлений як в різноманітних галузях чисто лінгвістичного аналізу (в теоріях формальних граматики, в «теорії мовленнєвих актів» і т. ін.), так і в дослідженнях по штучному інтелекту. Згідно з інтерпретацією в основі володіння мовою та її використанням лежить один і той же інтерпретуючий механізм, що обслуговує різні *сфери мовленнєвої діяльності* і при цьому використовує різні види знань. Серед цих сфер – мовлення, розуміння, редагування, коментування, перефразування, міркування, комунікація,

аргументація, навчання, переклад та ін. Сама ж інтерпретація, через яку і визначаються вказані сфери, представляє собою отримання на основі одного вихідного об'єкта (об'єкта, що інтерпретується) іншого, відмінного об'єкта, який допускається інтерпретатором як рівнозначного вхідному на фоні конкретної ситуації, набору знань [4]. Самі ж знання не входять до структури мови безпосередньо, а «залучаються» до інтерпретації, і тільки опосередковано визначають результат інтерпретації мовних висловлювань.

Однією з основних процедур обробки ПМТ є процедура розпізнавання, зокрема когнітивного розпізнавання. Під *когнітивним розпізнаванням* ПМТ розуміється процес формалізації добування й представлення знань предметної області, що містяться в ПМТ. Входом процедури розпізнавання є ПМТ, а виходом (результатом) – формально-логічне представлення. Воно представляє собою формалізоване представлення знань про світ (предметну область), відображене в певному ПМТ [12].

Процес розпізнавання та добування знань з ПМТ базується на комп'ютерному моделюванні інтелектуальних функцій людини, а саме – на моделюванні процесу розуміння людиною ПМТ. При цьому термін *розуміння* визначається за допомогою таких критеріїв: вміння переказати «своїми» словами, тобто іншими (лексичними, синтаксичними) засобами передати зміст вхідного тексту, вміння відповісти на запитання стосовно певного тексту та ін. Процедура розпізнавання базується на засобах формалізації (тобто розробки онтологічних моделей) знань про певну мову і знань про певну ПдО. Оскільки процедури розпізнавання та розуміння є базовими при лінгвістичній обробці ПМТ, розглянемо їх більш детально з методологічної точки зору.

Зазначимо, що когнітивне розпізнавання ПМТ має свої особливості, а саме:

- позамовні ситуації, описані в текстах визначається лише знаннями з певної ПдО;
- засоби вербалізації цих знань орієнтовані на певний рівень фахової підготовки;
- механізми взаємодії знань у тексті з когнітивною картиною світу ґрунтуються на моделі представлення людиною знань про певну мову (мовна картина світу) та знаннями про фрагменти реальної дійсності (бази знань предметних областей).

### *Прикладні аспекти лінгвістичного аналізу ПМТ*

Архітектура сучасних знання-орієнтованих інформаційних систем з природномовним представленням і обробкою знань включає онтологічну складову експліцитно, яку в загальному вигляді можна інтерпретувати як концептуальну базу знань. Така база знань представляється в вигляді орієнтованого графа, вершинами якого є фрейми, що описують концепти, а дугами – множина концептуальних відношень, що зв'язують між собою концепти. Іншою важливою особливістю вказаної архітектури є розділення

та окрема обробка семантики першого та другого ступеня [14], що в загальному випадку означає розділення внутрішньомовного та зовнішньомовного процесингу [7] і перехід до формально-логічного представлення вхідного тексту.

### *Графемно-морфологічний аналіз*

Як початковий етап лінгвістичного аналізу використовується графемно-морфологічний аналіз (ГМА). При виконанні графемного аналізу текст розглядається як різновидність знакової системи. Задачею даного етапу аналізу є дослідження властивостей одиниць мови й правил їх поєднання в аспекті їх знакової природи. В існуючих моделях графемний аналіз реалізовано тільки як засіб розв'язання протиріч аналізу текстових одиниць, що надходять на обробку морфологічного аналізатора і окреслюється відділенням аббревіатур, скорочень та деяких інших класів лексем. Це, в свою чергу, породжує протиріччя на етапі семантичного аналізу тексту, що призводить до помилок семантичного характеру, уникати яких користувачу пропонується самому, вручну маркувати текст.

Призначення *графемного аналізу* – побудова моделі графемної структури вхідного тексту, в якій виділені і зв'язані відношеннями (де це можливо) такі змістовні одиниці тексту як фрагмент, речення та лексема. Наприклад, для лексем: виділяються класи лексем, які відрізняються своєю графемною структурою і виконують різні функції в тексті. Крім того, аналізуються закономірності сполучуваності деяких лексичних одиниць, які уже на цьому етапі дозволяють об'єднувати декілька лексем в одну на тій основі, що вони виконують в тексті єдину функцію [13].

Внаслідок графемного аналізу (точніше, інтерпретації бази знань графічного зображення ПМО з вхідним текстом) текст перетворюється у впорядкований список складників ПМО.

Під *морфологічним аналізом* розуміють повну обробку послідовності (без будь-якого зв'язку з контекстом) словоформ (див. алгоритм морфологічного аналізу), сформованих на підетапі графемного аналізу. Внаслідок такої обробки кожна словоформа замінюється граматичною інформацією – ланцюжком символів, що визначають усі ті властивості даної словоформи, необхідні для подальшого аналізу. Морфологічна інформація сукупності словоформ, в свою чергу, використовується на наступному етапі створення програмної моделі ЛП – під час синтаксичного аналізу, внаслідок якого встановлюються зв'язки між усіма словоформами тексту та між частинами складних речень.

Вхідними даними процедури морфологічного аналізу є графемна структура тексту, база знань морфології ПМ та лексикографічна база даних. ЛБД включає словники лексем та словозмінну і словотвірну моделі вхідної мови, а також окремі таблиці для всіх частин мови. До кожної лексеми в

таблиці приєднуються, окрім традиційних морфологічних характеристик, набори синтаксичних і семантичних ознак [14-15].

Крім того, існують окремі таблиці відмінкових закінчень (для змінних частин мови) для формування слів форм лексеми. Всі лексичні одиниці в таблиці відповідним чином проіндексовані та мають однакове інтерпретаційне значення на всіх етапах лінгвістичного аналізу.

### *Синтаксичний аналіз*

Синтаксичним аналізом (parsing) називається процес структуризації лінійної репрезентації відповідно до заданої граматики. Лінійною репрезентацією речення природної мови будемо називати ланцюжок елементів, де кожен елемент є мінімальною синтаксичною одиницею. Мінімальна синтаксична одиниця може бути словоформою або оператором з певним набором характеристик. *Оператором* називається розділовий знак або сурядний сполучник. Процесом структуризації лінійної репрезентації речення називається побудова орієнтованого графа синтагм і орієнтованого графа сегментів. *Синтагма* визначає бінарне синтаксичне відношення вигляду  $R(A, B)$ , де  $A$  і  $B$  – словоформи, а  $R$  – тип синтаксичного відношення, яке відповідає імені синтагми, причому  $A$  керує  $B$ . Вузлами графа синтагм є термінальні одиниці. Зв'язаність не є обов'язковою умовою такого графа, оскільки синтагми спираються тільки на морфологічні репрезентації словоформи, лінійний порядок речення і, в деяких випадках, на примітивну модель управління.

Сегмент можна визначити як частину речення, яка виділена в тексті знаками пунктуації і описує окрему ситуацію. Кожен такий сегмент має як вершину явний предикат, який являє собою в більшості випадків фінітну форму дієслова або прихований предикат, який може бути виражений дієприслівником, дієприкметником або ім'ям з семантичною характеристикою дії; кожен такий предикат і задає ситуацію коли вузлами графа сегментів є нетермінальні одиниці.

В зарубіжній лінгвістичній традиції поняття «сегмент» еквівалентне терміну «клауза». Сегментація речення є першим етапом повного синтаксичного аналізу. Метою сегментації є виділення і класифікація сегментів у складі складного речення. Задачею другого етапу є усунення граматичної омонімії морфологічного рівня і побудова словосполучень. Об'єктом розпізнавання цього рівня є закономірності взаємодії лексем у межах мовної синтагми. Вхідними даними є результати попередніх етапів розпізнавання та база знань синтаксису ПМ. В ній закладені моделі синтаксичних правил, які визначають ознаки синтаксичного з'єднання лексем у словосполучення [47].

Серед мовних засобів синтаксичного аналізу можна виділити наступні:

1. **Словозмінні морфологічні засоби.** Для мов з розвиненою морфологією, якою є українська та російська мови, – це основний спосіб матеріалізації синтаксичних зв'язків. Показником морфологічної залежності

служить *флексія*, оскільки граматичні словоформи зазвичай приписані флексії, що дозволяє в деяких випадках виявити залежність між двома словоформами, відсутніми у словнику. В українській мові словоформи багатьох слів омонімічні (наприклад, це може бути омонімія числа і відмінка), що створює певні труднощі у процесі аналізу. Неоднозначність графем у ході автоматичного синтаксичного аналізу іноді призводить до виникнення синтаксичної омонімії і побудови альтернативного синтаксичного варіанту (графа синтагм). Взаємовплив двох рівнів аналізу морфологічного і синтаксичного полягає в тому, що графема, будучи ефективним засобом пошуку морфологічної залежності, а точніше реалізації синтаксичного відношення, може бути однозначно проінтерпретована тільки внаслідок фіксації цього відношення.

2. **Селективні ознаки:** Класифікуючі (селективні) ознаки приписуються лексемам у граматичному словнику. Найбільш важливою для синтаксису є класифікація лексем за категоріальними ознаками: іменник, дієслово, прикметник та ін. *Категоріальні ознаки задають потенційних учасників синтаксичного зв'язку і визначають у більшості випадків вершину в структурі, а також зумовлюють поняття керування та узгодження.*

3. **Службові слова:** прийменники, сполучники, частки та ін. – засоби, які слугують опорними точками аналізу. Так, наприклад, сполучник може бути використаний для визначення поверхневого типу сегмента, прийменники формують актант дієслова.

4. **Розділові знаки (оператори).** Кома, тире, крапка, знак питання та ін.

В першу чергу оператори визначають межі як сегментів, так і всього речення. Тире є вираженням еліптованого елемента речення і часто використовується у стратегіях пошуку неморфологічного предиката. Аналіз біфункціональності оператора (коли, наприклад, оператор є одночасно і правою межею сегмента, і оператором створення слів) – одне з найважчих завдань сегментації, яка задає рекурсивний характер, як граматичних стратегій аналізу, так і методів програмної реалізації.

5. **Порядок слів.** Лінійне розташування слів у реченні разом із селективними ознаками має домінуюче значення в проектуванні синтаксичних аналізаторів мов з бідною морфологією (англійська). У багатьох системах англійського синтаксису порядок слів задає напрям пошуку господаря або слуги для кожного класу лексем і типу зв'язку. Для флективних мов цей засіб аналізу має другорядне значення, хоча і застосовується у ряді випадків для встановлення синтаксичних зв'язків або оцінки омонімічних структур речення.

Незважаючи на вільний порядок слів в українській (російській) мові, деякі синтаксичні залежності мають обов'язковим критерієм виділення жорсткого лінійного порядку: генітивне визначення має слідувати за словом («ніжка стол-у», «син батьк-а»); прийменник передує іменнику («на стол-і», «у батьк-а»). Лінійний порядок речення в автоматичному синтаксичному аналізі використовується як показник найбільш вірогідного напрямку



пошуку слуги або господаря, і лише в окремих випадках – як обов’язковий критерій встановлення синтаксичної залежності.

Для перевірки і встановлення можливого синтаксичного відношення використовуються граматичні правила і стратегії. Кожне граматичне правило встановлює один тип синтаксичного відношення  $R(A, B)$  між двома одиницями аналізу й однозначно задає вершину. Кількість типів синтаксичних відношень, а також їх назви залежать від прикладної моделі і конкретної реалізації системи синтаксичного розбору.

Набір *універсальних синтаксичних відношень* для російської мови (може бути використаний і при аналізі української мови) наведений в багатьох теоретичних роботах, наприклад [16]: прийменникове (іменник і прийменник, який ним керує), визначальне (іменник і його узгоджене визначення), посесивне (іменник і його неузгоджене визначення), суб’єктне (присудок і підмет). В ролі одиниць аналізу на місці  $A$  і  $B$ , де  $A$  – вершина, а  $B$  – залежна одиниця аналізу, можуть виступати як окремі словоформи, так і цілі групи (фразові складові). Заповнення  $A$  і  $B$  багато в чому залежить від синтаксичного апарату, прийнятого в аналізаторі для опису структури. *Ідеальне граматичне правило* в автоматичному синтаксичному аналізі характеризується наступними критеріями:

- 1) описує тільки один тип синтаксичного відношення;
- 2) однонаправленість аналізу, тобто залежне  $B$  може знаходитися тільки зліва або тільки справа від вершини  $A$ ;
- 3) не містить рекурсивних викликів або викликів інших правил;
- 4) обробляє тільки компактно розташовані одиниці аналізу;
- 5) результат не залежить від порядку застосування правил.

Використання граматичних правил задає прозорість архітектури процесора і забезпечує стійкість системи до змін. Перераховані критерії не є строгими, але наближають правило до його ідеальної форми.

Результатом синтаксичного аналізу є побудований орієнтований граф синтагм і орієнтований граф сегментів речення українською або російською мовою.

### *Семантичний аналіз*

Під автоматичним семантичним аналізом будемо розуміти сукупність методів, що реалізовані у вигляді однозначної формальної процедури, за допомогою якої можливо з достатньою точністю представити зміст будь-якого висловлювання природною мовою у вигляді послідовності символів, які утворюють деяку формальну систему.

Найбільш простим і універсальним засобом представлення змісту деякого тексту в комп’ютерній системі є семантична мережа. У загальному випадку вона є орієнтований граф, вершини якого позначають сутності (об’єкти), а дуги – відношення (зв’язки) між ними. Імена вершин і дуг зазвичай збігаються з іменами відповідних сутностей і відношень, що використовуються у природній мові. Дуга і дві зв’язані з нею вершини

представляють мінімальну значиму інформацію – факт наявності зв'язку певного типу між відповідними об'єктами.

Виділяють два підетапи в проведенні семантичного аналізу. Перший підетап названо поверхневим семантичним аналізом. На цьому підетапі здійснюється усунення синтаксичної омонімії (якщо вона має місце), формування понять, відношень та їх характеристик (властивостей) в межах одного речення. Крім того, на цьому етапі проходить заміщення анафоричних зв'язків (зокрема, коли замість поняття підставляється займенник), узагальнення понять і відношень, пропусків понять і відношень, що притаманні природній мові. На другому підетапі семантичного аналізу будується власне онтолого-семантична структура тексту.

Згідно з [17], із множини слів російської мови (для української мови також) за семантичними ознаками можна виділити наступні категорії: слова з моделлю управління (предикати) та слова без моделі управління (поняття, характеристики та їх значення, чисельники).

Найбільш важливою є обробка предикатів дії та їх можливих актантів – понять, характеристик, їх значень, а також чисельників. Обробка функціональних і пустих предикатів може бути зведена до модифікації фрагментів семантичної мережі, побудованих для залежних від них предикатів дії.

Кожен предикат дії має одну або декілька моделей керування. Модель керування накладає синтаксичні і семантичні умови на можливі актанти (аргументи) даного предиката і відмічає їх семантичні ролі щодо предиката. У спрощеному виді модель керування можна представити як таблицю, кожна строчка в якій визначає один з можливих актантів. Ця строчка містить: прийменник (якщо є), частину мови та відмінок актанту, його семантичне метاپоняття (категорію), семантичну роль в предикаті й ознаку обов'язковості даної ролі. Аналіз моделей керування дозволяє природним чином перейти від синтаксичного дерева залежностей до фрагмента семантичної мережі. Для цього у простому випадку достатньо: 1) для кожного предиката відшукати в дереві залежностей всі присутні актанти; 2) створити по одній вершині на кожен актант і ще одну – для самого предиката; 3) провести від вершини-предиката дугу до кожної вершини-актанту; при цьому назва дуги (її семантична роль) вибирається з моделі керування залежно від актанту.

Якщо модель керування не допускає існування актанту ні в одній семантичній ролі, залишається дві можливості: або дана текстова одиниця насправді є стосовно предиката не актантом, або має місце помилка синтаксичної структури. Правильно розрізняти ці два випадки вдасться не завжди, проте можна вказати умову, за якої з великою вірогідністю присутня саме помилка синтаксичної структури. Мається на увазі ситуація, коли актант задовольняє морфологічним умовам моделі керування, а семантичним – ні. Оскільки синтаксичний аналізатор при групуванні не перевіряє семантичні ознаки, він не може самостійно виявляти помилки такого типу.

Таким чином, якщо помилка виявлена на етапі семантичного аналізу, необхідно повернутися назад до синтаксичного аналізу і спробувати знайти інший варіант синтаксичної структури, який би не містив невірною актанту у піддереві залежностей даного предиката.

Семантичний аналіз здатен виявляти деякі помилки синтаксичної структури (наприклад, пов'язані з помилковою моделлю керування). Таким чином, існування зворотного зв'язку між семантичним і синтаксичним аналізом – необхідна властивість лінгвістичного процесора, тому доцільно об'єднати ці два етапи аналізу та виконувати їх спільно в одному аналітичному блоці.

Як зазначалося раніше, існує зворотний зв'язок між семантичним і синтаксичним аналізом, тому можливо об'єднати ці дві процедури в одному блоці. Кінцевим завданням процедури синтаксичного аналізу є представлення кожного речення заданого природномовного тексту у вигляді синтаксичного дерева (лексем речення з семантико-синтаксичними відношеннями між ними).

Зв'язування слів у реченні відбувається поступово – від словосполучення до групи зв'язаних слів і, зрештою, до об'єднання всіх груп у реченні в одну структуру. Для встановлення зв'язку між окремими словами використовуються природномовні засоби вираження семантичних та синтаксичних відношень. У флективних мовах такими засобами є змінні частини повнозначних слів та службові слова.

Такі сегменти словосполучень, які кодують відношення між повнозначними словами, називаються *синтаксичними визначниками* [18]. Оскільки одному синтаксичному визначнику може відповідати декілька синтаксичних відношень, для однозначності визначення зв'язків між словами вводиться поняття кореляторів [18], які додатково враховують семантичні ознаки між словами.

Вихідні дані для блоку семантичного аналізу:

- результати попередніх етапів аналізу ПМТ (графемного і морфологічного);
- словник основ (містить основи слів та їх семантичні ознаки);
- список всіх можливих флексій слів;
- база даних з визначниками (містить синтаксичні визначники та списки кореляторів для кожного з них);
- база даних з кореляторами (кожен корелятор складається з назви відношення та списку пар семантичних ознак, між якими це відношення може існувати).

Розглянемо основні етапи роботи блоку синтактико-семантичного аналізу ПМТ.

*Перший етап.* Використовуючи словник основ та список флексій, у кожному слові речення виділяється його незмінна (основа) та флективна складові. Проводиться класифікація слів за семантичними ознаками відповідних основ у словнику. При цьому виникає задача можливої

неоднозначності виділення основи слова та визначення його семантичної ознаки. Можливими шляхами її вирішення є врахування характеристики слів, що стоять поруч у реченні, та розширення словника основ додатковими характеристиками.

Незмінна та флективна частини слова, його семантична ознака можуть бути отримані і на попередніх етапах аналізу (етап морфологічного аналізу). Синтаксичні характеристики в таблицях ЛБД, наприклад для дієслів, включають в себе лексико-семантичні валентності дієслів, що визначають ознаки найбільш вірогідного оточення, та ознаки семантичних інтерпретацій, які на основі розпізнаних синтаксичних правил визначають семантичні відношення між словами у словосполученні. Якщо речення складається з декількох синтагм, то завдяки інтерпретації певної частини бази знань синтаксису відбувається їх об'єднання в єдине синтаксичне дерево з відповідними семантико-синтаксичними відношеннями.

*Другий етап.* Зв'язування слів у реченні доцільно починати з словосполучення, що визначає голове відношення (відношення між підметом і присудком) в цьому реченні. У випадку, коли таке словосполучення встановити неможливо, речення аналізується зліва направо, починаючи з перших повнозначних слів. Для вибраного словосполучення формується синтаксичний визначник, який складається з службових слів та флективних частин повнозначних слів словосполучення. Наприклад, для виразу «права та свободи» таким визначником буде конструкція типу «-а та -и». Якщо сформований визначник існує в базі даних визначників, йому буде відповідати список кореляторів. Тоді, враховуючи семантичні ознаки слів у словосполученні, в базі даних з кореляторами знаходиться потрібний корелятор, який і встановить зв'язок між словами. Однозначність визначення такого зв'язку забезпечується тим, що для окремого визначника множини пар семантичних ознак для кореляторів з його списку не перетинаються.

Далі до словосполучення поступово приєднуються прилеглі повнозначні слова речення, шляхом встановлення зв'язку між новим словом та одним із слів опрацьованої частини речення. Таким чином створюється група зв'язаних слів. Важливим є вибір слова з групи зв'язаних слів, яке буде зв'язуватись з наступними словами. Це має бути або слово з головним відношенням, або останнє слово групи. У випадку, коли неможливо встановити зв'язок між новим словом та словами групи, створюється нова група зв'язаних слів. В кінці аналізу необхідно спробувати поєднати всі створені групи в одну, яка відобразить структуру зв'язків між всіма словами речення.

Неможливість встановлення зв'язку між окремими групами в реченні та їх об'єднання свідчить або про складне речення, частини якого не зв'язані (або зв'язані неявно) між собою, або про некоректні зв'язки між словами у групах. Для уникнення проблеми некоректного зв'язування слів в принципі необхідно проводити додатковий аналіз можливих зв'язків кожного наступного слова з словами групи та вибирати найбільш вірогідний, або ж

розглядати всі можливі варіанти зв'язків (однак, слід враховувати зростання кількості таких варіантів для кожного наступного слова).

Результатом роботи блоку семантичного аналізу є поняттєва (логіко-семантична) структура ПМТ. Для побудови такої структури спочатку необхідно сформулювати систему понять, відношень та визначити їх характеристики.

Однією з важливих є задача анафоричних зв'язків або заміни займенників в тексті на відповідні поняття (як правило, іменники). Заміщенню підлягають деякі особові, відносні, вказівні, присвійні та зворотні займенники. Алгоритм заміщення будується на основі аналізу закономірностей (відображених в базі знань синтаксису ПМ) вживання займенників в ПМ і відрізняється в залежності від типу та граматичних характеристик займенника.

Із практики відомо, що поняттям, якому відповідає займенник, є узгоджене з ним за граматичними характеристиками найближче повнозначне слово, що стоїть попереду, або слово, що входить до ядра даного чи попереднього речення. Тому вхідними даними для проведення заміни займенників є граматичні характеристики слів (отримані внаслідок морфологічного аналізу) та синтактико-семантичні зв'язки між ними (результат роботи синтаксичного блоку). Заміщення займенників дозволяє отримати додаткові зв'язки між відповідними поняттями в реченні та в усьому тексті, тому після цього доцільно провести повторний синтаксичний аналіз окремих речень.

Наступною задачею семантичного аналізу є виділення термінів природномовного об'єкта, з яких потім формуються поняття. Розглядаються терміни-іменники як однослівні, так і словосполучення. При автоматичному виділенні термінів аналізуються такі дані:

- семантичні ознаки слів ПМТ;
- синтактико-семантичні відношення між словами в реченні (для термінів, виражених словосполученням);
- віддаленість терміну від ядра фрази;
- частота використання термін в ПМТ;
- лінгвістична складова ПдО, до якої належить ПМТ.

Якість виділення таким чином термінів значною мірою залежить від якості роботи синтаксичного блоку, а саме від коректності класифікації слів за семантичними ознаками та встановлених зв'язків між словами.

### *Логіко-семантичні відношення в ПМТ*

Для виявлення логіко-семантичних відношень в ПМТ використовуються спеціальні діагностичні конструкції фраз та встановлені на етапі синтаксичного аналізу семантично-синтаксичні зв'язки між словами. Основні типи відношень логіко-семантичної структури наступні: відношення тотожності, включення, перетину та диз'юнкції, ієрархічні відношення, відношення агрегації та функціональні відношення [19].

*Ієрархічні відношення* належать до області, що може бути виражена семантикою дієслова «бути», і поділяються на три підвиди: рід – вид; ознака – значення; інваріант – варіант. Відношення *рід–вид*, в свою чергу, поділяються на структурні, функціональні та семантичні.

Для розпізнавання *родо-видових* відношень використовують наступні основні конструкції фрази:

– «*A* належить до *B*», «*A* є *B* з ознаками», «рядом з *A* є *B*», «*A* належить до класу/сімейства *B*» – для пошуку родового поняття *B* і видового поняття *A*;

– «*A* має наступні види:  $B_1, B_2, \dots$ », «до видів *A* належать  $B_1, B_2, \dots$ » – для пошуку видових понять *B* і родового поняття *A*.

Діагностичні конструкції відношення ознака – значення наступні (для пошуку *A* – поняття-сутності та *B* – імені сутності) наступні: «*A* – *B*», «*A* має ім'я *B*», «*A* іменується *B*», «*A* називається *B*».

Відношення інваріант – варіант знаходять за допомогою фраз «до варіантів *A* належать  $B_1, B_2, \dots$ » та «інваріантом *A* є *B*».

*Відношення агрегації* належать до області, що може бути виражена семантикою дієслова «мати». Виділяють чотири підвиди: 1) ціле – частина (компонент); 2) об'єкт – простір реалізації об'єкта (локалізації, позиції); 3) об'єкт – властивості/ознаки; 4) рівень – одиниця рівня.

Основні діагностичні конструкції для пошуку відношень агрегації:

Для підвиду 1):

– для цілого *B* і частини *A*:

«*A* входить до складу *B*», «*A* складає частину *B*», «*A* є частиною *B*»;

– для частини *B* і компонента *A*:

«*A* складається з *B*-х», «*A* містить в собі *B* як частина», «*A* містить у своєму складі *B* як частину».

Для підвиду 2):

– «місце *A* – *B*», «*A* відбувається в *B*», «*A* знаходиться в *B*», «*A* входить в *B*».

Для підвиду 3):

– «*A* має властивість *B*, має *B* як суттєву ознаку», «*A* характеризується наявністю *B*», «для *A* характерно *B*»;

Для підвиду 4):

– «*A* належить до рівня *B*», «*A* – одиниця рівня *B*», «*A* розглядається на рівні *B*».

*Функціональні відношення* – це відношення, які належать до сфери опису процесів. Таке відношення представляє собою багатомісний предикат, що описує деякий процес. У цьому предикаті виділяють наступні змінні: *A* – суб'єкт деякої сутності; *B* – інструмент, спосіб, метод або алгоритм виконання сукупності операцій процесу; *C* – початковий об'єкт або вихідні дані; *D* – кінцевий об'єкт або кінцеві дані; *E* – подія, що активує операцію.

Основні діагностичні вирази:

– для виявлення суб'єкта: «*A* здійснює *B*»;

- для інструмента, способу, методу або алгоритму – тут можуть бути вказані різнорідні відповідні терміни;
- для виявлення початкового об'єкта: «*A* виконується над *B*», «*A* застосовується до *B*», «*A* змінює *B*», «*AB*-і»;
- для виявлення кінцевого об'єкта: «внаслідок *A* виходить *B*», «внаслідок *A* утворюється *B*», «*A* призводить до *B*», «*A* перетворює *B*<sub>1</sub> в *B*<sub>2</sub>».

### **3.4. Методологічні засади проектування ОКІС з обробкою предметно-орієнтованих знань**

До онтологічних аспектів відноситься широке коло питань: від формальної побудови та опису комп'ютерних онтологій до області їх застосування. Головний вектор досліджень спрямований на формалізацію етапів побудови, структурування та представлення знань предметної області та інтегрованого з нею проблемного простору. У свою чергу, ефективна реалізація зазначених етапів і отримання кінцевого результату (у вигляді бази знань ПдО) неможливо без проведення системно-онтологічного аналізу заданої ПдО [2].

*Онтологічний підхід* (ОнП) до проектування інформаційної та архітектурної компоненти комп'ютерної системи виник як міждисциплінарна проблема формування, представлення, обробки та аналізу знань, моделі яких описують структуру і взаємозв'язки об'єктів відповідних предметних областей.

Будемо вважати, що ОКІС з обробкою знань, що містяться в ПМО, спроектована на основі онтологічного підходу, якщо вона має наступні характерні риси:

1. Комп'ютерні онтології (КО), забезпечують ефективну машинну обробку мовних і предметних знань (в [1] такі онтології названі відповідно мовно-онтологічної картиною світу і онтологією ПдО).

2. На відміну від емпіричного підходу, що є основою «творчості та мудрості», системно-онтологічний підхід [2] передбачає строгу (наскільки це можливо) систематизацію знань будь-якого рівня, у тому числі категоріального. Останній представляється категоріальною ґраткою, або онтологією верхнього рівня. Її проектування входить в загальний алгоритм синтезу ОКІС.

1) Архітектурно-технологічні особливості ОКІС:

- онтолого-керована архітектура (ontology-driven architecture – ODA) характеризується високим рівнем формалізації подання О ПдО і ефективними механізмами онтолого-керування, у тому числі підтримують розвиток системи;

- високий ступінь інтеграції міждисциплінарних знань.

2) Використання засобів підтримки автоматизованої побудови онтологій ПдО (методика, технологія та програмно-апаратна реалізація).

3) Прикладна спрямованість і сильний взаємозв'язок технології обробки інформації з архітектурно-структурною організацією ОКІС.

3. Проектування ОКІС виконується на основі принципів, методів і механізмів онто-логічного підходу (ОП) [1].

4. Функціонування ОКІС у двох режимах: накопичення предметних онтологічних знань та їх обробки.

### Онтологічний підхід до розробки архітектури ОКІС

Нижче викладаються основи так званого онто-логічного підходу (ОП) проектування ОКІС (далі для стислості будемо називати його онтологічним). Такий підхід враховує логіко-інформаційну та онтологічну концепції проектування, а також віртуальну парадигму, при якій архітектура комп'ютерної системи орієнтована на технологію реконфігурованого процесингу. Остання забезпечує адаптивність системи завдяки наявності в ОКІС архітектурних і технологічних можливостей налаштування в умовах апріорної та поточної невизначеності на основі навчання і досвіду.

Проектування ОКІС передбачає розробку двох взаємопов'язаних підсистем, відповідно для обробки знань в заданій ПдО та обробки текстів на основі «мовних» знань (або знань з ПдО «Комп'ютерна лінгвістика»). Зазначені інформаційні системи на рисунку 3.1 і 3.2 представлені відповідно як онтолого-керована ІС обробки знань в заданій ПдО та мовно-онтологічна ІС обробки ПМ-інформації на основі мовних знань.

Взаємодія між зазначеними підсистемами здійснюється шляхом реалізації процедури деякого напівформального відображення  $G$ , що використовує базу мовних знань (в основі якої лежить МОКС) [19, 20], відповідну формальну логічну теорію першого порядку і базу знань заданої ПдО.

Схема обробки предметних знань, що містяться в ПМО, представлена на рисунку 3.1.

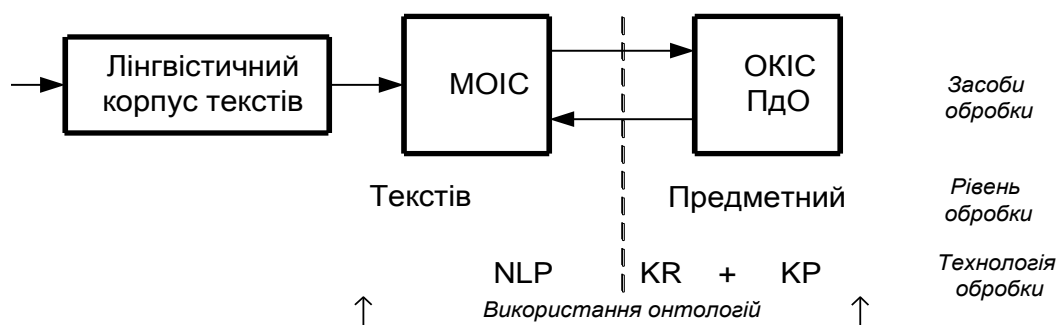


Рис. 3.1. Схема обробки знань, що містяться в ПМО

Онтологічний підхід в онтолого-класифікаційній схемі бере свій початок від двох підходів – логіко-інформаційного [21, 22] і онтологічного.



Коротко розглянемо суть *логіко-інформаційного методу* розробки та оцінки ефективності системи управління ІС з програмовою архітектурою.

Відповідно до логічної концепції логіко-інформаційної моделі процесор представляється відомою композицією операційного і керуючого автоматів. Останній моделює ієрархічну систему управління сучасних ІС, і, в свою чергу, описується багаторівневою композицією програмових автоматів.

Відповідно до інформаційної концепції процесор розглядається як інформаційна система, вся інформація в якій віднесена до трьох "сфер" станів: зберігання, транспортування та перетворення. Вочевидь, що при певних співвідношеннях між об'єктами інформації в цих сферах можна отримати оптимальні технічні параметри інформаційної системи.

Під рівнем управління  $\tau_i (\forall i = 0 \div h)$  розуміється сукупність структурних та інформаційних (програмних) засобів представлення й інтерпретації операторів мови відповідного рівня системи, а також формування необхідної послідовності їх виконання відповідно до заданого алгоритму. Найбільш відомими і поширеними в архітектурах сучасних процесорів і ІС є мікропрограмний, програмний та алгоритмічний рівні.

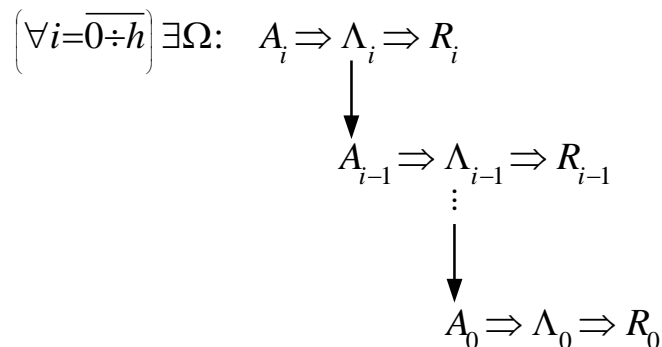
Таким чином, формальний опис моделі можна представити у вигляді

$$\forall_{i=0,h} \tau_i = \langle A_i, \Lambda_i, R_i \rangle, \quad (3.1)$$

Де  $A_i$  – множина алгоритмів, реалізованих і записаних в пам'яті на  $i$ -му рівні;  $\Lambda_i$  – множина операторів програмування  $i$ -го рівня, в термінах яких представлені ці алгоритми;  $R_i$  – множина інформаційно-кодових представлень операторів  $i$ -го рівня.

Оптимальною вважається така структурна реалізація моделі інформаційної системи, для якої відповідно до прийнятих критеріїв знайдені оптимальна кількість рівнів і оптимальні співвідношення між визначеними характеристиками компонент на кожному рівні, а також відповідними характеристиками компонент сусідніх рівнів.

Загальний підхід до проектування архітектури інформаційної системи описується наступною схемою:



де  $\Omega$  – символ відображення алгоритмічного опису інформаційної системи на структурне. Серед усіх варіантів становить інтерес таке  $\Omega$ , при якому оптимізується інтегральний критерій  $\theta$ , що враховує задані характеристики якості архітектури (апаратні витрати, обсяг інформації, швидкодню, функціональні можливості тощо).

Схема ґрунтується на наступних твердженнях. При заданих критеріях, по-перше, існує оптимальний зв'язок між алгоритмами і операторами, в яких вони представлені на кожному рівні, а також операторами та їх інформаційно-кодovими еквівалентами, по-друге, оператор будь-якого рівня можна представити у вигляді алгоритму, заданого на множині операторів нижчого рівня, по-третє, від інформаційно-кодovих представлень операторів на всіх рівнях існує безпосередній перехід до структурних компонентів і параметрів інформаційної системи.

Схема є ядром формалізованої методики проектування, яка являє собою систему взаємопов'язаних алгоритмів проектування. Основні з них відносяться до системного, алгоритмічного і логічного етапів проектування.

У даній роботі логіко-інформаційний підхід модифікується у відповідності з принципами та механізмами онтологічної концепції.

По суті, онтологічна модель ОКІС орієнтується на попередньо «онтологізовані» (на основі системно-онтологічного аналізу [2]) сфери обробки, зберігання та транспортування інформації, що є базовими поняттями в логіко-інформаційній моделі. При цьому умовно можна вважати, що ОП відповідальний за «операційну частину» ОКІС, а ЛП – за «керуючу».

«Онтологізація» сфер переробки інформації має такі характерні риси.

– у сфері зберігання – пам'ять ОКІС проектується з урахуванням онтологічної інформації – сегментується на пам'ять бібліотеки онтологій предметних областей, бібліотеки онтологій текстових документів, загальну базу фактів і правил виведення предметних областей;

– у сфері транспортування йдеться про складну траєкторію природномовної інформації, як носія знань, якій притаманні різного роду неоднозначності (неповнота, неточність) і, перш ніж перейти до машинної обробки цих знань, інформація «фільтрується» (ітераційно) лінгвістичними онтологіями мовних і предметних знань до тих пір, поки не будуть усунуті всі види неоднозначностей;

– у сфері обробки суттєвою є орієнтація на формалізовані предметні знання. Верхнім рівнем формалізації є онтології предметних доменів, в загальному випадку міждисциплінарного характеру, заснованого на системній інтеграції онтологічних знань вихідної множини предметних областей.

Важливою відмітною особливістю ОП в порівнянні з ЛП (крім онтологізації) є додавання ще одного (предметного) рівня управління ОКІС.

Вхідною інформацією для ОКІС є множина ПМО, і першим етапом її обробки є графемно-морфологічний аналіз. На цьому етапі виконується

сукупність програмних процедур, алгоритмічно не складних, але таких, що займають левову частку часу обробки. Істотно підвищити продуктивність (на два, три порядки) системи для зазначеного виду лінгвістичного аналізу можливо, здійснивши перехід з програмного рівня інтерпретації на апаратний рівень. І далі, аналіз показав, що принципово можливо, враховуючи сучасні архітектурні характеристики мікроелектронної бази (ПЛІС) і можливості САПР для проектування обчислювальних пристроїв на ПЛІС, реалізувати всі три етапи лінгвістичного аналізу (графемно-морфологічний, синтаксичний і поверхнево-семантичний) на апаратно-мікропрограмному рівні. При цьому виконується аналіз вхідної ПМ-інформації не послівний, а за реченнями. Речення є мінімальною синтаксичною одиницею, для якої може бути виконаний повний синтаксичний і поверхнево-семантичний аналіз.

Одним з варіантів апаратної підтримки найбільш трудомістких процедур є реалізація системи у вигляді автоматної мережі, включаючи варіант комбінаційних автоматів без пам'яті [23] на базі адаптивних логікових мереж. ЛПІ, а також відомі його модифікації довів свою ефективність при вирішенні задач архітектурного проектування комп'ютерних систем [23, 24].

У схемі для онтолого-керованих архітектур  $\tau_0, \tau_1, \tau_2, \tau_3, \tau_4$ , – рівні управління знаннями

- $\tau_0$  – фізичний рівень;
- $\tau_1$  – мікропрограмний рівень;
- $\tau_2$  – програмний рівень;
- $\tau_3$  – алгоритмічний рівень;
- $\tau_4$  – предметний рівень.

Під *рівнем керування* розуміється сукупність інформаційних та структурних засобів представлення й інтерпретації операторів мови відповідного рівня, а також засобів формування необхідної послідовності їх виконання відповідно до заданого на цій мові алгоритму. Кожен з рівнів управління, таким чином, відповідає визначеному рівню мови:

$$\tau_1 \rightarrow \Lambda_1, \tau_2 \rightarrow \Lambda_2, \tau_3 \rightarrow \Lambda_3, \tau_4 \rightarrow \Lambda_4. \quad (3.2)$$

#### ***а. Структурне подання онтологічної (О) моделі***

На рисунку 3.3 наведено структурне представлення О-моделі, що пояснює її логічну та інформаційну концепції. Кожен з рівнів управління ( $\tau_1, \tau_2, \tau_3, \tau_4$ ) відповідно до інформаційної концепції представлений перетворювачем інформації  $Pr_i, \forall i = \overline{1 \div 4}$  (сфера обробки) і вузлом пам'яті  $Pl_i$  (сфера зберігання). Сукупність зазначених чотирьох рівнів управління представляє керуючий автомат  $A$ . Операційний автомат  $B$  (виконавчий рівень) представлений двома складовими  $Pr_0$  і  $Pl_0$ , що відповідає сучасним тенденціям побудови операційних пристроїв.

Відомо, що в процесі створення ОКІС головними завданнями є розробка методів і підходів систематизації знань, що припускає розробку формальної теорії, у тому числі мови представлення знань (МПЗ), засобів обробки (інтерпретації) знань і механізмів виведення нових знань.

Аналіз процесу дослідження ПдО [5, 23] починається, перш за все, з фіксації певної частини об'єктивної реальності (предметної області), з наступною побудовою її узагальненої моделі (класу IDEF0). На наступному етапі – виявлення найбільш істотних рис, явищ ПдО, визначення елементарних понять і їх взаємозв'язків, аксіом і обмежень, керуючих розвитком даної ПдО, тобто створення деякого образу реальності, що відображає «не строго» суб'єктивне її сприйняття дослідником. Для фіксації добутих знань у вигляді тексту в рамках деякої знакової системи необхідно визначити відповідний формалізм, що забезпечує наочне і зручне їх подання. Він включає власне мову, по можливості більш строгу, ніж природна, і в той же час не дуже відмінна за сприйняттям від останньої, а також метод (методи), що дозволяють здійснити цю формалізацію. В якості такого може бути сполучення алгебро-логічного та аксіоматичного методів.

Наступний етап процесу проектування полягає в побудові формальної теорії досліджуваної ПдО. Ця теорія підлягає інтерпретації, тобто кожному виразу теорії, інакше, кожному поняттю необхідно порівняти конкретний об'єкт реального світу, а кожному судженню – зв'язки між реальними об'єктами, відповідними поняттями, які входять в дане судження, і т.п. Після цього можна визначити істинність або хибність висловлювань теорії, тобто судити про правильність міркувань про предметну область.

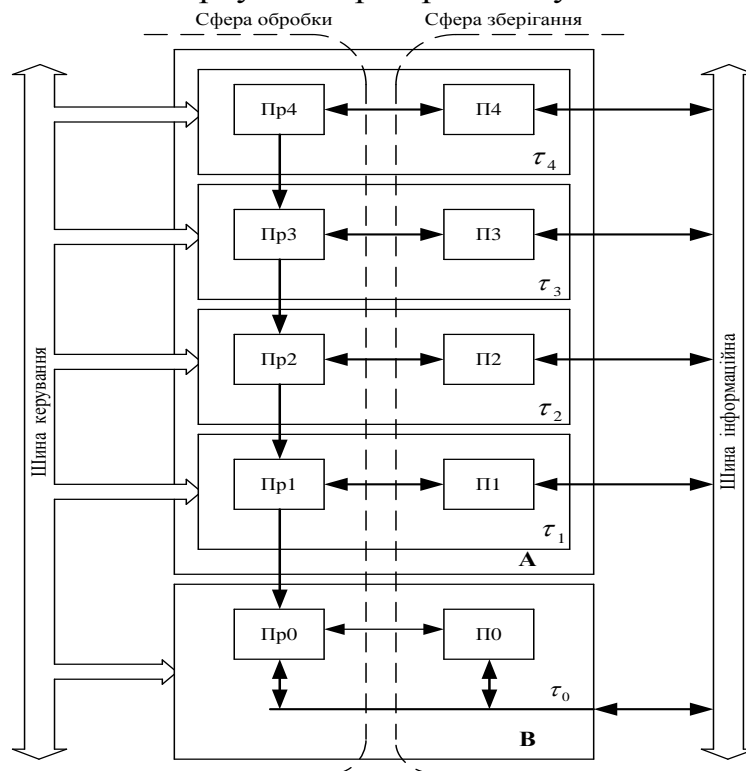


Рис. 3.2. Структура O-моделі з програмовою архітектурою

Далі можна досліджувати побудовану модель в строгому науковому сенсі, ставити по відношенню до неї питання щодо її повноти, несуперечливості, розв'язанні, простежувати її еволюцію і т. п. і, нарешті, інтерпретувати отримані знання в заданій ПдО. Процес синтезу структури ОКІС описується блок-схемою, представленою на рисунку 3.3. Невід'ємною компонентою будь-якої ОКІС є наявність в ній бази знань, в якій в формалізованому вигляді представлені знання предметної області. Саме наявність цих знань в системі дозволяє їй успішно вирішувати як традиційні, так і нові (проблемні) задачі, які раніше були винятково прерогативою наукової творчості. В процесі проектування БЗ все частіше застосовується онтологічний підхід і онтологічний опис предметної області. Виникло самостійне відгалуження в дослідженнях з побудови баз знань – «*онтологічні бази знань*». Онтологічні знання одночасно можуть виступати як інформаційна структура концептуальних знань ПдО і як один з головних компонентів ОКІС. Зазначене функціональне поєднання відображає зазначену вище подвійну парадигму і два аспекти проектування ОКІС – інформаційний об'єкт і інструментальний засіб обробки.

#### *Схема моделі ОКІС та критерії її ефективності*

Модель проектованої ОКІС для вирішення типового набору задач у заданій предметній області можна представити наступною системою

$$S = \langle M, A, X, P \rangle, \quad (3.2)$$

де:  $M$  – множина математичних методів, що лежать в основі рішення типового набору задач в заданій ПдО;

$A$  – множина алгоритмів, що реалізують множину методів  $M$ ;

$X$  – множина об'єктів і процесів, що описують задану ПдО і беруть участь у реалізації алгоритмів  $\{A_i\}$  вирішення  $m$ -го типового набору задач;

$P$  – процедура онтологічного опису об'єктів, процесів і задач заданої ПдО.

Абстрактна архітектура ОКІС описується трійкою

$$A^{OKIC} = \langle A^{MOIC}, A^{OKIC\ PдO}, G \rangle, \quad (3.3)$$

де  $A^{MOIC}$  – архітектура підсистеми обробки ПМО на основі мовних знань;

$A^{OKIC\ PдO}$  – архітектура підсистеми машинної обробки знань в заданій ПдО;

$G$  – відображення результату обробки ПМО в формалізмах деякої відповідної теорії (наприклад, логіки першого порядку).

Для ОКІС процес реалізації О-методу і узагальненого алгоритму рішення типового набору задач носить ітеративний характер. Критеріями

ефективності зазначеного методу (алгоритму) є узагальнені характеристики продуктивності, програмно-апаратних витрат, точності виконання задач, складності алгоритму, надійності проектованої системи або спеціальні критерії, такі як трудомісткість реалізації О-методу і узагальненого алгоритму, повнота обґрунтованих рекомендацій для прийняття рішень, рівень автоматизації побудови онтологічної бази знань ПдО, обмеження реального часу отримання результату та ін.

Результат проектування створюється на основі комплексного набору методів аналізу ПМО, класичної логіки першого порядку для формально-логічного представлення ПМО, структуризації онтологічних знань, а також процедур прикладного процесингу та відповідних архітектур спроектованих ОКІС, вибору підходящої пари (метод – архітектура) для конкретного кола (типового набору) проблемних задач.

### **в. Задачі, що впливають з формальної постановки проблеми аналізу ПМТ**

З формалізованої постановки проблеми аналізу ПМТ впливає необхідність розв'язання таких наступних задач:

- побудувати предметну модель  $A$  ця задача є основною і найбільш складною в зв'язку з тим, що предметна модель по суті є базою знань (побудова такої бази полягає в тому, щоб визначитися з об'єктами, які добуваються з тексту, з формальною логічною мовою, правилами виведення, аксіоматикою тощо);
- показати обчислюваність відношень  $\gamma$  і  $\varphi$  на предметній моделі  $A$  і побудувати алгоритми обчислення відношень  $\gamma$  і  $\varphi$ ;
- при обчисленні відношень  $\gamma$  і  $\varphi$  контролювати відповідність типів аргументів і предикатів;
- визначити взаємодію алгоритмів обчислення  $\gamma$  і  $\varphi$  з системами синтаксичного і семантичного аналізу тексту.

Другорядними, але теж важливими, є задачі пов'язані з:

- визначенням структури даних для словників;
- визначенням інформації, яка повинна знаходитися в словниках;
- визначенням режиму взаємодії з користувачем (автоматичний, автоматизований, діалоговий);
- визначенням мови інтерфейсу користувача й алгоритмів логічного виведення.

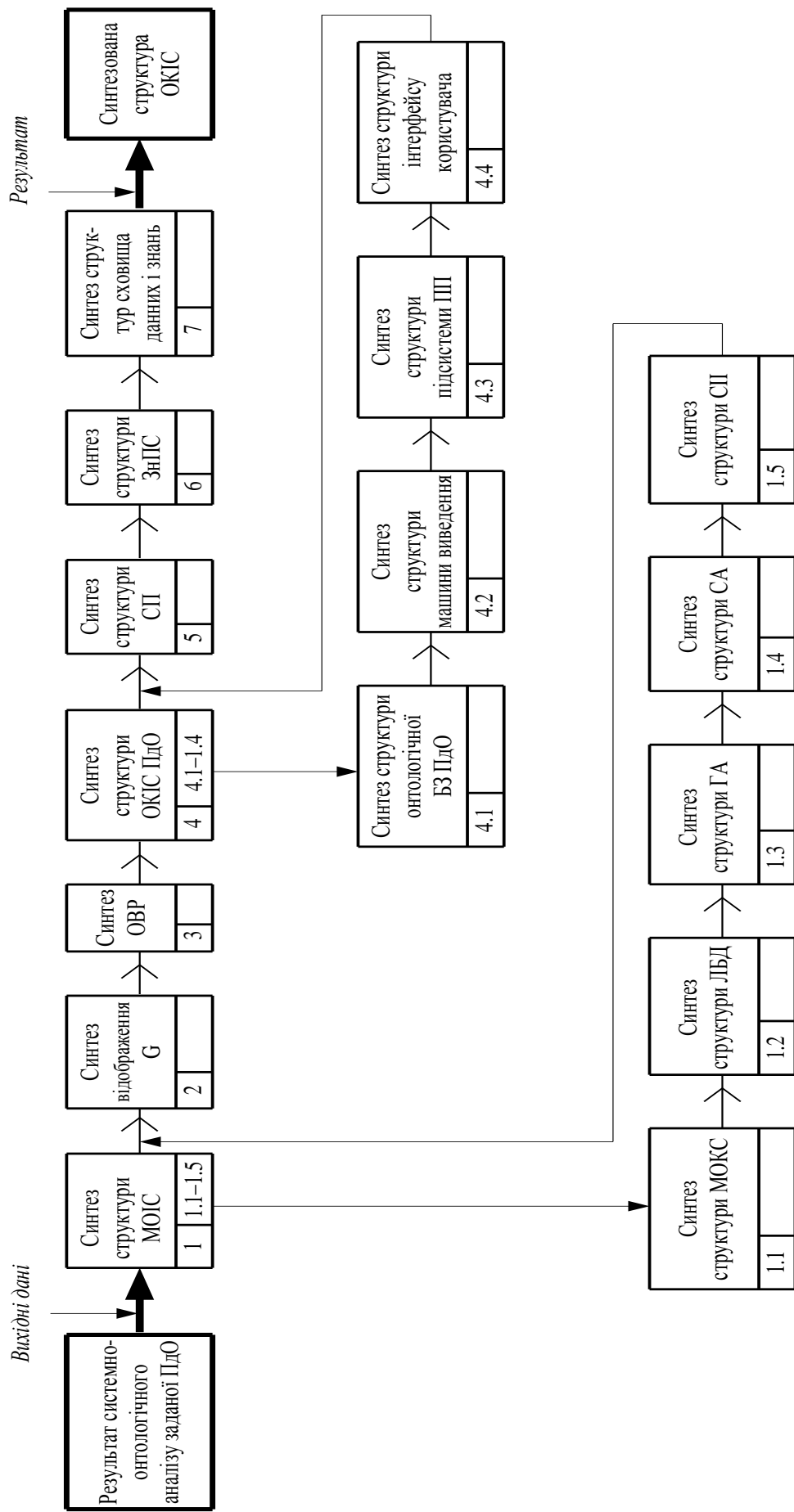


Рис. 3.3. Блок-схема процесу синтезу структури ОКІС

## Обмеження на вхідний текст

Розглянемо коротко структурні обмеження, які накладаються на вхідний текст. Розглянемо два види таких обмежень.

**Нормалізація ПМТ.** Більшість систем як інформаційного пошуку, так і обробки текстової інформації включають в якості основної компоненти систему аналізу, яка служить для виявлення «змісту» або «значення» заданої одиниці інформації. В звичайних системах такого типу аналіз може виконувати людина. При цьому він (аналіз) використовує завчасно розроблені таблиці або шаблони для визначення того, який ідентифікатор змісту за сенсом більше підходить для заданої одиниці інформації. Відомі системи так званого автоматичного індексування, в яких ідентифікатори змісту приписуються автоматично, виходячи із структури тексту документа і запиту.

В зв'язку з тим, що природна мова включає різного роду нерегулярні явища (зокрема, неоднозначності), які зустрічаються як в синтаксисі, так і в семантиці, то система смислового аналізу повинна приводити вхідні тексти до деякого нормалізованого вигляду, перетворюючи різні, можливо неоднозначні, структури на вході у фіксовані, стандартні ідентифікатори змісту. Такого типу процедури нормалізації мови часто використовують словники і списки слів, які включають допустимі ідентифікатори змісту, причому для кожного ідентифікатора наводиться відповідне означення з тим, щоб регулювати і контролювати його використання. Слід зауважити, що до появи поняття «онтологія» (та розробки «онтологізованих» систем обробки ПМТ) процедури аналізу ПМТ рідко виходили за рамки аналізу одного речення. Це пояснюється тим, що проблема аналізу ПМТ досить складна і приходить сильно обмежувати «свої запити» при спробі автоматизації такого роду аналізу, виконуючи деякий спрощений аналіз тексту (або накладати обмеження на стилі оброблюваних текстів).

**Обмеження задачі аналізу.** Конкретизація задачі аналізу спрощується до такої. Словник  $S$ , про який говорилося вище, є тлумачним словником мови  $L(X)$  (це може бути словник російської, української, англійської або будь-якої іншої природної мови). Текст  $T$  складається із речень мови  $L(X)$  і являє собою текст, який не включає ніяких символів, крім символів алфавіту  $X$  (тобто,  $T$  не включає формул, графіків, рисунків і т.п.).

Відношення  $\gamma$  складається із суперпозиції двох відношень  $\gamma_1 * \gamma_2$ , які виконуються послідовно. Відношення  $\gamma_1$  означає розпізнавання належності слова до даної мови і перевірку правильності написання слова  $t_j \in t_i$ , де  $t_i \in T$  у відповідності з написанням його в тлумачному словнику, тобто

$$\gamma_1(t_j) = \begin{cases} 1, & \text{якщо } t_j \in S; \\ 0, & \text{якщо } t_j \notin S. \end{cases} \quad (3.5)$$



Якщо слово  $t_{ij} \in t_i$  розпізнано в словнику  $S$ , то воно заноситься в словник  $T'$  правильних слів, а якщо це не так, то передбачається сигналізація про те, що дане слово відсутнє в словнику  $S$  і приймається рішення про додавання даного слова в словник або його виправлення (слово може бути неправильним, наприклад, внаслідок сканування тексту  $T$ ).

Словники  $S$  і  $T'$  є вхідними даними для відношення  $\gamma_2$ . Змістовний сенс відношення  $\gamma_2$  зводиться до того, що коли  $\gamma_1(t_{ij}) = 1$ , то  $\gamma_2(t_{ij})$  визначає його граматичну одиницю мови (іменник - ім'я власне, іменник - загальний, займенник тощо) а також можливі флексії слова  $t_{ij} \in t_i$ .

Областю інтерпретації тексту  $T$  є модель  $A=(D, \Pi)$ , де  $T$  – вихідний текст, можливо розширений деякою додатковою інформацією, а сигнатура  $\Pi$  визначається виходячи з тексту  $T$  в результаті використання інформації про різні входження слова  $t_{ij}$  в речення  $t_i \in T$ . Обчислення відношення  $\varphi$  обмежується окремо взятим реченням  $t_i \in T$ , яке визначається кожним входженням слова  $t_{ij}$  в текст  $T$ . У випадку труднощів у визначенні предиката  $\pi_i \in \Pi$ , передбачається діалоговий режим обчислення  $\varphi(\pi_i)$  і  $\gamma(\varphi(\pi_i))$ .

### **3.5. Використання онтологічного опису предметних областей при створенні інформаційних систем та роботи з контентом**

Створення онтологій як явних формальних описів предметних областей, доступних як людині, так і програмному агенту, може забезпечити фіксацію, спільне та повторне використання цих знань та їх трансляцію в специфікації програм.

Досягається за рахунок використання програмами онтологічних знань підвищення вбудованого «інтелекту» інструментальних засобів, призначених для створення інформаційних систем і роботи з їх контентом, може забезпечити вирішення наступних завдань:

1. Істотне зниження вимог до розробника, за рахунок зменшення обсягів як програмістських знань, так і знань про предметну область. Можуть бути розширені як можливості фахівця предметної області, недостатньо обізнаного в програмуванні, так і можливості програміста, не завжди знає предметну область у потрібному обсязі («диполь Тиугу»). Зокрема, можна істотно розширити можливості створення прототипів інформаційних систем фахівцями предметних областей.

2. Підвищення рівня мов, що використовуються для адміністрування і роботи зі змістом баз даних і знань, використовуваних інформаційними системами.

Доречне питання: якими повинні бути онтології, що забезпечують опис предметних областей? Сучасні онтології, як правило, зосереджені на концептах-сутності з дуже обмеженою специфікацією. Концепти-зв'язку досить примітивні. Концепти-слоти забезпечують зв'язування сутностей у

безліч ієрархій, але, наприклад, не дозволяють відобразити методи концептів, що представляють собою класи об'єктного програмування.

В багатьох роботах запропоновано деякі розширення поняття «концепт онтології». Зокрема, запропоновано в описі понять-сутностей виділяти чотири класи атрибутів:

- обов'язкові,
- необов'язкові,
- стану (з переходами між ними),
- ресурси.

У моделях понять-сутностей дозволяються будь-які класифікаційні структури на атрибутах, не виключаючи структури, що відображають можливі способи звуження / узагальнення понять. Виділяється явно клас сутностей-ресурсів, які висловлюються через атрибути інших сутностей, у тому числі і сутностей-ресурсів. Для них визначаються характеристики доступності та, якщо це можливо, закони збереження. Поняття-зв'язку, визначаються зазвичай за допомогою атрибутів, які задають боку зв'язку, і атрибутів, що визначають емерджентні властивості, що з'являються тільки при наявності зв'язку. Пропонується розширити поняття-зв'язки, допускаючи більше двох і додаючи такі класи атрибутів зв'язків:

- атрибути, що визначають умови задіяння зв'язку, зокрема, порого чутливості;
- атрибути, що визначають стан зв'язку, його історію.

Для вирішення згаданих вище завдань зниження вимог до розробника та підвищення рівня використовуваних мов залишається додати концептисценарії, що дозволяють описувати процедурні знання. Сценарії характеризуються атрибутами стану, атрибутами, що визначають запуск / зупинка / продовження роботи та атрибутами, обумовленими контекстом. Є підстави вважати, що сценарний підхід і використання деяких досить легко реалізуються особливостей природних мов, таких, як синонімія, анафора і катафора, мабуть, дозволять отримати невеликий і досить просто організований фрагмент природної мови, придатний для роботи з інформаційними системами і досить зручний для користувачів, не обізнаних у мовах програмування.

В даний час не існує інструментальних засобів, що дозволяють працювати з онтологіями описаного виду. Крім того, створення таких онтологій вимагатиме набагато більше зусиль, ніж зазвичай. Тому представляє інтерес використання звичайних онтологій (наприклад, OWL) з розширеннями, визначальними, принаймні, деякі з введених додаткових концептів.

Пропонується в якості інструменту використовувати Protégé у варіанті з базою знань, реалізованої у вигляді єдиної таблиці в деякій базі даних. Примітно, що ця таблиця без порушення працездатності Protégé може бути поповнена додатковими стовпцями, значення яких можуть бути використані тільки поза Protégé або ж спеціально розробленими підключаються до Protégé

програмними модулями.

Таблиця, що зберігає всі класи, слоти, фацет та примірники бази знань Protégé має схему:

```
<frame:integer,      frame_type:smallint,      slot:integer,      facet:integer,
is_template:smallint,      value_index:integer,      value_type:smallint,
slot_or_facet_value:varchar (N), slot_or_facet_value:longvarchar>.
```

Кожен рядок таблиці описує один з фреймів бази знань - клас, слот або фацет. У всіх структурах є слот: NAME, що містить їх назви. Класи мають слоти, що описують ієрархію успадкування. Значення у стовпцях frame\_type і value\_type кодуються. Зауважимо, що такий варіант таблиці представляє найпростіший варіант спеціалізованої універсальної схеми даних, не містить обмежень цілісності.

Обмін інформацією між базою даних програми і базою знань Protégé може виконуватися спеціальним модулем, що входять в додаток, що працює з базою даних. Необхідно основну таблицю доповнити двома довідковими таблицями, що містять кодування стовпців frame\_type, value\_type, і описати алгоритм заповнення таблиці, який на стороні програми буде імітувати дії Protégé.

Синоніми природної мови задаються множинним атрибутом Protégé, що є атрибутом класу (не шаблону). Концепти-зв'язку і концепти-сценарії кодуються аналогічно концептам-сутностей, але посилання на пов'язані з ними процедурні компоненти, не доступні Protégé, перебувають у додаткових стовпцях таблиці зберігає онтологію. Аналогічним чином організована робота з нестандартними для Protégé атрибутами - необов'язковими, стану і ресурсів.

Без доробок Protégé не дозволяє зберігати і візуалізувати такі мережеві структури, які відіграють визначальну роль у введених розширених онтологіях:

- мережі переходів між станами;
- мережі зв'язків, не утворюють ієрархій зі звичайною для Protégé структурою;
- мережі, що описують сценарії.

Введення в таблицю додаткових стовпців, що забезпечують розмітку, дозволяє зберігати довільні мережі. Для їх візуалізації необхідне створення додаткових модулів.

Запропонований підхід дозволяє використовувати звичайні онтології як основу для переробки в розширені онтології. Є можливість будувати фрагмент онтології за схемою бази і будувати фрагмент схеми по онтології.

### **3.6. Застосування теорії онтологій предметних областей для структурування інформації, яка представлена в Інтернеті**

Проблема пошуку інформації в мережі Інтернет у даний момент часу є досить актуальною. Вона вирішується за допомогою вільнодоступних, безкоштовних інформаційно-пошукових систем (Google.com, Yahoo.com,

Yandex.ru) і тематичних каталогів. Пошукові системи охоплюють більше ресурсів, ніж каталоги, але виникає складність при формулюванні запитів. Каталоги надають більш точні результати пошуку, проте відрізняються бідністю структури розділів і старінням інформації.

Один із варіантів вирішення проблеми пошуку інформації в Інтернеті полягає в створенні каталогів. Пропонується створити віртуальний каталог. Його ідея полягає в наступному: користувачеві надається інтерфейс звичайного тематичного каталогу. Після вибору користувача цікавлять його рубрики (або розділи) та відбувається генерація запиту до однієї або декількох пошукових системах. Такий запит формується автоматично віртуальним каталогом на підставі вибраної користувачем рубрики, ієрархії онтологій предметної області та спеціальних евристик.

Для досягнення поставленої мети вирішувалося завдання автоматичної або напівавтоматичної побудови структури розділів каталогу за певною науковою тематикою, а також створення інструментарію, який буде це виконувати. Предметна область, з якою ми працюємо: інформатика та інформаційні технології (тематика каталогу може бути будь-який).

У результаті була розроблена система напівавтоматичного складання структури розділів каталогу на основі онтології. Для цього була складена онтологія предметної області інформатика та інформаційні технології. У результаті роботи системи сформувався каталог, який представляє собою ієрархію розділів з даної тематики. Був розроблений інтерфейс каталогу. По ньому можна здійснювати навігацію: дивитися зміст розділів, визначення понять. Завдяки цьому користувач зможе уточнювати запити під час пошуку потрібної інформації в мережі.

Так само при виконанні роботи була вивчена предметна область, розглянуті існуючі рубрикатори з інформатики та інформаційних технологій, проведено огляд готових онтологій по заданій тематиці.

### **Постановка завдання**

Зараз в Інтернеті є значна кількість матеріалів з різної тематики, що охоплює практично всі галузі знань. Їх обсяг постійно зростає. Різноманітність Інтернет-ресурсів призводить до складності їх класифікації. У зв'язку з цим у користувачів часто виникають проблеми при знаходженні необхідної інформації.

Можна виділити наступні основні форми пошуку інформації в Інтернеті:

- використання інформаційно-пошукових систем;
- використання мета-пошукових систем;
- використання каталогів.

Проілюструємо їх основні переваги та недоліки нескладним експериментом. Будемо шукати інформацію за науковою тематикою в існуючих пошукових системах і каталогах. З отриманих результатів видно, що пошукові системи видають користувачеві значно більше інформації: цифри відрізняються один від одного в десятки тисяч разів. А деякої

інформації й зовсім не міститься в каталогах. Проте, при перегляді ресурсів, знайдених в каталогах та пошукових системах, можна зробити висновок, що каталоги надають більш точніші результати, ніж інформаційно-пошукові системи.

Розглянемо переваги і недоліки обох ресурсів:

### **Каталоги**

Переваги: точність результатів пошуку (ресурси, які заносяться в каталог, перевіряються вручну експертами)

Недоліки: бідність структури рубрик (вони не покривають усі існуючі області знання)

Бідність наповнення ресурсами (так як каталоги, в основному, складаються вручну, а хороших автоматизованих засобів ще не розроблено). Несвоєчасність розміщення інформації: неможливо знайти найновішу інформацію (новини), бо її не встигають заносити в каталог.

### **Інформаційно-пошукові системи**

Переваги: Великий обхват наявних ресурсів в Інтернеті: видають значно більше.

Недоліки: Багато зайвої інформації, не цікавою користувачеві: пошукові системи не перевіряють належність знайденого документа до тематики запиту.

Для вирішення проблеми як метод пошуку інформації в Інтернеті пропонується створення віртуального каталогу. Мета віртуального каталогу - об'єднати в собі переваги пошукових систем і каталогів. Його інтерфейс практично не буде відрізнятися від інтерфейсу звичайних каталогів. Для пошуку необхідної інформації користувач здійснює навігацію по розділах і підрозділах каталогу. На підставі обраної рубрики через використання ієрархії онтологій і з використанням евристик складається серія запитів до однієї або декількох пошукових системах. Можлива вторинна обробка результатів видачі пошуковика. В результаті користувач отримує весь існуючий набір ресурсів, виходячи з специфікованих ним умов. Таким чином, вирішується проблема повноти та актуальності інформації в каталогах.

### **Інструменти створення онтологій**

Опис предметної області займає багато часу і зусиль, тому що необхідно виділити всі ключові терміни, що характеризують область, визначити їх властивості та врахувати взаємовідносини між ними. Вручну це дуже складно зробити, тому для полегшення цього процесу на початку 90-х років почали з'являтися перші засоби розробки онтологій. Вони допомогли користувачеві створювати, редагувати онтології, робити перевірку на несуперечність. Наприклад, на сайті W3C налічується більше 50 таких інструментів.

Проведено короткий аналіз найбільш популярних програмних засобів розробки онтологій: система Ontolingua, система Protege, OntoEdit, OilEd, WebOnto, Ontosaurus, конструктор ODE.

Система Ontolingua розроблена в KSL (Knowledge Systems Laboratory)

Стенфордського університету і стала першим інструментом інженерії онтологій. Вона складається з сервера і мови представлення знань. Сервер Ontolingua організований у вигляді набору онтологій. Є редактор онтологій, який є Web-додатком. Ontolingua включає сервер ОКВС (доступ до онтологій Ontolingua по протоколу ОКВС) і Chimaera (інструмент для об'єднання онтологій в різних формалізми).

Protege - локальна, вільно розповсюджувана Java-програма, розроблена групою медичної інформатики Стенфордського університету. Програма призначена для побудови (створення, редагування та перегляду) онтологій прикладної області. Структура онтології зроблена аналогічно ієрархічній структурі каталогу. На основі сформованої онтології, Protege може генерувати форми отримання знань для введення екземплярів класів і підкласів.

OntoEdit виконує перевірку, перегляд та модифікацію онтологій. Підтримує мови представлення: FrameLogic, OIL, розширення RDFS і OntoEdit (OntoEdit's XML-based Ontology representation Language). Існує дві версії: вільно поширювана OntoEdit Free (обмежена 50 концептами, 50 відносинами і 50 екземплярами) і ліцензована OntoEdit Professional (немає обмежень на розмір).

OilEd - автономний графічний редактор онтологій, розроблений в Манчестерському університеті. Інструмент заснований на мові OIL. Надає засоби редагування онтологій, підтримує міркування, але не має можливості створення примірників класів.

WebOnto призначений для підтримки спільного перегляду, створення і редагування онтологій. Використовує мову OCML (Operational Conceptual Modeling Language).

OntoSaurus складається з сервера онтологій і Web-браузера для редагування та перегляду онтологій LOOM за допомогою HTML-форм, забезпечуючи для них графічний інтерфейс. Основна мета - перегляд онтологій, тому має обмежені засоби редагування.

Конструктор онтологій ODE (Ontological Design Environment) взаємодіє з користувачами на концептуальному рівні: пропонується набір таблиць для заповнення (концептів, атрибутів, відносин) і з них автоматично генерується код в LOOM, Ontolingua і FLogic.

### **Віртуальний каталог**

У віртуальному каталозі є можливість для створення великої кількості рубрик, які будуть мати коментарі на природній мові. При наведенні на рубрику користувачеві буде висвітлюватися інформація про неї. Такі описи розділів можуть надати істотну допомогу користувачу для швидкої та точної навігації.

Робота з пошуковою системою перетворюється на інтерактивний процес щодо уточнення, з високим ступенем конкретизації. Ця діяльність займає багато часу і сил і не обов'язково призводить до потрібних результатів. Крім того, навігацію по каталогу можна розглядати не тільки як

процедуру уточнення запиту, а й як формування розуміння у користувача, що йому, власне, потрібно. У цьому випадку не буде зайвою наявність текстових пояснень до рубрик.

Таким чином, ми маємо повну та актуальну інформацію, яка знаходиться в Інтернеті з даної тематики на даний момент. Такий спосіб має єдину відмінність від роботи з пошуковою системою: засобами, які надають пошукові системи, недосвідченому користувачеві досить складно описати вимоги до ресурсу, який він хоче знайти.

Опишемо структуру каталогу. Існує такі способи: подання рубрик за алфавітом (найпростіший спосіб вирішення), дерево рубрик. Використання деревовидної структури не допускає перетин рубрик, отже, потрібно шукати інші варіанти представлення каталогу.

Було вирішено розробити рубрикатор наступного вигляду:

- слой рубрик (набір областей науки, які мають між собою певні зв'язки, наприклад, частина-ціле);
- шар атомарних областей - це поняття, терміни (набір ключових слів для кожної рубрики);
- ресурси.

### 3.7. Інформаційна модель етапів лінгвістичного аналізу

Архітектура сучасних знання-орієнтованих інформаційних систем (ЗОІС) з природно-мовним представленням і обробкою знань включає онтологічну складову експліцитно, яку в загальному вигляді можна інтерпретувати як концептуальну базу знань. Така база знань представляється у вигляді орієнтованого графа, вершинами якого є фрейми, що описують концепти, а дугами – множина концептуальних відносин, що зв'язують між собою концепти. Іншою важливою особливістю вказаної архітектури є розділення обробки семантики першого і другого ступеня, що в загальному випадку означає розділення внутрішньомовного і позамовного процесингу [1] і перехід до формально-логічного представлення початкового тексту.

Вказані особливості архітектури сучасних ЗОІС трансформують традиційну модель обробки природно-мовних текстів у формальну модель наступного вигляду

$$F = \langle T, W, SS^1, O, S^2, I \rangle, \quad (3.6)$$

де:

$T$  – множина ПМТ;

$W$  – множина словоформ, що входять в  $T$ ;

$SS^1$  – множина синтактико-семантичних структур першого ступеня, що описують  $T$ ;

$O$  – множина онтологічних структур, що відображають множини  $W$  і  $SS^1$  в  $S^2$ ;

$S^2$  – множина семантичних структур другого ступеня, що описують множину сценаріїв  $T$ ;

$I$  – множина інформаційно-кодових представлень  $S^2$ .

Опишемо об'єкти формальної моделі.

Множина  $T$  представляє сукупність природно-мовних текстів, що характеризуються стилями ділового і науково-технічного характеру.

Ланцюжок  $W \rightarrow SS^1$  в класичному розумінні представляє граматичний аналіз ПМТ. На відміну від традиційних лінійного і сильнокодованого методів аналізу тут використовується змішаний метод аналізу. Суть його полягає в тому, що в лексикографічній базі даних повна множина  $W$  представлена в таблицях двох типів: таблиці лексем з відповідними морфологічними, синтаксичними і семантичними характеристиками і таблиці флексій для всіх повнозначних, частин мови, що змінюються. При цьому алгоритми формування парадигми лексем прості: у таблицях лексики вказані основи лексем і відповідні коди для вибору записів з таблиць флексій. Нефлексійні зміни враховуються відповідними алгоритмами.

Описана структура граматичного аналізу однозначно відповідає ефективному відображенню функціональних операторів на програмно-апаратний рівень реалізації.

Множина  $O$  онтологічних структур в ідеалі представляє мовно-онтологічну картину світу [5, 24].

Множина  $SS^1$  формується і інтерпретується ітераційно підсистемою синтактико-семантичного аналізу типу «Конспект» [2]. Основною операцією синтактико-семантичного аналізу є розпізнавання синтаксичних і семантичних відносин, що зв'язують слова тексту. Для кожного речення початкового тексту будується дерево розбору. Визначення семантичної неоднозначності здійснюється шляхом звернення до множини онтологічних структур  $O$ . На основі побудованих дерев розбору фраз будується категоріальна мережа, що є семантичним простором  $S^2$  тексту. Для комп'ютерного представлення тексту зручно використовувати, зокрема, семантичну мережу множини інформаційно-кодових представлень  $I$ , яка організована на основі пірамідальної мережі (підсистема «Конфор» [25]), рецептори якої відповідають іменам об'єктів, класів об'єктів, властивостей, станів, дій, відносин, семантичних відмінків і модифікаторів.

Ланцюжок перетворення інформації  $T \rightarrow W \rightarrow SS^1$  і  $O \rightarrow S^2 \rightarrow I$ , по суті, представляють (відповідно) базові процедури аналізу і розуміння ЕЯТ, засобами інтерпретації яких є граматичний і семантичний процесори.

У додатках пошуку і обробки великого об'єму текстових документів доцільно використовувати знання-орієнтовану пошукову систему, що забезпечує початковий і кінцевий етапи обробки документів – пошук в Інтернет і збереження документів в базі даних у вигляді їх конспектів, які згенеровані підсистемою «Конспект».



## Графемно-морфологічний аналіз

Як початковий етап лінгвістичного аналізу використовується графемно-морфологічний аналіз (ГМА). При виконанні графемного аналізу текст розглядається як різновидність знакової системи. Задачею даного етапу аналізу є дослідження властивостей одиниць мови і правил їх поєднання в аспекті їх знакової природи. В існуючих моделях графемний аналіз реалізовано тільки як засіб розв'язання протиріч аналізу текстових одиниць, що надходять на обробку морфологічного аналізатора і окреслюється відділенням аббревіатур, скорочень та деяких інших класів лексем. Це, в свою чергу, породжує протиріччя на етапі семантичного аналізу тексту, що призводить до помилок семантичного характеру, уникати яких користувачу пропонується самому, вручну маркірувати текст.

Призначення *графемного аналізу* – побудова моделі графемної структури вхідного тексту, в якій виділені і зв'язані відношеннями (де це можливо) такі змістовні одиниці тексту як фрагмент, речення та лексема. Наприклад, для лексем: виділяються класи лексем, які відрізняються своєю графемною структурою і виконують різні функції в тексті. Крім того, аналізуються закономірності сполучуваності деяких лексичних одиниць, які уже на цьому етапі дозволяють об'єднувати декілька лексем в одну на тій основі, що вони виконують в тексті єдину функцію [24-25].

Внаслідок графемного аналізу (точніше, інтерпретації бази знань графічного зображення ПМТ з вхідним текстом) текст перетворюється у впорядкований список складників ПМТ.

Під *морфологічним аналізом* розуміють повну обробку послідовності (без будь-якого зв'язку з контекстом) словоформ, сформованих на під етапі графемного аналізу. Внаслідок такої обробки кожна словоформа замінюється граматичною інформацією – ланцюжком символів, що визначають усі ті властивості даної словоформи, необхідні для подальшого аналізу. Морфологічна інформація сукупності словоформ, в свою чергу, використовується на наступному етапі створення програмної моделі ЛП – під час синтаксичного аналізу, внаслідок якого встановлюються зв'язки між усіма словоформами тексту та між частинами складних речень.

Вхідними даними процедури морфологічного аналізу є графемна структура тексту, база знань морфології ПМ та лексикографічна база даних (ЛБД).

ЛБД включає словники лексем та словозмінну і словотвірну моделі вхідної мови. Вона включає окремі таблиці для всіх частин мови. До кожної лексеми в таблиці приєднується, окрім традиційних морфологічних характеристик, набори синтаксичних і семантичних ознак [26-29].

Крім того, існують окремі таблиці відмінкових закінчень (для змінних частин мови) для формування словоформ лексеми. Всі лексичні одиниці в таблиці відповідним чином проіндексовані та мають однакове інтерпретаційне значення на всіх етапах лінгвістичного аналізу.

## Синтаксичний аналіз

Синтаксичним аналізом (parsing) називається процес структуризації лінійної репрезентації відповідно до заданої граматики. Лінійною репрезентацією речення природної мови будемо називати ланцюжок елементів, де кожен елемент є мінімальною синтаксичною одиницею. Мінімальна синтаксична одиниця може бути словоформою або оператором з певним набором характеристик. Оператором називається розділовий знак або сурядний сполучник. Процесом структуризації лінійної репрезентації речення називається побудова орієнтованого графа синтагм і орієнтованого графа сегментів. *Синтагма* визначає бінарне синтаксичне відношення вигляду  $R(A, B)$ , де  $A$  і  $B$  - словоформи, а  $R$  - тип синтаксичного відношення, який відповідає імені синтагми, причому  $A$  керує  $B$ . Вузлами графа синтагм є термінальні одиниці. Зв'язаність не є обов'язковою умовою такого графа, оскільки синтагми спираються тільки на морфологічні репрезентації словоформи, лінійний порядок речення і, в деяких випадках, на примітивну модель управління.

*Сегмент* можна визначити як частину речення, яка виділена в тексті знаками пунктуації і описує окрему ситуацію. Кожен такий сегмент має як вершину явний предикат, який являє собою в більшості випадків фінітну форму дієслова або «прихований» предикат, який може бути виражений дієприслівником, дієприкметником або ім'ям з семантичною характеристикою дії; кожен такий предикат і задає ситуацію. Вузлами графа сегментів є нетермінальні одиниці. У західній лінгвістичній традиції поняття сегмент еквівалентно терміну клауза. Сегментація речення є першим етапом повного синтаксичного аналізу. Метою сегментації є виділення і класифікація сегментів у складі складного речення. Задачею другого етапу є усунення граматичної омонімії морфологічного рівня і побудова словосполучень. Об'єктом розпізнавання цього рівня є закономірності взаємодії лексем у межах мовної синтагми. Вхідними даними є результати попередніх етапів розпізнавання та база знань синтаксису ПМ. В ній закладені моделі синтаксичних правил, які визначають ознаки синтаксичного з'єднання лексем у словосполучення.

Серед мовних засобів синтаксичного аналізу можна виділити наступні:

1. *Словозмінні морфологічні засоби*. Для мов з розвинутою морфологією, якою є українська та російська мови, - це основний спосіб матеріалізації синтаксичних зв'язків. Показником морфологічної залежності служить флексія, оскільки граматичні словоформи звичайно приписані флексії, що дозволяє в деяких випадках виявити залежність між двома словоформами, відсутніми у словнику. В українській мові словоформи багатьох слів омонімічні (наприклад, це може бути *омонімія числа і відмінка*), що створює певні труднощі у процесі аналізу. Неоднозначність графем у ході автоматичного синтаксичного аналізу іноді призводить до виникнення синтаксичної омонімії і побудови альтернативного синтаксичного варіанту (графа синтагм). Взаємовплив двох рівнів аналізу

морфологічного і синтаксичного полягає в тому, що грамема, будучи ефективним засобом пошуку морфологічної залежності, яка служить одним із способів реалізації синтаксичного відношення, може бути однозначно проінтерпретована тільки внаслідок фіксації цього відношення.

2. *Селективні ознаки*: Класифікуючі (селективні) ознаки приписуються лексемам у граматичному словнику. Найбільш важливою для синтаксису є класифікація лексем за категоріальними ознаками: іменник, дієслово, прикметник та ін. *Категоріальні ознаки задають потенційних учасників синтаксичного зв'язку і визначають у більшості випадків вершину у структурі, а також зумовлюють поняття управління й узгодження.*

3. *Службові слова*: прийменники, сполучники, частки й т. ін. – засоби, які слугують опорними точками аналізу. Так, наприклад, сполучник може бути використаний для визначення поверхневого типу сегмента, прийменник формує актант дієслова.

4. *Розділові знаки (оператори)*: кома, тире, крапка, знак питання й т. ін. У першу чергу оператори визначають межі як сегментів, так і всього речення. Тире є виразом еліптованого елемента речення і часто використовується у стратегіях пошуку неморфологічного предиката. Аналіз біфункціональності оператора (коли, наприклад, оператор є одночасно і правою межею сегмента, і оператором творення слів) – одне з найважчих завдань сегментації, яка задає рекурсивний характер як граматичних стратегій аналізу, так і методів програмної реалізації.

5. *Порядок слів*: Лінійне розташування слів у реченні разом із селективними ознаками має домінуюче значення у проектуванні синтаксичних аналізаторів мов з бідною морфологією (англійська). У багатьох системах англійського синтаксису порядок слів задає напрям пошуку господаря або слуги для кожного класу лексем і типу зв'язку. Для флективних мов цей засіб аналізу має другорядне значення, хоча і застосовується у ряді випадків для встановлення синтаксичних зв'язків або оцінки омонімічних структур речення.

Незважаючи на вільний порядок слів в українській мові, деякі синтаксичні залежності мають обов'язковим критерієм виділення жорсткого лінійного порядку: генітивне визначення має слідувати за словом ('ніжка стол-а', 'син батьк-а'); прийменник передує іменнику ('на стол-і', 'у батьк-а'). Лінійний порядок речення в автоматичному синтаксичному аналізі використовується як показник найбільш вірогідного напрямку пошуку слуги або господаря, і лише в окремих випадках – як обов'язковий критерій встановлення синтаксичної залежності.

Для перевірки і встановлення можливого синтаксичного відношення використовуються граматичні правила і стратегії. Кожне граматичне правило встановлює один тип синтаксичного відношення  $R(A, B)$  між двома одиницями аналізу й однозначно задає вершину. Кількість типів синтаксичних відношень, а також їх назви залежать від прикладної моделі і конкретної реалізації системи синтаксичного розбору. Набір універсальних синтаксичних відношень для російської мови (може бути використаний при

аналізі української мови) наведений в багатьох теоретичних роботах, наприклад [8]: відприйменникове (іменник і прийменник, який ним керує), визначальне (іменник і його узгоджене визначення), посесивне (іменник і його неузгоджене визначення), суб'єктне (присудок і підмет). В ролі одиниць аналізу на місці А і В, де А - вершина, а В - залежне, можуть виступати як окремі словоформи, так і цілі групи (фразові складові); заповнення А і В багато в чому залежить від синтаксичного апарату, прийнятого в аналізаторі для опису структури. Ідеальне граматичне правило в автоматичному синтаксичному аналізі характеризується наступними критеріями: 1) описує тільки один тип синтаксичного відношення; 2) однонаправленість аналізу, тобто залежне В може знаходитися тільки зліва або тільки справа від вершини А; 3) не містить рекурсивних викликів або викликів інших правил; 4) обробляє тільки компактно розташовані одиниці аналізу; 5) результат не залежить від порядку застосування правил. Використання граматичних правил задає прозорість архітектури процесора і забезпечує стійкість системи до змін. Перераховані критерії не є строгими, але наближають правило до його ідеальної форми.

Результатом синтаксичного аналізу є побудований орієнтований граф синтагм і орієнтований граф сегментів речення українською мовою.

### **Семантичний аналіз**

Під автоматичним семантичним аналізом будемо розуміти сукупність методів, що реалізовані у вигляді однозначної формальної процедури, за допомогою якої можливо з достатньою точністю представити зміст будь-якого висловлювання природною мовою у вигляді послідовності символів, які утворюють деяку формальну систему.

Найбільш простим і універсальним засобом представлення змісту деякого тексту в комп'ютерній системі є семантична мережа. У загальному випадку вона є орієнтований граф, вершини якого позначають сутності (об'єкти), а ребра – відношення (зв'язки) між ними. Імена вершин і ребер звичайно збігаються з іменами відповідної сутності і відношення, які використовуються у природній мові. Ребро і дві зв'язуванні їм вершини представляють мінімальну значеннєву інформацію – факт наявності зв'язку певного типу між відповідними об'єктами. Виділяють два підетапи у проведенні семантичного аналізу. *Перший* підетап названо поверхневим семантичним аналізом. На цьому підетапі здійснюється усунення синтаксичної омонімії (якщо вона має місце), формування понять, відношень та їх характеристик (властивостей) у межах одного речення. Крім того, на цьому етапі проходить заміщення анафоричних зв'язків (зокрема, коли замість поняття підставляється займенник), узагальнення понять і відношень, пропусків понять і відношень, що притаманно природній мові. *На другому* підетапі семантичного аналізу будується власне онтолого-семантична структура тексту.

Згідно з [16], із множини слів української мови (для російської мови також) за семантичними ознаками можна виділити наступні категорії: слова з

моделлю управління (предикати) та слова без моделі управління (поняття, характеристики та їх значення, числівники).

Для лінгвістичного процесора найбільш важливою є обробка предикатів дії та їх можливих актантів – понять, характеристик, їх значень, а також числівників. Обробка функціональних і порожніх предикатів може бути зведена до модифікації фрагментів семантичної мережі, побудованих для залежних від них предикатів дії.

Кожен предикат дії має одну або декілька моделей управління. Модель управління накладає синтаксичні і семантичні умови на можливі актанти (аргументи) даного предиката і зазначає їх семантичні ролі щодо предиката. У спрощеному виді модель управління можна представити як таблицю, кожен рядок в якій визначає один з можливих актантів. Цей рядок містить: прийменник (якщо є), частину мови і відмінок актанта, його семантичне метаяпоняття (категорію), семантичну роль у предикаті і ознаку обов'язковості даної ролі.

Аналіз моделей управління дозволяє природним чином перейти від синтаксичного дерева залежностей до фрагмента семантичної мережі. Для цього у простому випадку достатньо: 1) для кожного предиката відшукати в дереві залежностей всі присутні актанти; 2) створити по одній вершині на кожен актант і ще одну – для самого предиката; 3) провести від вершини-предиката дугу до кожної вершини-актанта; при цьому назва дуги (її семантична роль) вибирається з моделі управління залежно від актанта. Якщо модель управління не допускає існування актанта ні в одній семантичній ролі, залишається дві можливості: або дана текстова одиниця насправді є стосовно предиката не актантом, а обставиною (тоді цей зв'язок можна відобразити дугою відповідного типу), або має місце помилка синтаксичної структури.

Правильно розрізняти ці два випадки вдається не завжди, проте можна вказати умову, за якої з великою вірогідністю присутня саме помилка синтаксичної структури. Мається на увазі ситуація, коли актант задовольняє морфологічним умовам моделі управління, а семантичним – ні. Оскільки синтаксичний аналізатор при угрупуванні не перевіряє семантичні ознаки, він не може самостійно виявляти помилки такого типу. Таким чином, якщо помилка виявлена на етапі семантичного аналізу, необхідно повернутися назад до синтаксичного аналізу і спробувати знайти інший варіант синтаксичної структури, який би не містив невірною актанта у піддереві залежностей даного предиката.

Семантичний аналіз здатен виявляти деякі помилки синтаксичної структури (наприклад, пов'язані з невірною моделлю управління). Таким чином, існує зворотний зв'язок між семантичним і синтаксичним аналізом – необхідна властивість *лінгвістичного процесора*, тому доцільно об'єднати ці два етапи аналізу та виконувати їх спільно в одному аналітичному блоці.

## Інформаційно-структурна організація етапів лінгвістичного аналізу

Практична цінність одержуваних результатів при обробці ПМО, в основному, залежить від повноти інтерпретаційних моделей семантичних структур ПМО та їхнього формального представлення. Під повнотою ми розуміємо включення в модель як складової семантики першого ступеня (або об'єктової складової), так і складової семантики другого ступеня (або акторної складової). Такий розподіл семантики добре узгоджується як із онтологічною ієрархією концептуальних категорій, так і зі складністю виконання обчислювальних процедур при комп'ютерній обробці ПМО.

З погляду лінгвістики, семантична складова першого ступеня описується на рівні граматики окремих частин мови, в той час як складова другого ступеня описується синтаксичними конструкціями таких одиниць синтаксису як речення, абзац, розділ і текст. З погляду математичної логіки, якщо перший ступінь можна описати (досить умовно) численням висловлювань, то другий ступінь має описуватися численням предикатів з квантифікованими змінними.

Найбільшої повноти (і відповідно найбільшого ступеню складності) набувають моделі, що описують ПМО в цілому. Такі моделі описують, зокрема, деякий сценарій (вищу категорію, що описує явище, взаємовідношення об'єктів, що перебувають у постійному русі), що відображає зміст ПМО. У свою чергу, як ПМО ділиться на синтаксичні одиниці, так і загальний сценарій розділяється на окремі сценарії, ситуації та елементарні ситуації.

Описана істотна різниця між об'єктовою та акторною складовими семантики, а також морфолого-синтаксичним аналізом, зокрема в складності їхніх інтерпретаційних моделей, обумовила виділення для моделювання й інтерпретації семантики окремого функціонального модуля – семантичного процесора. Морфологічний і синтаксичний аналіз при цьому виконується граматичним процесором, а точніше, окремими його блоками графемно-морфологічного та синтаксичного аналізу.

Основним призначенням *семантичного процесора* (СП) є побудова формалізованого опису вихідного ПМО і його відображення в онтологічному дереві МОКС. Інакше кажучи, головним завданням СП є відображення структури тексту на онтологічну структуру МОКС і фіксація семантико-синтаксичної структури окремих речень і текстових фрагментів у вигляді відповідних сукупностей індексів, що зв'язують відношеннями повну множину лексем та їхніх значень, представлених у МОКС.

Знання про навколишній світ в МОКС вичерпуються системою сформульованих певною мовою на рівні здорового глузду та зв'язаних між собою максимально повною системою відношень понять, що відображають навколишній світ з повною множиною його об'єктів і явищ, тобто являють собою лінгвістичну проекцію буття людини, де зафіксований досвід її взаємодії з навколишньою дійсністю. Прототипами таких МОКС можуть

служити тезауруси, ідеографічні (синоптичні) або тлумачні словники, енциклопедії, кожний з яких – деяка поняттєво-структурна класифікація об'єктів реальності з фіксованою ієрархією категорій, понять, класів, родів, видів, понятійних семантико-синтаксичних груп та ін.

У даному випадку МОКС – складова прагматичної моделі мовної свідомості, що є ключовою компонентою сучасних інтелектуальних інформаційних систем із природномовним представленням, обробкою й актуалізацією знань. Безумовно, при цьому не забувається, що поряд з лінійними лексикографічними конструкціями людська свідомість реально має справу також з більш складними представленнями об'єктів навколишнього світу – ментальними образами.

На основі виконаного аналізу моделей та загальних принципів комп'ютерної обробки ПМО на рисунку 3.4 синтезовано структурно-логічну схему етапів лінгвістичного аналізу.

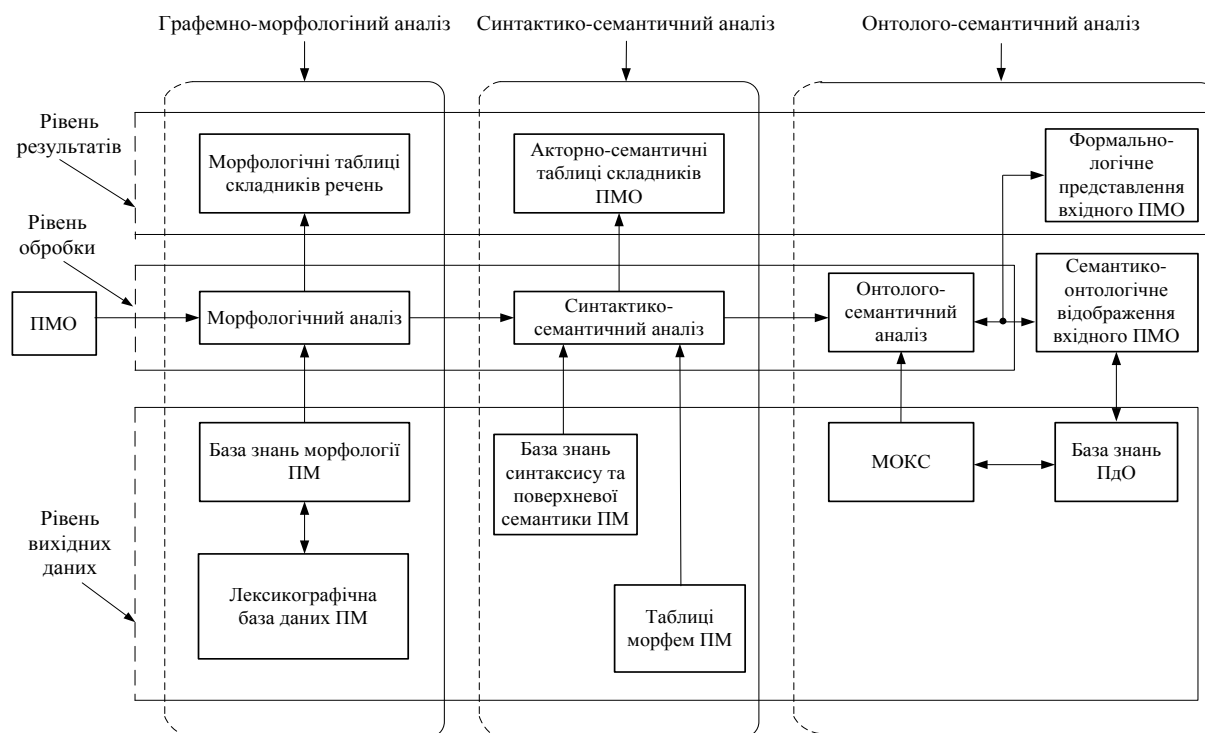


Рис. 3.4. Структурно-логічна схема етапів лінгвістичного аналізу

На схемі база знань ПМ складається з МОКС та баз знань етапів графемного-морфологічного та синтактико-семантичного аналізу. База знань частини реального світу, до якого належить лінгвістичний корпус ПМО, відображена в блоці «База знань ПдО».

### 3.8. Алгоритми семантичної інтерпретації морфологічних властивостей одиниць природної мови

В загальному випадку лінгвістичний аналіз складається з чотирьох етапів – графемного, морфологічного, синтаксичного та семантичного.

Більшість експериментальних досліджень в галузі машинної морфології починалися зі створення машинного словника. Алгоритми програм, що працюють без машинного словника, використовують ймовірно-статистичні методи та лексикони основ чи квазіоснов, суфіксів чи квазісуфіксів, побудованих емпіричним шляхом. Наприклад, в [30] описано працюючу модель морфологічного аналізу, яка не потребує об'ємних словників основ слів. Вона використовує наступні лексикони: закінчень і рефлексивів, суфіксів, квазікоренів, префіксів та основ. Кожній одиниці такого лексикону приписано всі можливі (але не вичерпні) граматичні характеристики словоформ, частиною якої може бути вказана одиниця.

Моделі, які використовують словники, здатні дати більш повний аналіз словоформи (тобто оперувати більшим числом граматичних ознак). Ступінь точності такого аналізу вище у порівнянні з моделями, які не використовують словники.

На теперішній час існують електронні версії повного лексикографічного словника природної мови (зокрема для російської та української мов), які практично виключають ручну розробку машинного словника.

Значні досягнення в області мікроелектроніки дозволяють розробляти апаратні лінгвістичні процесори з достатнім об'ємом пам'яті для збереження машинних словників і на 1-2 порядки вищою швидкістю.

Архітектурно-структурна організація вказаних апаратних засобів та сучасні засоби проектування (САПР ПЛІС – системи автоматизованого проектування програмовних логічних інтегральних схем) дозволяють в короткий термін реалізувати ефективні механізми асоціативного пошуку (базової процедури співвіднесення словоформ вхідного ПМО з лексемами словника) та паралельного аналізу словоформ речення в цілому, що в перспективі може привести до створення паралельних (на апаратній основі) лінгвістичних процесорів.

Отже, задача автоматизованого аналізу ПМО зводиться до двох параметрів: якості, що визначається парою – точність (рівень помилок у побудованих лінгвістичних структурах речень) і повнота (ступінь покриття тексту синтаксичними зв'язками, або зв'язність графа речення), та швидкодії, поки що недостатній для ряду прикладних задач.

Описані два підходи до побудови моделей лінгвістичного аналізу названо в [31] відповідно: граматичним та безпосереднього розпізнавання. Вони, як вказувалося вище, мають свої недоліки та переваги, а тому оптимізація параметрів «Якість ↔ Швидкодія» для кожного конкретного проекту повинна виконуватись індивідуально, але з врахуванням наступних зауважень.

Розробка моделей лінгвістичного аналізу без словника потребує праці групи висококваліфікованих лінгвістів. Експериментальні дослідження моделі з граматичним розпізнаванням показали її високу ефективність для етапів графемного та морфологічного аналізу. При цьому вказані етапи можливо об'єднати в один (графемно-морфологічний) і виконувати їх майже



паралельно. Для етапів синтаксичного і семантичного аналізу переваги моделі дещо нівелюються за рахунок морфологічної та синтаксичної омонімії.

Виконаний аналіз двох підходів до побудови моделей лінгвістичного аналізу, їх переваг та недоліків, можливостей усунення недоліків, дозволяє стверджувати про можливий третій, *комбінований підхід*. Його особливості та практичну реалізацію буде розглянуто нижче.

В якості початкового етапу лінгвістичного аналізу виконується *графемно-морфологічний аналіз*.

Призначенням *графемного аналізу* є побудова моделі графемної структури вхідного тексту, в якій виділені і зв'язані відношеннями (де це можливо) такі змістовні одиниці тексту, як фрагмент, речення та лексема. Наприклад, для лексем: виділяються класи лексем, які відрізняються своєю графемною структурою і виконують різні функції в тексті. Крім того, аналізуються закономірності сполучуваності деяких лексичних одиниць, які уже на цьому етапі дозволяють об'єднувати декілька лексем в одну на тій основі, що вони виконують в тексті єдину функцію.

В результаті графемного аналізу текст перетворюється у впорядкований список складників ПМТ.

Під *морфологічним аналізом* розуміють повну обробку послідовності (без будь-якого зв'язку з контекстом) словоформ, сформованих на етапі графемного аналізу. В результаті такої обробки кожна словоформа доповнюється граматичною інформацією – ланцюжком символів, що визначають усі властивості словоформи, які необхідні для подальшого аналізу. Морфологічна інформація сукупності словоформ, в свою чергу, використовується на наступному етапі створення програмної моделі лінгвістичного процесору – під час синтаксичного аналізу, в результаті якого встановлюються зв'язки між усіма словоформами тексту та між частинами складних речень.

Вхідними даними процедури морфологічного аналізу є графемна структура тексту, база знань морфології природна мова та лексикографічна база даних, яка включає словники лексем та словозмінну і словотвірну моделі вхідної мови. До кожної лексеми в лексикографічній базі даних приєднується, окрім традиційних морфологічних характеристик, набори синтаксичних і семантичних ознак. Крім того, існує окрема таблиця відмінкових закінчень (для змінних частин мови) для формування словоформ лексеми. Всі лексичні одиниці в таблиці відповідним чином проіндексовані та мають однакове інтерпретаційне значення на всіх етапах лінгвістичного аналізу.

Існує зворотний зв'язок між семантичним і синтаксичним аналізом, тому було вирішено об'єднати ці дві процедури в одному блоці – *синтактико-семантичного аналізу*. Кінцевим завданням процедури аналізу є представлення кожного речення заданого ПМТ у вигляді синтаксичного дерева (лексем речення з синтактико-семантичними відношеннями між ними).

При побудові дерева розбору речення зв'язування слів у реченні відбувається поетапно, від пошуку словосполучень до визначення груп зв'язаних слів і, зрештою, до об'єднання всіх груп у реченні в одну структуру. Для встановлення зв'язку між окремими словами використовуються природномовні засоби вираження семантичних та синтаксичних відношень. У флективних мовах такими засобами є змінні частини повнозначних слів та службові слова.

Такі сегменти словосполучень, які кодують відношення між повно значними словами, називаються *синтаксичними визначниками* [32]. Оскільки одному синтаксичному визначнику може відповідати декілька синтаксичних відношень, для однозначності визначення зв'язків між словами вводиться поняття *кореляторів* [32], які додатково враховують семантичні ознаки слів у словосполученні.

Вихідними даними для блоку синтаксичного аналізу є:

- результат попередніх етапів аналізу ПМТ (графемного і морфологічного);
- словник основ (містить основи слів та їх семантичні ознаки);
- список всіх можливих флексій слів;
- база даних з визначниками (містить синтаксичні визначники та списки кореляторів для кожного з них);
- база даних з кореляторами (кожен корелятор складається з назви відношення та списку пар семантичних ознак основ, між якими це відношення може існувати).

Необхідність побудови омонімічних синтаксичних структур була передумовою автоматичного створення додаткових словників морфів, що значно підвищило швидкодію етапів синтаксичного і семантичного аналізу.

Незмінна та флективна частини слова та його семантична ознака для всіх слів ПМТ передаються з попереднього етапу аналізу (етап морфологічного аналізу). Синтаксичні характеристики в таблицях лексикографічної бази даних, наприклад дієслів, включають лексико-семантичні валентності дієслів, що визначають ознаки найбільш вірогідного оточення та ознаки семантичних інтерпретацій, які на основі розпізнаних синтаксичних правил визначають поверхневі семантичні відношення між словами у словосполученні. Якщо речення складається з декількох синтагм, то завдяки інтерпретації певної частини бази знань синтаксису відбувається їх об'єднання у єдине синтаксичне дерево з відповідними семантико-синтаксичними відношеннями.

Нижче описані основні етапи роботи блоку *синтактико-семантичного аналізу* природномовного тексту.

#### *Перший етап.*

Проводиться класифікація слів у реченні за семантичними ознаками відповідних основ у словнику. Зв'язування слів у реченні починається зі словосполучення, що визначає головне відношення (відношення між підметом і присудком) у цьому реченні. У випадку, коли таке

словосполучення встановити неможливо, речення аналізується зліва направо, починаючи з перших повнозначних слів.

Для вибраного словосполучення формується синтаксичний визначник, який складається зі службових слів та флексивних частин повнозначних слів словосполучення. Наприклад, для виразу «права та свободи» таким визначником буде конструкція типу «-а та -и». Якщо сформований визначник існує в базі даних визначників, йому буде відповідати список кореляторів. Тоді, враховуючи семантичні ознаки слів у словосполученні, у базі даних з кореляторами знаходиться потрібний корелятор, що і встановить зв'язок між словами. Однозначність визначення такого зв'язку забезпечується тим, що для окремого визначника множини пар семантичних ознак для кореляторів з його списку не перетинаються.

Далі до словосполучення поступово приєднуються прилеглі повнозначні слова речення, шляхом встановлення зв'язку між новим словом та одним із слів опрацьованої частини речення. Таким чином створюється група зв'язаних слів. Важливим є вибір слова з групи зв'язаних слів, яке буде зв'язуватись з наступними словами. Це має бути або головне, або останнє слово групи. У випадку, коли неможливо встановити зв'язок між новим словом та словами групи, створюється нова група зв'язаних слів. В кінці аналізу всі створені групи поєднуються, за можливістю, в одну, яка відображає структуру зв'язків між всіма словами речення.

Неможливість встановлення зв'язку між окремими групами в реченні та їх об'єднання свідчить або про складне речення, частини якого не пов'язані (або пов'язані неявно) між собою, або про некоректні зв'язки між словами у групах. Для уникнення проблеми некоректного зв'язування слів необхідно проводити додатковий аналіз можливих зв'язків кожного наступного слова із словами групи і вибирати найбільш вірогідний, або ж розглядати всі можливі варіанти зв'язків (що недоцільно, враховуючи зростання кількості таких варіантів для кожного наступного слова).

#### *Другий етап.*

На другому етапі вирішується задача встановлення анафоричних зв'язків або заміни займенників у тексті на відповідні поняття (як правило іменники) [33]. Заміщенню підлягають деякі особові, відносні, вказівні, присвійні та зворотні займенники. Алгоритм заміщення будується на основі аналізу закономірностей (відображених в базі знань синтаксису природної мови) вживання займенників в природній мові і відрізняється в залежності від типу та граматичних характеристик займенника.

Із практики відомо, що поняттям, якому відповідає займенник, є узгоджене з ним за граматичними характеристиками найближче повнозначне слово, яке стоїть попереду, або слово, що входить до ядра даного чи попередніх речень. Тому вихідними даними для проведення заміни займенників є граматичні характеристики слів (отримані в результаті морфологічного аналізу) та синтаксично-семантичні зв'язки між ними (результат роботи синтаксичного блоку). Заміщення займенників дозволяє отримати додаткові зв'язки між відповідними поняттями у реченні та у

всьому тексті, тому після цього проводиться корекція синтактико-семантичних зв'язків у окремих реченнях (із займенниками).

Результатом роботи блоку синтактико-семантичного аналізу є поняттєва (логіко-семантична) структура ПМО.

### **3.9. Моделі представлення багатослівних термінів предметної області**

Задача виділення термінів з тексту виникає при побудові онтологій, формуванні та поповненні словників предметної області, рубрикаторів та будь-якій інтелектуальній обробці текстових документів.

Для побудови понятійного апарату з текстів ПдО необхідно використовувати модель субстантивних іменникових словосполучень, що виражаються схемою: іменник + слово, що погоджується з іменником. У цій моделі іменник є головним словом, а слово, що погоджується, - залежним і може виражатися як прикметником, так і іменником [34]. У загальному випадку субстантивні іменникові словосполучення можуть включати в свій склад наступні класи слів: іменники, прикметники, прийменники, сурядні союзи і прислівники.

Кількість слів в іменникових словосполученнях коливається від двох до п'ятнадцяти і в середньому складає три слова [35]. У російській мові синтаксична структура термінів ПдО більш ніж в 90 відсотків випадків має конструкцію, що відповідає наступним п'яти шаблонам [36]:

1. Одиночні іменники, прикметники, і скорочення;
2. Іменникова група: іменник + іменник в родовому відмінку;
3. Іменникова група: прикметник + іменник
4. Іменникова група: прикметник + прикметник + іменник
5. Іменникова група: іменник + прикметник + іменник в родовому відмінку.

У роботі [37] приводиться 18 шаблонів іменникових словосполучень, які можна використовувати для виділення термінів ПдО.

Разом з тим існують складні словосполучення, що використовуються для позначення понять і термінів і складаються з більше трьох значимих слів. Експерименти по виділенню термінів показали, що в українській мові для стилю «ділова проза» доцільно збільшити кількість слів у шаблонах синтаксичної структури іменникових груп до п'яти. Словосполучення довжиною в п'ять слів використовуються в назвах організацій, у визначенні понять, що відносяться до фінансово-економічної сфери діяльності організацій. Позначення понять і термінів словосполученнями довжиною більше п'яти слів, з використанням сполучників і прийменників зустрічається рідко, особливо такими словосполученнями, у яких частини мови не чергуються (наприклад, прикметник + прикметник + прикметник + іменник + іменник у родовому відмінку).

Структура шаблонів словосполучень, які доцільно використовувати для пошуку термінів, наведені в Таблиці 3.1.

## Структура шаблонів словосполучень пошуку термінів

№	Структура шаблону словосполучення	Приклад терміну
1	Абревіатура	СНГ; МВФ
2	Іменник	Система; держава
3	Іменник+ іменник в родовому відмінку	Президент України; Голова держави
4	Прикметник+ іменник	Конституційний Суд; економічне зростання
5	Іменник +іменник в родовому відмінку + іменник в родовому відмінку	Міністр фінансів України; Указ Президента України
6	Прикметник + іменник + іменник в родовому відмінку	Фінансова система держави; Верховна Рада України
7	Іменник+ прикметник в родовому відмінку + іменник в родовому відмінку	Органи виконавчої влади; Девальвація національної валюти
8	Прикметник + прикметник + іменник	Середньомісячна заробітна плата
9	Іменник + іменник в родовому відмінку + іменник в родовому відмінку + іменник в родовому відмінку	Постанова Кабінету Міністрів України; архів Міністерства оборони України
10	Прикметник + іменник + іменник в родовому відмінку + іменник в родовому відмінку	Національна програма імунопрофілактики населення
11	Іменник+прикметник в родовому відмінку + іменник в родовому відмінку + іменник в родовому відмінку	Кошти Державного бюджету України; Голова Верховної Ради України
12	Прикметник+ прикметник + іменник + іменник в родовому відмінку	Державна податкова адміністрація України
13	Іменник + іменник в родовому відмінку + прикметник в в родовому відмінку + іменник в родовому відмінку	Система використання бюджетних коштів; орган управління державною службою
14	Прикметник +іменник + прикметник в родовому відмінку + іменник в родовому відмінку	Центральні органи виконавчої влади
15	Іменник + прикметник в родовому відмінку + прикметник в родовому відмінку + іменник в родовому відмінку	Зона надзвичайної екологічної ситуації наслідки світової фінансової кризи

<b>№</b>	<b>Структура шаблону словосполучення</b>	<b>Приклад терміну</b>
16	Прикметник + прикметник + прикметник + іменник	Незалежний державний фінансовий контроль
17	Прикметник+ іменник + іменник в родовому відмінку + іменник в родовому відмінку + іменник в родовому відмінку	Фінансово-економічна діяльність Міністерства оборони України
18	Іменник+ прикметник в родовому відмінку + іменник в родовому відмінку + іменник в родовому відмінку+ іменник в родовому відмінку	Комітет фінансового контролю Міністерства фінансів
19	Прикметник+ прикметник+ іменник + іменник в родовому відмінку+ іменник в родовому відмінку	Знецінені грошові заощадження громадян України
20	Іменник+ іменник в родовому відмінку + прикметник в давальному відмінку + прикметник в давальному відмінку+іменник в давальному відмінку	Фонд сприяння молодіжному житловому будівництву;
21	Прикметник + іменник + прикметник в родовому відмінку + іменник в родовому відмінку + іменник в родовому відмінку	Спеціальний фонд державного бюджету України
22	Іменник+ прикметник в орудному відмінку + прикметник в орудному відмінку + іменник в орудному відмінку + іменник в родовому відмінку	Управління державними фінансовими ресурсами України
23	Іменник + іменник в родовому відмінку + іменник в родовому відмінку+ прикметник в родовому відмінку + іменник в родовому відмінку	Джерело фінансування дефіцита Державного бюджету; реформа системи контролю державних фінансів
24	Прикметник + прикметник + прикметник+ іменник + іменник в родовому відмінку	Державний позабюджетний Пенсійний фонд України
25	Прикметник+ іменник + іменник в родовому відмінку + прикметник в родовому відмінку + іменник в родовому відмінку	Головний розпорядник коштів Державного бюджету
26	Іменник + прикметник в родовому відмінку + іменник в родовому відмінку + прикметник в родовому відмінку + іменник в родовому відмінку	Контроль економічної діяльності комерційних організацій
27	Прикметник+ прикметник+ іменник+ прикметник в родовому відмінку + іменник в родовому відмінку	Міжнародний аудиторський контроль фінансової діяльності

№	Структура шаблону словосполучення	Приклад терміну
28	Іменник+ іменник в родовому відмінку + прикметник в орудному відмінку + прикметник в орудному відмінку + іменник в орудному відмінку	Система управління державними фінансовими ресурсами
29	Прикметник + іменник + прикметник в родовому відмінку + прикметник в родовому відмінку + іменник в родовому відмінку	Державне регулювання зовнішньої економічної діяльності
30	Іменник +прикметник в родовому відмінку+ прикметник в родовому відмінку + прикметник в родовому відмінку + іменник в родовому відмінку	Орган зовнішнього державного фінансового контролю
31	Іменник + іменник в родовому відмінку+ іменник в родовому відмінку+ іменник в родовому відмінку+ іменник в родовому відмінку	Програма діяльності Кабінета Міністрів України

### 3.10. Алгоритм автоматичного виділення багатослівних термінів з природномовних об'єктів предметної області

Розглянемо алгоритм автоматичного виділення однослівних та багатослівних термінів, який, крім шаблонів, використовує результати синтактико-семантичного аналізу тексту [38]. Розпізнання поверхневих семантичних відношень здійснюється на попередньому етапі аналізу тексту.

Процедура виділення багатослівних термінів поділяється на два етапи. На першому відбувається безпосередній пошук у тексті слів та словосполучень – кандидатів у терміни. В якості однослівних термінів вибираються іменники та аббревіатури. Багатослівні терміни формуються за допомогою визначених типів відношень між словами речення, шляхом поступового приєднання слів до однослівного терміну-іменника. Для термінів іменникового типу використовуються такі типи відношень між словами: «об'єктне», «приналежність» (між двома іменниками), «визначальне» (між прикметником та іменником), «однорідні слова» (між двома іменниками або двома прикметниками) та інші. Виділені групи слів перевіряються на відповідність визначеним шаблонам іменникових словосполучень, загальна кількість яких складає 31.

Шаблони термінів мають вигляд послідовності морфологічних характеристик слів (наприклад, іменник\_прикметник\_іменник – шаблон для терміну «онтологія предметної області»). Слова, що складають термін, в реченні не обов'язково стоять поруч, але обов'язково зв'язані між собою певним типом відношень. Це дозволяє, наприклад, з речення «Побудова онтології зазначеної предметної області» виокремити термін «онтологія

предметної області». Якість виділення таким способом термінів в значній мірі залежить від якості роботи синтактико-семантичного блоку, а саме, від коректності класифікації слів за семантичними ознаками та встановлення зв'язків між словами.

На другому етапі список кандидатів у терміни фільтрується: враховується значимість виділених словосполучень (наближеність у дереві розбору до підмета чи присудка речення) та частота, з якою вони зустрічаються у тексті. Відфільтрований список термінів виводиться на користувачу.

### *Логіко-семантичні відношення в ПМО*

Для виявлення логіко-семантичних відношень в ПМО використовуються спеціальні діагностичні конструкції фраз та встановлені на етапі синтаксичного аналізу семантично-синтаксичні зв'язки між словами. Основні типи відношень логіко-семантичної структури ПМО наступні: логічні відношення тотожності, включення, перетину та диз'юнкції, ієрархічні відношення, відношення агрегації та функціональні відношення [24].

*Ієрархічні відношення* належать до області, що може бути виражена семантикою дієслова «бути», і поділяються на три підвиди: рід – вид; ознака – значення; інваріант – варіант. Відношення *рід – вид* в свою чергу поділяються на структурні, функціональні та семантичні.

Для розпізнавання родовидових відношень використовують наступні основні конструкції фрази:

- «А належить до В», «А є В з ознаками», «поруч з А є В», «А належить до класу/сімейства В» - для пошуку родового поняття В і видового поняття А;

- «А має наступні види: В<sub>1</sub>, В<sub>2</sub>...»; «до видів А належать В<sub>1</sub>, В<sub>2</sub>...» - для пошуку видових понять В і родового поняття А.

Діагностичні конструкції відношення *ознака – значення* наступні (для пошуку А – поняття-сутності та В – імені сутності): «А - В», «А має ім'я В», «А йменується В», «А називається В».

Відношення *інваріант – варіант* знаходять за допомогою фраз «до варіантів А належать В<sub>1</sub>, В<sub>2</sub>...» та «інваріантом А є В».

*Відношення агрегації* належать до області, що може бути виражена семантикою дієслова «мати». Виділяють чотири підвиди: ціле - частина (компонент); об'єкт – простір реалізації об'єкта (локалізації, позиції); об'єкт – властивості/ознака; рівень – одиниця рівня.

Основні діагностичні конструкції для пошуку відношень агрегації:

- для цілого В та частини А: «А входить до складу В», «А складає частину В», «А є частиною В»;

- для частини В - компоненту А: «А складається з В-в», «А містить в собі В як частину», «А містить в своєму складі В як частину»;



- для відношення *об'єкт – простір реалізації об'єкта (локалізації, позиції)*: «місце А – В», «А відбувається в В», «А знаходиться в В», «А входить до В»;

- для відношення *об'єкт – властивості/ознака*: «А має властивість В, має В, як суттєву ознаку», «А характеризується наявністю В», «для А характерно В»;

- для відношення *рівень – одиниця рівня*: «А належить до рівня В», «А – одиниця рівня В», «А розглядається на рівні В».

*Функціональні відношення* – це відношення, які належать до сфери опису процесів. Таке відношення представляє собою багатомісний предикат, що описує деякий процес. У цьому предикаті виділяють наступні змінні: А – суб'єкт деякої сутності; В – інструмент, спосіб, метод або алгоритм виконання сукупності операцій процесу; С – початковий об'єкт або вихідні дані; D – кінцевий об'єкт або кінцеві дані; Е – подія, що активує операцію.

Основні діагностичні вирази:

- для виявлення суб'єкта: «А здійснює В»;

- для інструменту, способу, методу, або алгоритму – тут можуть бути вказані різномірні відповідні терміни;

- для виявлення початкового об'єкта: «А виконується над В», «А застосовується до В», «А змінює В», «А В-є»;

- кінцевого об'єкта: «внаслідок А виходить В», «внаслідок А утворюється В», «А призводить до В», «А перетворює В<sub>1</sub> у В<sub>2</sub>».

## РОЗДІЛ 4. МЕТОДОЛОГІЧНІ ОСНОВИ СТВОРЕННЯ СИСТЕМ УПРАВЛІННЯ ЗНАННЯМИ

### 4.1. Особливості систем управління знаннями

З точки зору системного аналізу, система становить певне цілісне середовище, яке з позицій досягнення цілей вибирає, формує та створює людина. В праці всі можливі системи діляться на дві підмножини – «жорсткі» і «м'які» системи. Теорія жорстких систем вимагає строгих кількісних побудов. У теорії м'яких систем розглядаються системи, здатні адаптуватися до умов зовнішнього середовища, зберігаючи при цьому свої характерні риси. М'які системи, піддаючись тривалим змінам, зберігають свою внутрішню сутність і здатність до розвитку. Відповідно до класифікації проведеної П.Чеклендом соціальні системи разом з біологічними і психологічними відносяться до м'яких. В цій роботі також підкреслюється, що проблеми, з якими стикається людина, не можуть бути вирішені раз і назавжди. На відміну від парадигми жорстких систем (системи, створені людиною або природою), що розглядає дійсність як системну, парадигма м'яких систем розглядає світ як проблематичний, можливо системний, але слабкоструктурований, що припускає різні інтерпретації.

В праці відзначається, що в дослідженні м'яких систем не можна використовувати тільки формальні методи, значну роль грають евристичні міркування й інтуїція. Дослідник змушений робити висновки на підставі невеликого числа спостережень, які, як правило, не можуть бути відтворені. Для цього йому слід використовувати методологію м'яких систем (ММС). Застосовуючи ММС, він повинен не тільки правильно описати поведінку системи, але й передбачати позицію включеного в неї людського фактора. Таким чином, людський вимір у ММС грає досить помітну роль, внаслідок чого в аналізі м'яких систем ММС поряд з апробованим інструментарієм можуть знайти досить широке застосування евристичні методи й моделі, особливо ті, що враховують притаманну таким системам нечіткість.

Ці міркування стосуються, зокрема, *систем управління знаннями* (СУЗ). Вибір (ухвалення) рішення щодо подальших дій в таких системах має зробити людина, яка є експертом в певній предметній області.

Однак існують природні межі людської здатності сприймати й обробляти інформацію. Роботу експертів лімітують як міжособистісні відносини, так і внутрішні психологічні й фізіологічні причини. До того ж людина здатна одночасно оперувати лише обмеженою кількістю операндів (понять, ідей, моделей, альтернатив); користуючись термінологією психологів, має місце так званий «закон сім плюс-мінус два». Це обмеження можна подолати шляхом введення ієрархії оцінювання ситуації та прийняття рішень, яка є природною та притаманна таким системам, як СУЗ.

Необхідно відзначити, що СУЗ притаманні всі характерні риси, які виділені в працях, як характерні ознаки складних систем:

1) *Унікальність*. Унікальним є кількість та види взаємодії елементів системи. Наслідком унікальності є неможливість застосування однакових математичних моделей для моделювання складних систем.

2) *Відсутність строго формалізованих та єдиних глобальних цілей існування*. Відсутність строгої формалізації робить неможливою чітку математичну постановку задачі, яка б враховувала всі необхідні параметри в чисельному вигляді для забезпечення адекватності моделі. Це викликає необхідність оперування з лінгвістичними формулюваннями, нечіткими постановками задач, до того ж багато результуючих параметрів є випадковими, а оцінка стану та особливостей роботи об'єкта проводиться експертами, виходячи виключно з власного досвіду.

3) *Багатокритеріальність або відсутність критеріїв оптимальності*. Сама по собі багатокритеріальність у складних системах вносить невизначеність. Відсутність критеріїв оптимальності, породжує множинність рівнозначних рішень, стосовно яких важко зробити оптимальний вибір.

4) *Відкритість та динамічність*. Як правило, протягом свого функціонування складні системи, засновані на знаннях, не є незмінними. Складні системи високоорганізовані і можуть змінюватись і еволюціонувати під впливом внутрішніх параметрів.

5) *Неповнота та невизначеність*, що викликана різними причинами:

– неповнота або недостатність знань про природу процесів в складних системах;

– лінгвістична невизначеність, що виникає внаслідок розпливчатості і неоднозначності словесних виразів;

– необхідність врахування одночасно кількісної і якісної інформації;

– наявність суб'єктивних факторів, яка пов'язана з тим, що в СУЗ на кожному рівні ієрархії рішення приймаються або виконуються людиною.

В процесі моделювання та управління СУЗ необхідно враховувати невизначеність, що обумовлена їх властивостями. Присутність всіх типів невизначеності не обов'язкова для кожної СУЗ, але кілька типів невизначеності завжди мають місце. Наприклад, ще не винайшли технічний засіб, який би дозволяв виміряти соціальні явища.

Зазначимо, що під невизначеністю мається на увазі властивість інтерпретації виразів коли їм приписується оцінки правдоподібності відмінні від абсолютної істини і неправди. А під лінгвістичною невизначеністю мається на увазі невизначеність, що виникає внаслідок розпливчатості та неоднозначності словесних виразів.

Слід додати, що описання вхідної та результуючої інформації в СУЗ важко піддається формалізації, тому що логіка прийняття рішення та сама система управління визначається та формується виходячи з експертних знань.

Труднощі розробки й ефективного використання інформаційних методів полягають також у тому, що для СУЗ дуже складно визначити цілі в конкретних цифрах. Цілі можуть бути або тільки позитивними, тобто спрямованими на досягнення нового, або також і негативними,

спрямованими на запобігання небажаного розвитку або невірному рішенню . У такому випадку при побудові узагальненої характеристики об'єкту управління має місце протиріччя цілей та наявність певних невизначеностей. Результати діяльності, таким чином, не можна визначити точно.

Для невірних об'єктів, до яких відносяться СУЗ, характерна наявність не функціональних цілей, що характеризують якість, виміряти яку дуже складно, а іноді взагалі неможливо. Організація пов'язана з великою кількістю зв'язаних груп, кожна з яких характеризується власною динамікою й власними цілями, які не завжди збігаються із загальною метою функціонування.

Разом з тим не завжди можна визначити ціль у вигляді кількісних показників. У такому випадку необхідно поряд з кількісними розробити систему якісних показників. Актуальність цілей необхідно перевіряти в міру зміни основних впливових параметрів. Нестабільність різних зовнішніх і внутрішніх явищ змушує гнучко ставитися до поставлених завдань. Таким чином, цілі і їхні пріоритети не є постійними величинами, вони можуть час від часу переглядатися у зв'язку зі змінами зовнішніх і внутрішніх умов. Отже, слід провести аналіз математичних методів, які враховують описані вище особливості та які можна використовувати при обробці інформації.

Зазначимо, що існуючі програмні системи дозволяють аналізувати дані, що зберігаються в базах даних, однак для роботи з ними необхідні спеціальні знання в області інформаційних технологій. Це значить що не підготовлений в галузі інформаційних технологій фахівець не зможе використовувати цю систему, або лише в досить обмеженому функціонально обсязі. У таких ситуаціях необхідно виділяти фахівця, який би працював із цією програмою, що не завжди прийнятно або можливо.

До того ж існуючі програмні модулі не дозволяють будувати та використовувати математичні моделі та методи, які б враховували природну ієрархічність складних систем управління знаннями, а також те, що такі системи працюють в умовах невизначеності. Розробка нових програмних модулів обробки інформації (ПМОІ) для певної установи потребує значних фінансових та людських ресурсів, до того ж математичні моделі прив'язані до конкретної предметної області і розроблені тільки для неї. Тому необхідно провести аналіз існуючих технологій створення ПМОІ, таких щоб побудовані за допомогою них ПМОІ враховували всі особливості проведення обробки інформації в СУЗ, були гнучкими до змін середовища та надавали засоби для проведення обробки інформації користувачем, який не є спеціалістом в галузі інформаційних технологій (ІТ) за рахунок ефективної побудови та використання відповідних математичних моделей.

## 4.2. Систематизація і аналіз методів обробки інформації в системах управління знаннями

На різних рівнях функціонування систем управління знаннями, та в залежності від обраної для оцінювання властивості необхідні різні математичні моделі, що відрізняються рівнем абстракції, кількістю параметрів, що враховуються, необхідними комп'ютерними ресурсами і т.п.

Впровадження теоретичних результатів математичного моделювання у повсякденну практику суттєво відстає від існуючих потреб. Це пояснюється високою складністю об'єктів, що впливає на складність і трудомісткість роботи з математичними моделями.

Виникає питання про можливість використання класичних методів для побудови моделей для оцінювання інформації та формування нових знань.

Вітчизняні та закордонні вчені досягли значних результатів у розв'язанні задач аналізу жорстких складних систем та методів і алгоритмів побудови їх математичних моделей. Зокрема ці проблеми розглядаються у роботах В.М. Глушкова, М.З. Згуровського, О.А. Павлова, Н.Д. Панкратової, Ю.П. Зайченка, І.В. Сергієнка, Г.Є. Цейтліна, О.М. Новикова, Е.А. Трахтенгерца, С.Ф. Теленика, В.М. Томашевського, С.М. Гриши, Т. Кохонена, Т. Сааті та інших учених.

Кількісні методи, які добре працюють при описі технічних систем та певних знань щодо них, які піддаються формалізації, у силу певних недоліків не можуть бути використані для опису систем управління знаннями. Це пов'язано насамперед з тим, що інформація в таких системах може мати не тільки кількісний вигляд, але і надаватись у вигляді знань – правил, евристик, наборів ранжованих альтернатив, семантичних оцінок кількісних параметрів і інших характеристик, які описуються за допомогою значень лінгвістичних змінних (ЛЗ), висловлювань і т.д.

Оскільки інформація характеризується тим чи іншим ступенем невизначеності, необхідно використовувати такі математичні методи і підходи, які включають засоби для представлення і оперування інформацією за наявності різних типів невизначеності у описанні об'єкту, а саме лінгвістичної невизначеності та необхідності врахування як кількісної так і якісної інформації. Серед таких методів та підходів можна виділити три великі групи: логічні, мережеві та еволюційне моделювання (таблиця 4.1).

Проведений аналіз методів та підходів до вирішення задач моделювання та обробки інформації в СУЗ дає уявлення про їх широкий спектр. Кожен із розглянутих напрямків має свої переваги і недоліки.

З наведених трьох основних типів методів лише логічні та мережеві методи можуть використовуватися для побудови моделі обробки інформації в умовах притаманних СУЗ, оскільки методи еволюційного моделювання не мають власної моделі знань.

Методи еволюційного моделювання є за своєю природою оптимізаційними методами та можуть використовуватися при побудові та настроюванні структури моделі.

**Методи та підходи, які можна використовувати для вирішення задач моделювання та обробки інформації в СУЗ**

Тип підходу	Підхід
логічні методи	звичайне використання предикатів
	логіка немонотонних міркувань
	логіки ґрунтовані на мінімальних моделях
	стенфордська теорія фактору впевненості
	логіки, що базуються на методі виведення Демпстера-Шейфера
	логіки, що базуються на суб'єктивному байєсівському виведенні
	системи нечіткої логіки
	логічні засади фізичних процесів (методи якісної фізики)
мережеві методи	активна семантична мережа
	спінові інформаційні моделі
	нейронна мережа
	нечітка нейронна мережа
методи еволюційного моделювання	генетичні алгоритми
	еволюційні стратегії
	еволюційне програмування

Додатково зазначимо, що серед розглянутих в таблиці 4.1 логічних методів найбільш придатними для моделювання в СУЗ є такі: звичайне використання предикатів, системи нечіткої логіки, логічні засади фізичних процесів. Оскільки при побудові моделі необхідно використовувати евристичні та експертну інформацію, яка подається у вигляді правил сформованих природною мовою у вигляді, найчастіше, продукційних правил, то в якості моделі для оцінки слід використовувати логічну модель.

З метою виділення саме тих методів та підходів, які слід застосовувати для моделювання в СУЗ запропоновано такі критерії:

- можливість використання в умовах лінгвістичної невизначеності та необхідності врахування як кількісної так і якісної інформації;
- можливість навчання;
- можливість використання евристик та експертних знань;
- врахування природної ієрархічності СУЗ;
- надання пояснення отриманого результату;
- наявність власної моделі знань;
- виявлення найвпливовішого параметру.

Виходячи з цих критеріїв проведено аналіз математичних методів та підходів з метою їх застосування для логічного моделювання в СУЗ, результати якого наведено в таблиці 4.2.

Як видно з цього аналізу жоден з методів в повному обсязі не

задовольняє всім висунутим критеріям, однак найбільш придатними для застосування є системи, що базуються на нечіткій логіці та нечіткі нейронні мережі. Тому слід більш детально розглянути ці методи з метою виявлення існуючих недоліків та запропонувати підходи, щодо вдосконалення цих методів та їх комплексного використання.

Для моделювання швидкоплинних локальних у часі процесів застосування математичних методів і моделей, які потребують навчання для виявлення закономірностей функціонування системи неможливо. У цьому випадку основним способом формування моделі в умовах невизначеності стає формалізація експертних знань і представлень у вигляді баз знань (БЗ), використовуючи методи нечіткої логіки. Основна ідея полягає в тому, що настроюючи базу нечітких знань (БНЗ), можна ідентифікувати нелінійні залежності з необхідною точністю.

Перевагою даного підходу є також можливість урахування недетермінованих впливів, структурних і якісних змін у складі системи в процесі моделювання. Також дуже корисною особливістю цього підходу є можливість представлення результатів моделювання у відносно простій і наочній формі, яка буде зрозумілою людині. До недоліків цього підходу можна віднести невпорядкованість та надлишковість початкової БНЗ, яка формується виходячи з експертних знань, та внаслідок цього важкість відповідей на питання «чому?» і «як?», тобто неможливість визначення всієї ієрархії параметрів, що впливають на значення виходу системи.

Це ускладнює процес прийняття рішень, щодо необхідних напрямків управління знаннями.

При побудові моделей для проведення обробки інформації в СУЗ з відносно постійним складом учасників на середньо- і довгостроковому часовому інтервалі з'являється можливість застосовувати в якості вихідних даних моделі більш об'єктивну інформацію, а саме інформацію про стан системи в періоди її стійкості. Тому додатково до сформованої за допомогою методів нечіткої логіки моделі можна використовувати методи теорії штучних нейронних мереж та еволюційних обчислень для проведення настроювання. Такий підхід дає можливість значно зменшити суб'єктивність і суперечливість експертних оцінок і внаслідок цього поліпшити якість прогнозування. Тому актуальною є задача розробки підходу до обробки інформації, до складу якого повинні входити метод обробки інформації на основі деревоподібних баз нечітких знань, моделі та алгоритми, що дозволяють впорядковувати БЗ введенням ієрархії правил, проводити настроювання БЗ у відповідності до вхідних даних та простежувати шлях виведення в зворотному напрямку з метою визначення всієї послідовності вершин, які вплинули на остаточне значення результуючої змінної.

## Методи вирішення задач логічного моделювання та оцінювання в СУЗ

Тип підходу	Критерій/ Підхід	використання в умовах лінгвістичної невизначеності	Можливість навчання	Використання евристик та експертних знань	Враховування природної ієрархічності СУЗ	Надання пояснення отриманого результату	Наявність власної моделі знань	Виявлення найвпливовішого параметру
логічні методи	звичайне використання предикатів	-	-	+	+/-	+/-	+	-
	системи нечіткої логіки	+	-	+	+/-	+/-	+	-
	логічні засади фізичних процесів	+/-	-	+	+/-	+/-	+	-
Мережеві методи	активна семантична мережа	-	+	-	+/-	-	+	-
	спінові інформаційні моделі	+	-	+	+/-	-	+	-
	нейронна мережа	-	+	-	-	-	+	-
	нечітка нейронна мережа	+	+	+	+/-	-	+	-

«-» – не задовольняє критерію

«+» – задовольняє критерію

«+/-» – задовольняє критерію не в повному обсязі

### 4.3. Обробка інформації на основі підсистеми нечіткого виведення загального виду

Для вирішення задач представлення знань в умовах невизначеності в працях пропонується будувати системи нечіткого виведення на основі баз нечітких знань. БНЗ в таких системах представляють собою сукупність лінгвістичних висловлювань типу ЯКЩО <входи>, ТО <виходи> та є частиною підсистем нечіткого виведення (ПНВ).



### Підсистема нечіткого виведення загального виду

В загальному випадку ПНВ, що використовується для проведення обробки інформації, складається з трьох блоків (рисунок 4.1).

- блоку кодування (БКД) – фаззифікатора;
- блоку обробки та виведення (БОВ), який є внутрішньою частиною нечіткої системи, та складається з БЗ та блоку нечіткого виведення (БНВ);
- блоку декодування (БДК) – дефаззифікатора.

Блок кодування та блок декодування ще називаються інтерфейсами нечіткої моделі, через які вона контактує з зовнішнім середовищем. На вхід БКД подається кількісна інформація  $\{x_i | i = \overline{1, v}\}$ , якій на виході ставиться у відповідність сукупність лінгвістичних змінних.

$$\text{БКД} : x_i \xrightarrow{\text{фаззифікація}} X_i \quad (4.1)$$

де  $X_i$  – лінгвістична змінна, що відповідає вхідній змінній  $x_i$ , та яка в даному випадку означає назву множини термів. Кожен терм розглядається як нечітка множина, якщо  $x_i$  кількісна змінна визначена на неперервній універсальній множині, або (4.1), якщо  $x_i$  якісна або кількісна змінна визначена на скінченій універсальній множині.

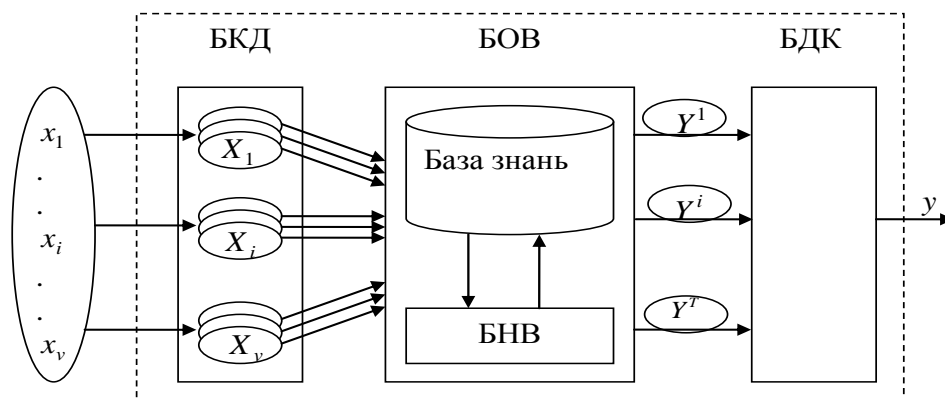


Рис. 4.1. Загальна структура підсистеми нечіткого виведення

$$X_i^p = \int_{x_i}^{\overline{x_i}} \mu^{X_i^p}(x_i) / x_i, \quad p = \overline{1, d_i}, \quad i = \overline{1, v} \quad (4.2)$$

де  $\mu^{X_i^p}(x_i)$  – функція належності значення вхідної змінної  $x_i \in [\underline{x_i}, \overline{x_i}]$  терму  $X_i^p \in X_i$ ,  $p = \overline{1, d_i}$ ,  $i = \overline{1, v}$ .

$$X_i^p = \sum_{k=1}^{q_i} \mu^{X_i^p}(u_{ik}) / u_{ik}, \quad p = \overline{1, d_i}, \quad i = \overline{1, v} \quad (4.3)$$

де  $\mu^{X_i^p}(u_{ik})$  – ступінь належності елемента  $u_{ik} \in U_i$  терму  $X_i^p \in X_i$ ,  $p = \overline{1, d_i}$ ,  $i = \overline{1, v}$ .

Далі ці множини  $\{X_i | i = \overline{1, v}\}$  подаються на вхід БОВ, де дані поступають до БНВ. Виходячи з отриманих даних та тих правил, що зберігаються в БНЗ, БНВ здійснює нечітке виведення з отриманням на виході тих нечітких множин, які відповідають термам лінгвістичної змінної  $Y$ , що відповідає модельованій змінній предметної області  $y$  (4.4) заданої користувачем для визначення. Лінгвістична змінна  $Y$  також відповідає термножині, яка визначається (4.5), якщо  $y$  кількісна змінна визначена на неперервній універсальній множині та (4.6) якщо якісна або кількісна змінна визначена на скінченій універсальній множині.

$$\begin{aligned} \text{БОВ} : \{X_i\} &\xrightarrow{\text{нечітке логічне виведення}} Y \\ Y &= \{Y^j | j = \overline{1, T}\} \end{aligned} \quad (4.4)$$

де  $Y^j$  - терм лінгвістичної змінної  $Y$ ,  $T$  - кількість термів.

$$Y^j = \int_{\underline{y}}^{\overline{y}} \mu^{Y^j}(y) / y, \quad j = \overline{1, d_Y} \quad (4.5)$$

де  $\mu^{Y^j}(y)$  - функція належності значення результуючої змінної  $y = [\underline{y}, \overline{y}]$  терму  $Y^j \in Y$ ,  $j = \overline{1, d_Y}$ .

$$Y = \{Y^1, Y^2, \dots, Y^{d_Y}\} \quad (4.6)$$

де  $Y$  - терм-множина змінної  $y$ ,  $Y^j$  -  $j$ -й лінгвістичний терм змінної  $y$ ,  $j = \overline{1, d_Y}$ .

Терм  $Y^j$  є також нечіткою множиною.

З БОВ множина  $Y$  передається на вхід дефаззифікатора БДК, який перетворює отриманні нечіткі дані, в формат, який відповідає формату даних модельованої змінної  $y$  (4.7).

$$\text{БДК} : Y \xrightarrow{\text{дефаззифікація}} y \quad (4.7)$$

Для випадку підсистеми, що описана вище, задачу обробки інформації з використанням нечіткої моделі можна описати відношенням (4.8).

$$\text{ПНВ} : \{x_i | i = \overline{1, v}\} \longrightarrow y \quad (4.8)$$

Слід відмітити, що БОВ є по суті експертною системою прямого пошуку, тобто такою, яка здійснює міркування від даних. У випадку рішення задач класифікації, коли результуюча змінна є якісною за своєю природою, застосовується нечітке виведення Мамдані, яке здійснюється в чотири етапи, що наведені на рисунку 4.3

### 1. Введення нечіткості.

Знаходяться ступені істинності для передумов кожного правила.

### 2. Нечітке виведення.

Знаходяться рівні «відсікання» для передумов кожного з правил (з використанням операції мінімум), та знаходяться «усічені» функції належності для вихідних змінних.

### 3. Композиція.

Проводиться об'єднання усічених функцій з використанням операції максимум, що призводить до одержання підсумкової нечіткої підмножини для змінної виходу з відповідною функцією належності.

### 4. Приведення до чіткості.

Додатковий етап, який використовується при необхідності перетворення нечіткої множини виведення до чіткого числа.

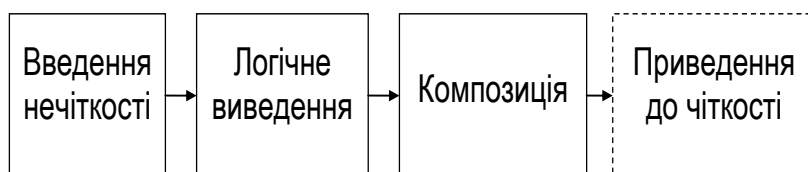


Рис. 4.2. Схема нечіткого виведення

Таким системам притаманні деякі недоліки. По-перше, міркування на основі даних здійснюється менш цілеспрямовано ніж міркування на основі цілі. Причина цього полягає в тому, що в системі, яка оснований на цілі, міркування направлені на рішення конкретної задачі. Ця задача розбивається на більш дрібні, а ці підзадачі, в свою чергу, можуть дробитися й далі. В результаті пошук завжди направлено по ієрархії, що веде від цілі до її складових частин. В системах на основі даних орієнтація на ціль відсутня. Пошук проводиться по дереву, що залежить лише від порядку слідування правил та появи нової інформації. В підсумку процес часто здається розпливчастим та не сфокусованим.

По-друге, пояснення, що видаються користувачу в процесі пошуку, дуже обмежені. В системах на основі правил у відповідь на питання «чому?» користувачу надається лише те правило, що розглядається. Але, якщо послідовність правил точно не відслідковується, то це пояснення не можна розвинути. «Розпливчатість» пошуку на основі даних ускладнює подібні пояснення. І, нарешті, при досягненні цілі системі важко дати вичерпну відповідь на питання «як?». В якості часткового або повного пояснення можна лише використати вміст робочої пам'яті або список активізованих правил. Але ж ці пояснення не надають повного погодженого обґрунтування.

## 4.4. Проблеми використання неупорядкованої бази знань

При заповненні бази знань ПНВ, яка базується на правилах, інженер по знанням (експерт) подає свої знання у вигляді правил «ЯКЩО..., ТО...».

Вхідні дані, в цьому випадку, пов'язані або напряму з загальним об'єктом (властивістю), для класифікації якої використовується ПНВ, або через якісь проміжні властивості. Тоді вирішення задачі класифікації реалізується шляхом виведення на невпорядкованій множині правил (рисунок 4.3). У випадку коли необхідно визначення ступенів належності термів шуканої лінгвістичної змінної, тобто вирішення для неї задачі класифікації, проводиться пошук від даних, який, як було відмічено вище має ряд недоліків. З того що множина правил є невпорядкованою слідує, що не можна бути впевненими у вірності результату доти, доки не були залучені всі правила. Тобто виведення в цьому випадку проходить доти, доки не залучені всі правила і не отримані всі можливі значення функцій належності для кожного терму шуканої лінгвістичної змінної і лише після цього проводиться композиція, тобто знаходження максимального значення функцій належності для кожного терму та дефаззіфікація, яка в випадку, що розглядається, представляє з себе знаходження максимального значення функцій належності серед усіх термів шуканої лінгвістичної змінної. Ще однією проблемою, яка пов'язана з невпорядкованістю правил є можливість наявності петель, тобто можливість зациклення алгоритму виведення.

Простір станів невпорядкованої БНЗ показано на рисунку 4.4 Кожна вершина графу відповідає терму лінгвістичної змінної, причому вершини нижнього рівня відповідають лівій частині правил „ЯКЩО..., ТО...» із зв'язкою ТА між ними, а пов'язана з ними вершина наступного рівня – правій.

$$(P_l^z)_j = \text{ЯКЩО} \left( X_{i_{j1}} = X_{i_{j1}}^{k_{i_{j1}}} \right) \text{ТА} \left( X_{i_{j2}} = X_{i_{j2}}^{k_{i_{j2}}} \right) \text{ТА} \dots \text{ТА} \left( X_{i_{jn}} = X_{i_{jn}}^{k_{i_{jn}}} \right) \quad (4.1)$$

*ТО*  $X_l = X_l^z$  із ступенем впевненості  $\mu_{l_zj}$

де  $(P_l^z)_j$  -  $j$ -е правило ( $j = \overline{1, M}$ ,  $M = M_{l_z}$ ) для визначення  $z$ -го терму ЛЗ з ідентифікатором  $l$ ,

$X_{i_{js}}$  - лінгвістична змінна з ідентифікатором  $i_{js} = i_{l_zjs}$ ,

$s = \overline{1, n_j}$  - номер ЛЗ в лівій частині  $j$ -го правила,

$n_j = n_{l_zj}$  - кількість ЛЗ, що знаходяться в  $j$ -му правилі,

$X_{i_{js}}^{k_{i_{js}}}$  - терм ЛЗ  $X_{i_{js}}$  з номером  $k_{i_{js}}$ .

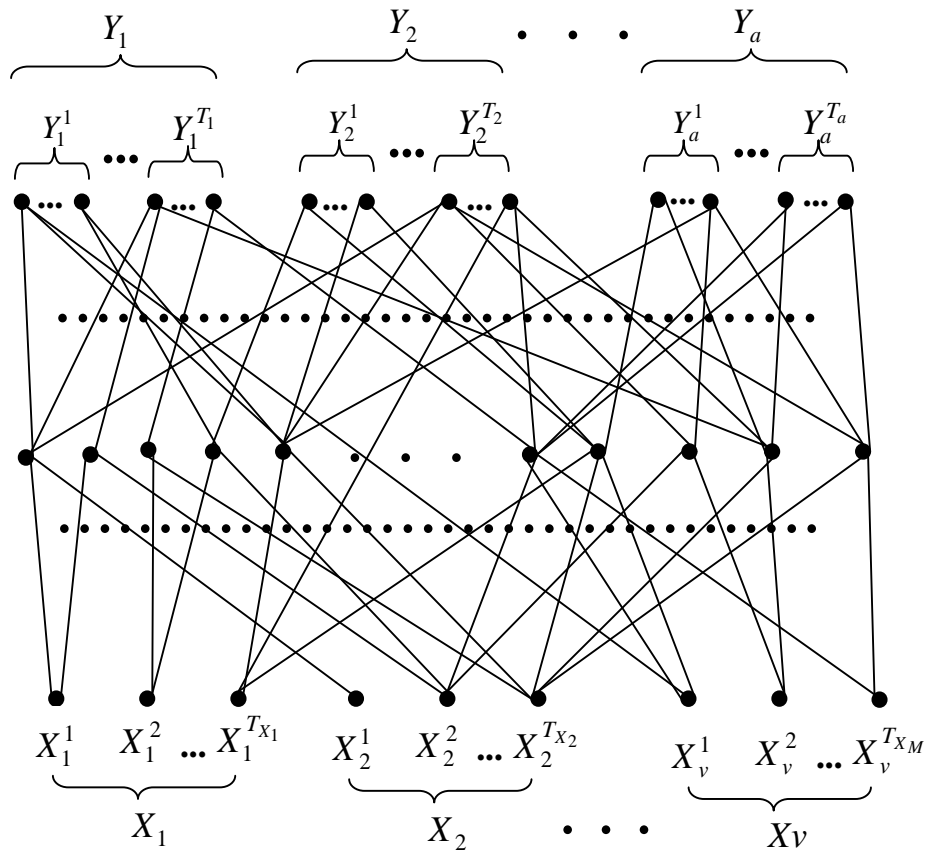


Рис. 4.3. Простір станів неупорядкованої БНЗ

Вершини найнижчого рівня відповідають термам лінгвістичних змінних  $X_i, i = \overline{1, v}$ , які можна виміряти або оцінити експертно напряму. Вершини найвищого рівня представляють собою терми кінцевих лінгвістичних змінних  $Y_r, r = \overline{1, a}$ , такі що не входять в жодну ліву частину жодного правила вигляду „ЯКЩО.., ТО..» (4.7). Тобто існує такий набір  $l_r, r = \overline{1, a}$ , що  $X_{l_r} = Y_r$ .

Оскільки лінгвістична змінна  $X_{i_{l_2 j_2 s_2}}$  може входити в ліву частину різних правил, то можливі ситуації коли  $X_{i_{l_1 j_1 s_1}} = X_{i_{l_2 j_2 s_2}}$ , тобто  $i_{l_1 z_1 j_1 s_1} = i_{l_2 z_2 j_2 s_2}$  при  $l_1 \neq l_2$  та (або)  $z_1 \neq z_2$  та (або)  $j_1 \neq j_2$  та (або)  $s_1 \neq s_2$ .

Оскільки деякі правила можуть мати однакові праві частини, то в просторі станів (рисунок 4.3) вершини можуть повторюватись.

В багатьох при виведенні на основі багаторівневих БНЗ пропонується обчислювати значення функції належності висновку через знаходження мінімуму функцій належності посилок з врахуванням ступеню впевненості, який пов'язаний з правилом. В якості операцій диз'юнкції і кон'юнкції використовуються максимум і мінімум відповідно. Таке виведення аналогічно нечіткому виведенню Мамдані за умови, що лінгвістична змінна, яка знаходиться в правій частині правила «ЯКЩО.., ТО..», має лише якісні значення. В роботі під нечітким виведенням буде розумітись саме нечітке виведення Мамдані.

Для кожного  $j$ -го правила, що відповідає  $z$ -му терму  $l$ -ї лінгвістичної змінної значення функції належності буде обчислюватись за формулою (4.8):

$$(\mu_{X_l^z})_j = \mu_{X_{i_{j1}}^{k_{ij1}}} \wedge \mu_{X_{i_{j2}}^{k_{ij2}}} \wedge \dots \wedge \mu_{X_{i_{jn}}^{k_{ijn}}} \wedge \mu_{l_{zj}}$$

$$X_{i_{jg}}^{k_{ijg}} = X_{i_{l_{zjg}}}^{k_{il_{zjg}}}, \quad n = n_{l_{zj}} \quad (4.2)$$

де  $X_{i_{jg}}^{k_{ijg}}$  –  $k_{ijg}$ -й терм ЛЗ  $X_{i_{jg}}$  з ідентифікатором  $i_{jg} = i_{l_{zjg}}$ ,  
 $n$  – кількість ЛЗ в правилі.

На останньому кроці виведення знаходиться максимальне значення функції належності для кожного терму шуканої змінної, та з цього робиться висновок про результуюче значення, тобто про розв'язок задачі класифікації (4.3).

$$\mu_{Y_w^s} = \mu_{X_{Nw}^s} = \bigvee_{j,h} \left( \mu_{X_{i_{j1}}^{k_{ij1}}} \wedge \mu_{X_{i_{j2}}^{k_{ij2}}} \wedge \dots \wedge \mu_{X_{i_{jn}}^{k_{ijn}}} \wedge \mu_{wsj} \right)_h$$

$$X_{i_{jg}}^{k_{ijg}} = X_{i_{wsjg}}^{k_{iwsjg}}, \quad n = n_{wsj} \quad (4.3)$$

де  $h$  – це номер набору значень функцій належності термів лівої частини правила  $j$ .

В працях показано, що використання ваги правила змінює семантику лінгвістичних змінних, які складають правило, та може статися, що однакові терми лінгвістичних змінних мають різні функції належності, якщо вони знаходяться в різних правилах. До того ж показано, що замість зміни ваги правила при настроювання правил, слід змінювати вигляд функції належності та рекомендовано не використовувати вагу правил. Тоді вирази (4.1, 4.8), слід переписати наступним чином 4.10, 4.11).

$$(P_l^z)_j = \text{ЯКЩО} \left( X_{i_{j1}} = X_{i_{j1}}^{k_{ij1}} \right) \text{ТА} \left( X_{i_{j2}} = X_{i_{j2}}^{k_{ij2}} \right) \text{ТА} \dots \text{ТА} \left( X_{i_{jn}} = X_{i_{jn}}^{k_{ijn}} \right) \quad (4.4)$$

$$\text{ТО } X_l = X_l^z$$

$$(\mu_{X_l^z})_j = \mu_{X_{i_{j1}}^{k_{ij1}}} \wedge \mu_{X_{i_{j2}}^{k_{ij2}}} \wedge \dots \wedge \mu_{X_{i_{jn}}^{k_{ijn}}}$$

$$X_{i_{jg}}^{k_{ijg}} = X_{i_{l_{zjg}}}^{k_{il_{zjg}}}, \quad n = n_{l_{zj}} \quad (4.5)$$

де  $(P_l^z)_j$  –  $j$ -е правило ( $j = \overline{1, M}$ ,  $M = M_{l_z}$ ) для визначення  $z$ -го терму ЛЗ з ідентифікатором  $l$ ,

$X_{i_{js}}$  – лінгвістична змінна з ідентифікатором  $i_{js} = i_{l_{zjs}}$ ,

$s = \overline{1, n_j}$  – номер ЛЗ в лівій частині  $j$ -го правила,

$n_j = n_{l_zj}$  – кількість ЛЗ, що знаходяться в  $j$ -му правилі,

$X_{i_{js}}^{k_{i_{js}}}$  – терм ЛЗ  $X_{i_{js}}$  з номером  $k_{i_{js}}$ .

Формула (4.9) перепишеться наступним чином:

$$\mu_{Y_w^s} = \mu_{X_{Nw}^s} = \bigvee_{j,h} \left( \mu_{X_{i_{j1}}^{k_{i_{j1}}}} \wedge \mu_{X_{i_{j2}}^{k_{i_{j2}}}} \wedge \dots \wedge \mu_{X_{i_{jn}}^{k_{i_{jn}}}} \right)_h$$

$$X_{i_{jg}}^{k_{i_{jg}}} = X_{i_{wsjg}}^{k_{i_{wsjg}}}, \quad n = n_{wsj} \quad (4.6)$$

Відзначимо, що в загальному випадку на кожному рівні простору станів можливі ситуації, коли значення функції належності терму лінгвістичної змінної, що знаходиться в посиланні правила для одного й того ж правила будуть різні, в залежності від того яким чином його було отримано, тобто які значення функцій належності термів лінгвістичних змінних, що знаходяться в посиланні були використані для обчислення значення функції належності результуючого терму.

Серед проблем, що також виникають при використанні невпорядкованої БНЗ, що розглядається, можна виділити наступні. Якщо в нас є не всі вхідні дані, то при виведенні на основі невпорядкованої БЗ буде задіяне кожне з правил, в яких посилання містить не доступні на даний момент дані. Якщо нам треба відстежити який з елементів наступний за впливом на батьківську вершину, то слід буде знову робити повний перебір всіх правил, можливо, виключивши ті, які були використані на попередньому кроці.

Виходом з такої ситуації є розробка підходу до зведення БНЗ до деревоподібної структури, який дозволить зробити пошук більш цілеспрямованим, підходу до модифікації схеми нечіткого виведення з метою відстеження всієї ієрархії параметрів, що вплинули на результуюче значення вихідної змінної та надання відповіді на питання «чому?» і «як?», дослідженню умов, при виконання яких частину БНЗ можна розглядати як класичну логічну, що дозволить удосконалити нечітку логічну модель, та розробка методу обробки інформації на основі деревоподібної БНЗ, який би використовував удосконалення нечіткої логічної моделі через використання змішаної, а саме чіткої і нечіткої, схеми виведення.

Додатково слід провести аналіз існуючих способів завдання функцій належності з метою виділення найкращого для використання при моделюванні в СУЗ.

#### 4.5. Методи та способи завдання функцій належності

За визначенням функція належності характеризує суб'єктивну міру (в діапазоні  $[0,1]$ ) впевненості  $\mu^T(x)$  в тому, що чітке значення  $x$  відповідає нечіткому терму  $T$ .

Можна виділити прямі та непрямі методи завдання функцій належності. При прямому методі експерт або просто задає для  $x \in E$  значення  $\mu_A(x)$ , де  $E$  – універсальна множина, а  $A$  – нечітка множина, або визначає функцію сумісності. Як правило, прямі методи завдання функції належності використовуються для понять, які можна виміряти, таких як швидкість, час, відстань, тиск, температура і тому подібне, або коли виділяються полярні значення.

В багатьох задачах при характеризуванні об'єкту можна виділити набір ознак і для кожного з них визначити полярні значення, які відповідають значенням функції належності 0 або 1.

Непрямі методи визначення значень функції належності використовуються в тих випадках, коли відсутні елементарні властивості, що вимірюються, і через які визначається нечітка множина, що нас цікавить. Як правило, це метод парних порівнянь запропонований Сааті. Якщо б значення функцій належності були нам відомі, наприклад,  $\mu_A(x_i) = w_i, i = \overline{1, n}$ , то парні порівняння можна представити матрицею  $A = \{a_{ij}\}$ , де  $a_{ij} = w_i / w_j$  (операція ділення).

На практиці експерт сам формує матрицю  $A$ , при цьому мають на увазі, що діагональні елементи дорівнюють 1, а для елементів, які симетричні від діагонали  $a_{ij} = 1/a_{ji}$ , тобто якщо один елемент оцінюється в  $\alpha$  разів сильніше за інший, то цей останній повинен бути  $1/\alpha$  разів сильніше за першого. В загальному випадку задача зводиться до пошуку вектора  $w$ , який задовольняє рівняння вигляду  $A_w = \lambda_{\max} w$ , де  $\lambda_{\max}$  – найбільше власне значення матриці  $A$ . Оскільки матриця  $A$  додатна в побудові, рішення даної задачі існує і є додатним. Складність використання цього методу зумовлена необхідністю знаходження власного вектору матриці парних порівнянь  $A = \{a_{ij}\}$ . Причому складність зростає із зростанням розмірності універсальної множини, на якій задається лінгвістичний термін.

Для подолання цієї складності в праці запропоновано метод, який також використовує матрицю парних порівнянь, але не потребує знаходження власного вектора.

Питання про побудову функцій (F-функцій) є одним з найбільш важливих питань в теорії нечітких множин, але ця проблема не є унікальною тільки для теорії нечітких множин.

В роботі запропонована модель фундаментального вимірювання нечіткості. Нечіткість вимірювання інтенсивності деякої властивості  $P$ , якою володіє деякий об'єкт, може полягати в складності, неточному вимірюванні, або через те, що індивідууми можуть по різному сприймати інтенсивність



цієї властивості  $P$ . Для представлення функції належності нечіткої множини використовується техніка теорії вимірювання і шкалування. Аналогічні проблеми виникають при побудові характеристичних функцій теорії корисності.

В праці проведено аналіз прямого та зворотного оцінювання, групових та індивідуальних функцій належності, а в працях зображено, як можна побудувати чисельне представлення належності на шкалі інтервалів і шкалі відношень.

Питання наближення функцій належності і нечітких відношень, а також вибору серед функцій належності, узгоджених з заданим співвідношенням, такої, яка найкращим чином наближує потрібну, розглянуті в роботі [110].

Проблеми побудови  $F$ -функцій пов'язані з методами обробки експертних даних і результатів опитування спеціалістів.

Представлені методи дозволяють включити в склад обмежень на систему деякі суб'єктивні обмеження, які можуть бути спричинені тим, що людина має ряд неформальних відомостей про систему на природній мові.

Використання функцій належності надає змогу більш повно відзеркалити думку одного чи декількох експертів. Це пов'язано з неспроможністю людини формулювати власне кількісне відчуття у вигляді одинарного числа.

Нехай є  $m$  експертів, частина з яких на питання про належність елемента  $x \in X$  нечіткій множині  $A$  дає позитивну відповідь. Позначимо їх кількість через  $n_1$ . Інша частина експертів  $n_2 = m - n_1$  відповідає на це питання негативно. Тоді приймають, що  $\mu_A(x) = n_1 / (n_1 + n_2)$ .

Якщо є звичайна функція ефективності (наприклад, функція витрат енергії, коштів і тому подібне), яка встановлює лінійну впорядкованість на множині альтернатив, то вона може бути трансформована в функцію належності за допомогою нормалізації, яка зберігає лінійну впорядкованість.

Імовірнісний підхід побудови функції належності базується на подібності понять нечіткості та ймовірності. Обидва ці поняття використовуються при наявності невизначеності в системах, яка є наслідком випадкових факторів, неточності нашого знання або принципіальної неможливості та непотрібності отримання точних значень. Але навіть в тих випадках, коли нечіткість в процесі прийняття рішення може бути представлена імовірнісною моделлю, звичайно зручніше оперувати з нею методами теорії нечітких множин без залучення апарату теорії ймовірностей.

Багато авторів підкреслюють суб'єктивність функції належності на відміну від об'єктивності ймовірності як характеристики. Але дуже природно визначити функцію належності таким чином, щоб мало ймовірна з позиції теорії ймовірностей подія мала й малу ступінь належності. «Нестатична» природа побудови деяких функцій належності для якісних параметрів не означає, що в якості таких функцій не можна використовувати функції розподілу ймовірностей, які нормовані визначеним чином, тобто в

залежності від типу нечітких змінних можна виділити два способи їх побудови – статистичний та нестатистичний.

Нечіткість та випадковість можуть використовуватись в процесі прийняття рішень способами та в вигляді різних комбінацій.

Якщо для параметру, що контролюється, відома ймовірність переходу системи з допустимого стану до недопустимому в залежності від призначеного для системи режиму роботи, то функція належності, яка характеризує ступень ризику рішення, що приймається, може бути визначена з умови переваги тих режимів, де ступень ризику менша. Умова переваги може записана в наступній формі:

$$x_1 > x_2 \Leftrightarrow [1 - P(x < x_1)] \geq [1 - P(x < x_2)].$$

Тоді в якості функції належності можна прийняти функцію

$$\mu_D(x_a) = [1 - P(x < x_a)].$$

Ця функція повністю задовольняє основним вимогам до функції належності. Практично  $P(x < x_a)$  визначається на основі побудови діаграми накопичених частот переходу системи в недопустимі стани для різних режимів роботи. Існує також підхід представлення стохастичних величин як нечітких з використанням можливостей теорії Заде.

Альтернативою методу представлення функції належності шляхом запису для неї деякого рівняння, є визначення функції належності як сукупності  $r$ -рівневих множин, або шляхом розбиття області значень функції на ряд сегментів, що не перетинаються. Підхід сегментного розбиття дозволяє охарактеризувати нечітку множину за допомогою ряду функцій належності кожного з таких сегментів.

В практиці застосовується методи визначення характеристик функцій (або побудови їх оцінок) по вибіркам та на основі апріорної інформації, в яку входять обмеження на ці функції. Якщо апріорної інформації про властивості характеристичних функцій недостатньо для побудови визначених функцій, які були б «оптимальні» в якомусь розумінні, потрібно використовувати евристичні методи знаходження цих функцій з наступною експериментальною перевіркою «якості» обраних.

В багатьох практичних ситуаціях функція належності повинна бути оцінена, виходячи з часткової інформації про неї, скажімо такої, як значення, які вона приймає на кінцевій множині опорних точок  $x_1, x_2, \dots, x_n$ . В цьому випадку кажуть, що вона є частково визначеною за допомогою «прикладу, що пояснює».

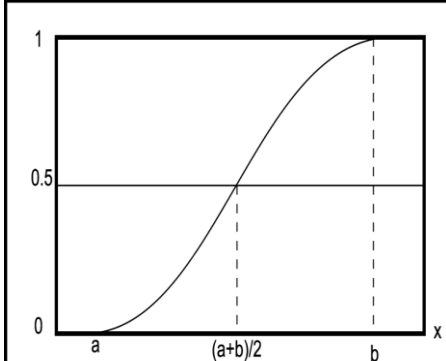
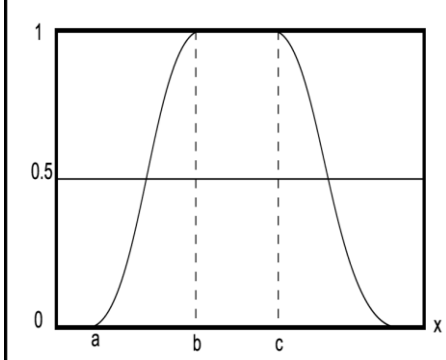
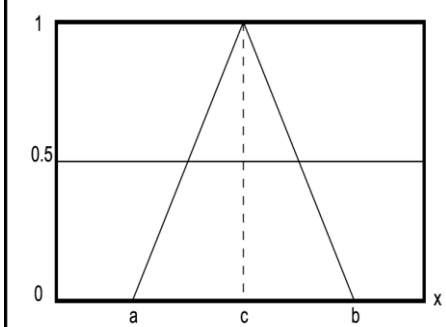
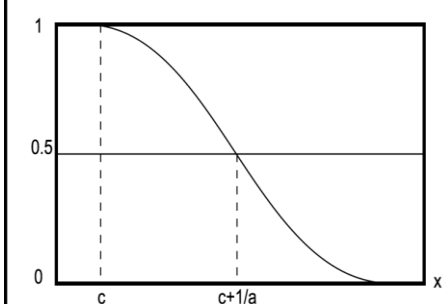
Є ряд робіт, де авторами пропонуються методи побудови та інтерпретації  $F$ -функцій на основі об'єктивної технологічної інформації.

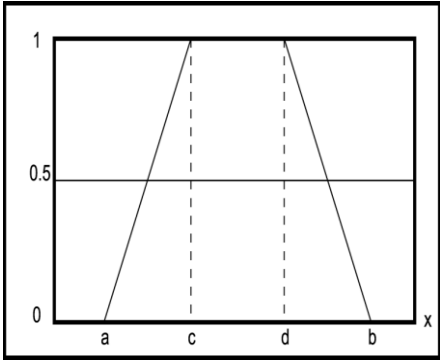
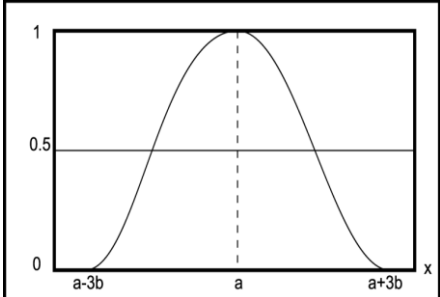
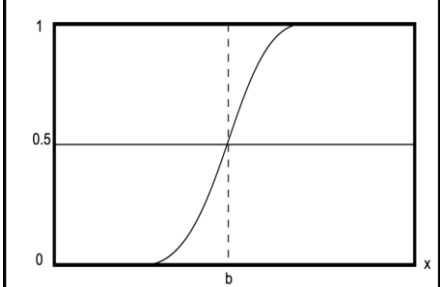
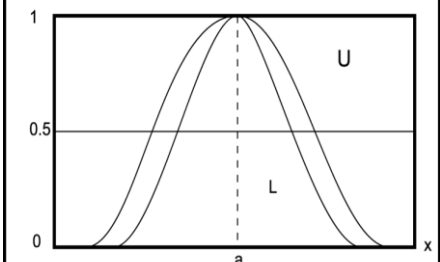
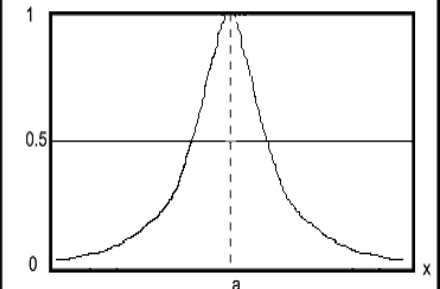
Однак, частіш за все при оцінці параметрів відомі лише інтервальні обмеження і найбільше допустиме значення. Наприклад, експерт може не

вказати мінімально та максимально допустимі режими  $[a, b]$  і номінальну продуктивність системи  $c$ . Якщо в нього більше нема інформації, то єдиною прийнятною апроксимацією є лінійна. В цьому разі параметр, що розглядається, можна охарактеризувати трикутною функцією належності  $\mu_3(x)$  (таблиця 4.3).

Таблиця 4.3

**Види функцій належності**

Математичне подання функцій належності	Графічне подання функцій належності
$\mu_1(x,a,b) = \begin{cases} 0, & \text{якщо } x \leq a \\ \frac{2(x-a)^2}{(b-a)^2}, & \text{якщо } a \leq x \leq \frac{a+b}{2} \\ 1 - \frac{2(x-a)^2}{(b-a)^2}, & \text{якщо } \frac{a+b}{2} < x < b \\ 1, & \text{якщо } x \geq b \end{cases}$	
$\mu_2(x,a,b,c) = \begin{cases} \mu_1(x,a,b), & \text{якщо } x < b \\ 1, & \text{якщо } b \leq x \leq c \\ 1 - \mu_1(x,c,c+b-a), & \text{якщо } x > c \end{cases}$	
$\mu_3(x,a,b,c) = \begin{cases} 0, & \text{якщо } x \leq a \\ \frac{x-a}{c-a}, & \text{якщо } a < x \leq c \\ \frac{b-x}{b-c}, & \text{якщо } c < x < b \\ 0, & \text{якщо } x \geq b \end{cases}$	
$\mu_4(x,a,b,c) = \begin{cases} 1, & \text{якщо } x \leq c \\ \left\{ 1 + [a(x-c)]^b \right\}^{-1}, & \text{якщо } x > c \end{cases}$	

Математичне подання функцій належності	Графічне подання функцій належності
$\mu_5(x, a, b, c, d) = \begin{cases} 0, & \text{якщо } x \leq a \\ \frac{x-a}{c-a}, & \text{якщо } a < x < c \\ 1, & \text{якщо } c \leq x \leq d \\ \frac{b-x}{b-d}, & \text{якщо } d < x < b \\ 0, & \text{якщо } x \geq b \end{cases}$	
$\mu_6(x, a, b) = \exp\left[-\frac{(x-a)^2}{2b^2}\right]$	
$\mu_7(x, a, b) = \{1 + \exp[-a(x-b)]\}^{-1}$	
$\mu_8\left(x, a, \left[\underline{b}, \bar{b}\right]\right) = \left[\mu_6^L(x, a, \underline{b}), \mu_6^U(x, a, \bar{b})\right]$	
$\mu_9(x, a, b) = \left(1 + \left(\frac{x-a}{b}\right)^2\right)^{-1}$	

В табл. 4.2 наведено всі основні види функцій належності, які застосовуються в теорії нечітких множин.

При побудові функцій належності результуючої та вхідних змінних пропонується їх конкретний вид визначати виходячи з часткової інформації про них, такої як значення, які вони приймають на кінцевій множині опорних точок  $x_1, x_2, \dots, x_n$ . Тоді можна подати функцію належності у вигляді ломаної. Або, якщо треба то можна побудувати або підібрати відому згладжену функцію, виходячи з різних допоміжних припущень о властивостях цієї функції (симетричність, монотонність, неперервність першої похідної і тому подібне) з урахуванням специфіки існуючих невизначеностей і реальної ситуації на об'єкті. Як показано в роботі, найчастіше використовуються функції належності трикутного вигляду, які, до того ж, є універсальним апроксиматором. Тому в подальшому викладенні будуть використовуватися саме такі функції належності.

## РОЗДІЛ 5. МЕТОДИ ОБРОБКИ ІНФОРМАЦІЇ В СИСТЕМАХ УПРАВЛІННЯ ЗНАННЯМИ

### 5.1. Зведення бази нечітких знань до деревоподібної структури

Задля зменшення кількості перебору та усунення недоліків, що пов'язані з невпорядкованістю БНЗ, сформулюємо задачу зведення невпорядкованої БНЗ до деревоподібної структури.

Постановка задачі:

Дано:

1)  $BZ = \bigcup_{l,z,j} (P_l^z)_j$  – невпорядкована БНЗ, де

$(P_l^z)_j$  –  $j$ -е правило ( $j = \overline{1, M}$ ,  $M = M_{l_z}$ ) для визначення  $z$ -го терму ЛЗ з ідентифікатором  $l$

2)  $I = \{I_i \mid i = \overline{1, N}\}$  – множина з уведеною на ній ієрархією, що відповідає ієрархії оцінювання ситуації

3) Відповідність елементів ієрархії оцінювання ситуації лінгвістичним змінним  $X_i$ ,  $i = \overline{1, N}$

Отримати:

$$BZ = \bigcup_{n,g} BZ_{(n+1),g} = \bigcup_{n,g} \hat{R}_{(n+1),g} \quad (5.1)$$

де  $BZ$  – деревоподібна БНЗ,

$$\begin{aligned} & \text{ЯКЩО } (X_{n1} = X_{n1}^{k_1^{z1}}) \text{ТА } (X_{n2} = X_{n2}^{k_2^{z1}}) \text{ТА} \dots \text{ТА } (X_{nS} = X_{nS}^{k_S^{z1}}) \text{АБО} \\ & (X_{n1} = X_{n1}^{k_1^{z2}}) \text{ТА } (X_{n2} = X_{n2}^{k_2^{z2}}) \text{ТА} \dots \text{ТА } (X_{nS} = X_{nS}^{k_S^{z2}}) \text{АБО} \dots \end{aligned}$$

$$\hat{R}_{(n+1),g} = \text{АБО } \left( X_{n1} = X_{n1}^{k_1^{zM^z}} \right) \text{ТА } \left( X_{n2} = X_{n2}^{k_2^{zM^z}} \right) \text{ТА} \dots \text{ТА } \left( X_{nS} = X_{nS}^{k_S^{zM^z}} \right) \quad (5.2)$$

$$\begin{aligned} \text{ТО } X_{(n+1),g} &= X_{(n+1),g}^z \\ z &= \overline{1, Z_{(n+1),g}} \end{aligned}$$

де  $\hat{R}_{(n+1),g}$  – часткова БНЗ для визначення залежності ЛЗ  $X_{(n+1),g}$  від ЛЗ  $\{X_{ns} = X_{nq_s} \mid s = \overline{1, S}\}$ ,  $n$  – номер рівня,  $q_s$  – номер ЛЗ на рівні  $n$ ,  $s$  – номер ЛЗ у рядку-кон'юнкції,  $S = S_{(n+1),g}^z$  – кількість ЛЗ у рядку-кон'юнкції,  $M^z = M_{(n+1),g}^z$  – кількість рядків кон'юнкцій для визначення  $z$ -го терму ЛЗ  $X_{(n+1),g}$ ,  $X_{ns}^{k_s^{zr}}$  – терм ЛЗ  $X_{ns}$ ,  $k_s^{zr} = k_{ns}^{zr}$  – номер терму.

Для рішення цієї задачі нижче запропоновано підхід до зведення БНЗ до деревоподібної структури за рахунок використання природної ієрархічності, а саме ієрархії оцінювання ситуації, та притаманних їм евристик (рисунок 5.1). Він полягає у співставленні з кожною ЛЗ, окрім тих,

що відповідають вхідним змінним, частини БЗ, яка визначатиме залежність цієї ЛЗ від ЛЗ попереднього рівня.

Таку частину БЗ будемо називати частковою БЗ. Тоді для кожного набору правил, що відповідають одному терму ЛЗ буде проводитись композиція, яка буде полягати в знаходженні максимуму серед всіх отриманих значень ФН. Нижче підхід до зведення БЗ до деревоподібної структури наводиться більш детально.

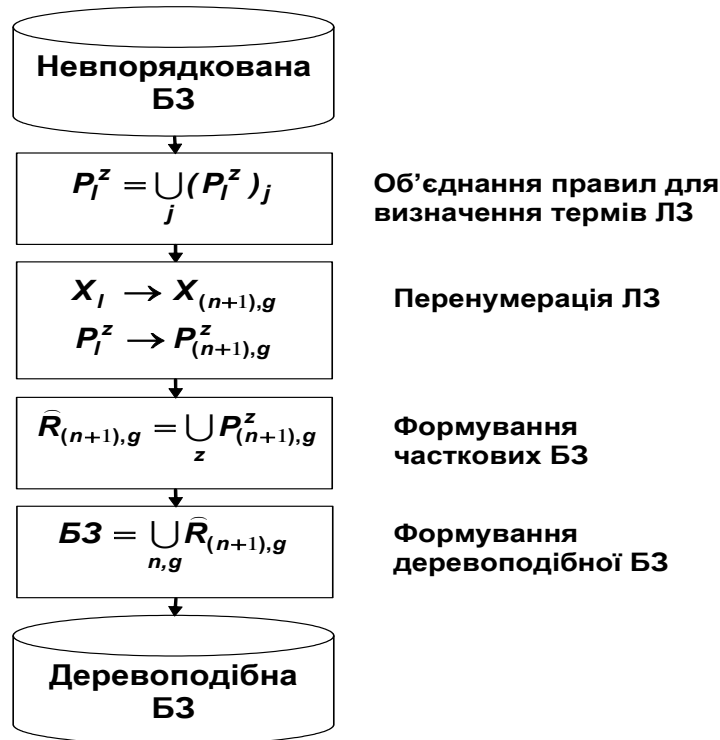


Рис. 5.1. Кроки вирішення задачі зведення нечіткої БЗ до деревоподібної структури

Частину неупорядкованої БЗ, що буде визначати терм ЛЗ поточного рівня можна подати співвідношенням (5.3), яке графічно представляється у вигляді дворівневого графу «ТА/АБО» (рисунок 5.2).

$$\begin{aligned}
 & \text{ЯКЩО} \left( X_{i_{11}} = X_{i_{11}}^{k_{i_{11}}} \right) \text{ТА} \left( X_{i_{12}} = X_{i_{12}}^{k_{i_{12}}} \right) \text{ТА} \dots \text{ТА} \left( X_{i_{1m_1}} = X_{i_{1m_1}}^{k_{i_{1m_1}}} \right) \text{АБО} \\
 & \left( X_{i_{21}} = X_{i_{21}}^{k_{i_{21}}} \right) \text{ТА} \left( X_{i_{22}} = X_{i_{22}}^{k_{i_{22}}} \right) \text{ТА} \dots \text{ТА} \left( X_{i_{2n_2}} = X_{i_{2n_2}}^{k_{i_{2n_2}}} \right) \text{АБО} \dots \\
 & \text{АБО} \left( X_{i_{M1}} = X_{i_{M1}}^{k_{i_{M1}}} \right) \text{ТА} \left( X_{i_{M2}} = X_{i_{M2}}^{k_{i_{M2}}} \right) \text{ТА} \dots \text{ТА} \left( X_{i_{Mm_M}} = X_{i_{Mm_M}}^{k_{i_{Mm_M}}} \right)
 \end{aligned} \tag{5.3}$$

$$\text{ТО } X_l = X_l^z$$

де  $X_{i_{js}}^{k_{i_{js}}}$  – означає лінгвістичну змінну,

$i_{js} = i_{lzs}$  – ідентифікатор лінгвістичної змінної, та визначає ту змінну,

що знаходиться на  $s$ -му місці ( $s = \overline{1, n_j}$ ) в лівій частині  $j$ -го ( $j = \overline{1, M}, M = M_{l_z}$ ) правила „ЯКЩО.., ТО..” для визначення  $z$ -го терму лінгвістичної змінної з ідентифікатором  $l$ ,

$n_j = n_{l_z j}$  – кількість змінних в лівій частині  $j$ -го рядка кон’юнкції,

$X_{i_{js}}^{k_{ijs}}$  – терм лінгвістичної змінної  $X_{i_{js}}$  з номером  $k_{ijs}$ .

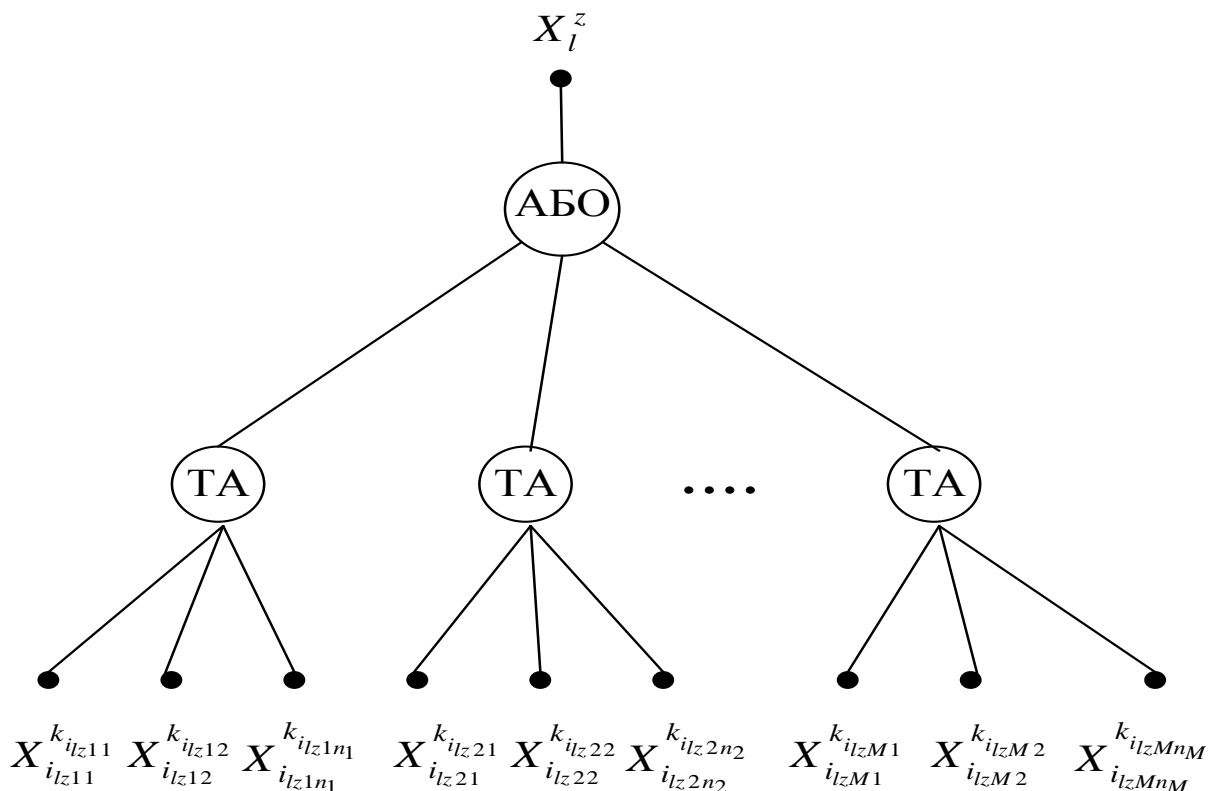


Рис. 5.2. Граф «ТА/АБО», що відповідає вершині терму ЛЗ більш ніж першого рівня

Тоді ФН терму  $X_l^z$  буде обчислюватись за формулою (5.3).

$$\mu_{X_l^z} = \bigvee_{j=1}^M (\mu_{X_l^z})_j = \bigvee_{j=1}^M \left( \mu_{X_{i_{zj1}}^{k_{i_{zj1}}}} \wedge \mu_{X_{i_{zj2}}^{k_{i_{zj2}}}} \wedge \dots \wedge \mu_{X_{i_{zjn_j}}^{k_{i_{zjn_j}}}} \right) \quad (5.4)$$

де  $(\mu_{X_l^z})_j$  – значення функції належності терму  $X_l^z$ , яке отримано при обчисленні  $j$ -го правила.

Виходячи із структури інформації та її ієрархічності перенумеруємо ЛЗ наступним чином:  $X_{ni}$ , де  $n = \overline{1, N}$  – номер рівня, на якому знаходиться ЛЗ, а  $i = \overline{1, M_n}$  – номер ЛЗ на цьому рівні. Тоді часткову БНЗ, для визначення залежності ЛЗ  $(n+1)$ -го рівня від ЛЗ  $n$ -го рівня можна записати у вигляді формули (5.2)

В праці [2] пропонується записувати БНЗ у вигляді матриць знань для спрощення сприйняття. Тоді матрицю знань, яка відповідає (5.2), можна



записати у вигляді формули (5.5) А значення функції належності термів лінгвістичної змінної  $X_{(n+1),l}$  буде обчислюватись за формулою (5.6).

При такому впорядкуванні БЗ для обчислення функції належності терму лінгвістичної змінної необхідно буде задіяти лише  $M = M_{lz}$  правил.

Тоді БНЗ можна подати у вигляді деревоподібної структури (рисунок 5.3) в якій можна виділити дві частини: нижню і верхню. Нижня частина деревоподібної БНЗ визначає залежність ЛЗ другого рівня від ЛЗ першого рівня, які пов'язані з вхідними змінними. Верхня частина деревоподібної БНЗ визначає залежність результуючої ЛЗ від ЛЗ другого рівня через ЛЗ проміжних рівнів.

$$\text{Matr}_{X_{(n+1),g}} = \begin{array}{|c|c|c|c|c|}
 \hline
 X_{n1} & X_{n2} & \dots & X_{nS} & X_{(n+1),g} \\
 \hline
 X_{n1}^{k_1^{11}} & X_{n2}^{k_2^{11}} & \dots & X_{nS}^{k_S^{11}} & \vdots \\
 \dots & \dots & \dots & \dots & X_{(n+1),g}^1 \\
 X_{n1}^{k_1^{1M^1}} & X_{n2}^{k_2^{1M^1}} & \dots & X_{nS}^{k_S^{1M^1}} & \vdots \\
 \dots & \dots & \dots & \dots & \dots \\
 X_{n1}^{k_1^{z1}} & X_{n2}^{k_2^{z1}} & \dots & X_{nS}^{k_S^{z1}} & \vdots \\
 \dots & \dots & \dots & \dots & X_{(n+1),g}^z \\
 X_{n1}^{k_1^{zM^z}} & X_{n2}^{k_2^{zM^z}} & \dots & X_{nS}^{k_S^{zM^z}} & \vdots \\
 \dots & \dots & \dots & \dots & \dots \\
 X_{n1}^{k_1^{Z1}} & X_{n2}^{k_2^{Z1}} & \dots & X_{nS}^{k_S^{Z1}} & \vdots \\
 \dots & \dots & \dots & \dots & X_{(n+1),g}^Z \\
 X_{n1}^{k_1^{ZM^Z}} & X_{n2}^{k_2^{ZM^Z}} & \dots & X_{nS}^{k_S^{ZM^Z}} & \vdots \\
 \hline
 \end{array} \tag{5.5}$$

$$\mu_{X_{(n+1),g}^z} = \bigvee_{j=1}^{M^z} \left( \mu_{X_{n1}^{k_1^{zj}}} \wedge \mu_{X_{n2}^{k_2^{zj}}} \wedge \dots \wedge \mu_{X_{nS}^{k_S^{zj}}} \right) \tag{5.6}$$

$$z = \overline{1, Z}, Z = Z_{(n+1),g}$$

Зауважимо, що виходячи з особливостей ПО застосовувати автоматичні механізми формування та настроювання можливо лише для нижньої частини деревоподібної БНЗ. В роботі запропоновано розглядати нижню частину БЗ у вигляді нечіткої нейромережі аналогічної NEFCLASS, що дозволило використати розроблені для нечітких нейромереж механізми навчання.

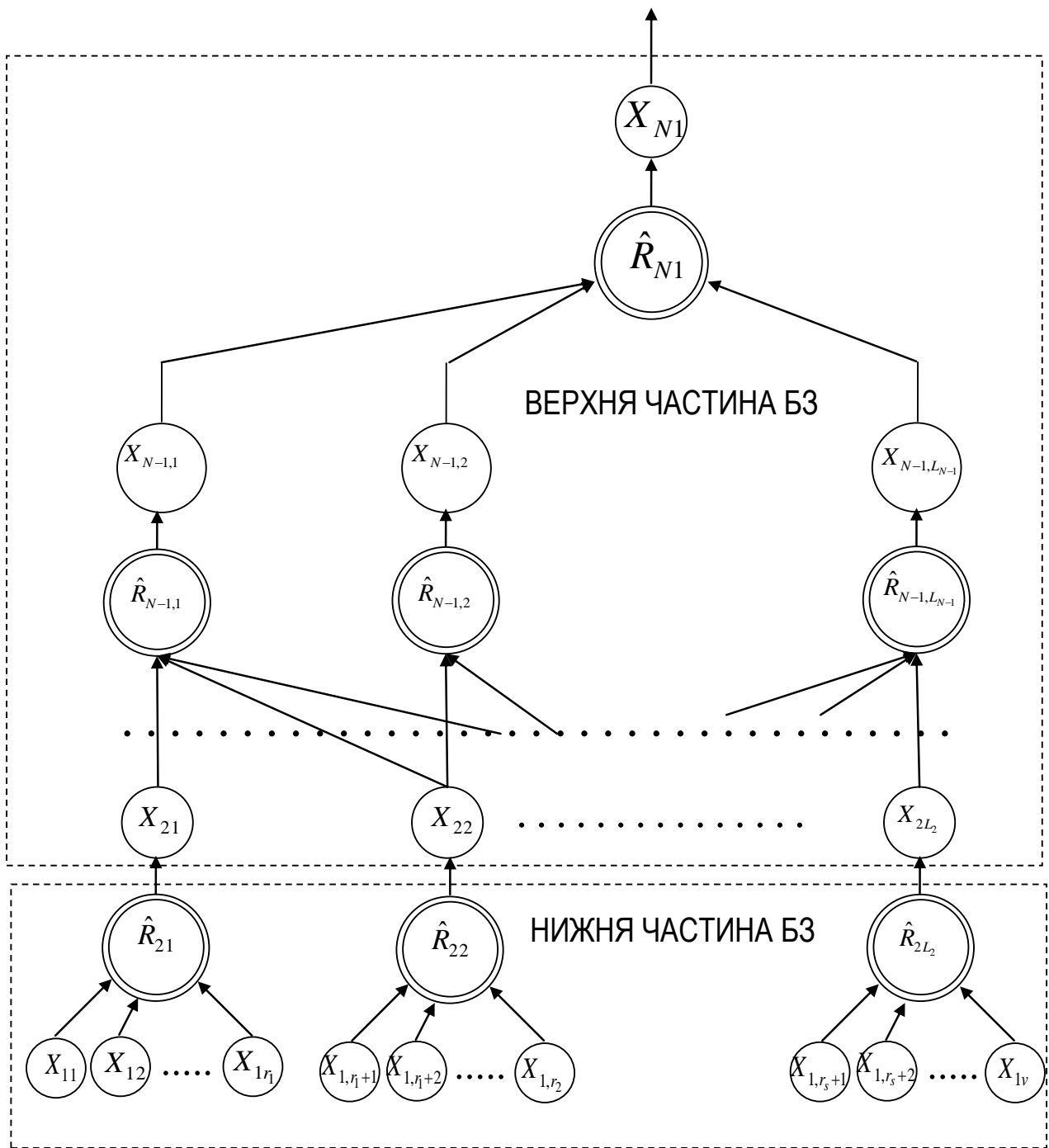


Рис. 5.3. Деревоподібна БНЗ

Знайдемо обчислювальну складність виведення на основі неупорядкованої БНЗ та на основі деревоподібної БНЗ. Для цього сформулюємо такі твердження.

Твердження 2.1 Алгоритм нечіткого виведення на основі неупорядкованої БНЗ має обчислювальну складність

$$O \left( (I-1) \cdot \sum_{i=1}^N \left( I^{\frac{i-1}{1-I}} \cdot K^{I^{i-1}} \right) + \left( I^{\frac{I^{N-1}-1}{1-I}} K^{I^{N-1}-1} - 1 \right) \right),$$

де  $I$  - загальна кількість лінгвістичних змінних на рівні,  $K$  - розмір терм-множин лінгвістичних змінних,  $N$  - кількість рівнів.

### Доведення

Оскільки має місце ієрархія оцінювання ситуації, то і для неупорядкованої БНЗ вірно припущення, що всі терми кожної ЛЗ знаходяться тільки на одному рівні та всі терми ЛЗ, для якої вирішується задача класифікації є кінцевими.

Для неупорядкованої БЗ впорядкуємо ЛЗ на рівні. Тоді лінгвістичну змінну можна буде однозначно визначити за допомогою двох індексів номер рівня  $n$  та номер змінної на рівні  $i$ , тобто  $X_{ni}$ . Нехай лінгвістична змінна  $X_{ni}$  має  $K_{ni}$  термів. Терм лінгвістичної змінної  $X_{ni}$  будемо позначати  $X_{ni}^{k_{ni}^{lj}}$ , де  $k_{ni}^{lj}$  – номер терму лінгвістичної змінної, яка стоїть в  $j$ -му правилі для визначення  $z$ -го терму  $l$ -тої лінгвістичної змінної  $(n+1)$ -го рівня. Тут  $l$  означає не ідентифікатор, а номер на рівні. Позначимо кількість різних правил, які визначають терм  $X_{ni}^{k_{ni}^{lj}}$ , тобто в яких він стоїть в правій частині правила, через  $M_{ni}^{k_{ni}^{lj}}$ .

Тоді вираз 5.1 можна записати таким чином:

$$\begin{aligned} & \text{ЯКЩО} \left( X_{nq_1} = X_{nq_1}^{k_{nq_1}^{lj}} \right) \text{ТА} \left( X_{nq_2} = X_{nq_2}^{k_{nq_2}^{lj}} \right) \text{ТА} \dots \\ & \text{ТА} \left( X_{nq_s} = X_{nq_s}^{k_{nq_s}^{lj}} \right) \text{ТО} X_{(n+1),l} = X_{(n+1),l}^z \\ & \quad \overline{j = 1, M_{(n+1),l}^z} \end{aligned} \quad (5.7)$$

де  $q_s$  – номер лінгвістичної змінної на рівні  $n$ , а  $s = \overline{1, S}$ ,  $S = S_{(n+1),lj}^z$  – номер лінгвістичної змінної у  $j$ -му правилі для визначення  $z$ -го терму  $l$ -тої лінгвістичної змінної  $(n+1)$ -го рівня.

Або у вигляді функцій належності:

$$\begin{aligned} \mu_{X_{(n+1),l}^z} &= \mu_{X_{nq_1}^{k_{nq_1}^{lj}}} \wedge \mu_{X_{nq_2}^{k_{nq_2}^{lj}}} \wedge \dots \wedge \mu_{X_{nq_s}^{k_{nq_s}^{lj}}} \\ & \quad \overline{j = 1, M_{(n+1),l}^z}, S = S_{(n+1),lj}^z \end{aligned} \quad (5.8)$$

Кількість значень функції належності терму  $X_{ni}^{k_{ni}^{lj}}$ , тобто кількість виконаних правил на  $(n-1)$ -му рівні для отримання функції належності терму  $X_{ni}^{k_{ni}^{lj}}$   $n$ -го рівня позначимо  $A_{ni}^{k_{ni}^{lj}}$  ( $n \geq 2$ ).

Кількість значень функції належності терму  $X_{n+1,l}^{k_{n+1,l}^z}$ , які будуть отримані для нього на  $(n+1)$ -му рівні визначається наступним чином:

$$A_{n+1,l}^z = \sum_{j=1}^{M_{(n+1),l}^z} \prod_{i=1}^{S_{(n+1),lj}^z} A_{ni}^{k_{ni}^{lj}} \quad (5.9)$$

Необхідно знайти кількість використаних правил та кількість проведених операцій для обчислення значень функцій належності термів лінгвістичної змінної  $Y = X_{Nw}$ . Величини  $A_{Nw}^b$ , навіть при фіксованому першому та останньому рівнях, залежать від кількості лінгвістичних змінних і термів на проміжному рівні ( $1 < n < N$ ).

Знайдемо максимальну кількість лінгвістичних змінних, що знаходяться в лівій частині правила для визначення терму лінгвістичної змінної, що знаходиться в правій частині правила, та позначимо її  $S$ . Визначимо максимальну кількість термів, що відповідають одній лінгвістичній змінній, тобто розмір її терм-множини, та позначимо його  $K$ .

Припустимо, що

$$M_{ni}^c = M, S_{nij}^c = S, K_{ni} = K$$

$$j = \overline{1, M_{ni}^c}, i = \overline{1, I_n}, c = \overline{1, K_{ni}}, n = \overline{2, N},$$

де  $I_n$  – кількість лінгвістичних змінних на рівні  $n$ .

Цього можна досягти додавши до відповідних правил віртуальні (неіснуючі) терми лінгвістичних змінних.

Враховуючи введені припущення можна сказати, що для кожного терму рівня  $n$  буде однакова кількість отриманих значень функцій належності, яку будемо позначати  $A_n$ . Доведемо за індукцією, що

$$A_n = M \frac{1-S^{n-1}}{1-S} \quad (5.10)$$

На першому рівні для кожного терму існує лише одне значення функції належності, тобто  $A_1 = 1$ . Підставивши  $n = 1$ , отримуємо

$$A_1 = M \frac{1-S^{1-1}}{1-S} = M^0 = 1$$

Отже формула (5.10) вірна для  $n = 1$ .

Робимо індуктивне припущення, що формула (5.10) вірна для  $n = N - 1$ . Доведемо, що вона вірна і для  $n = N$ .

Залежність  $A_N$  від  $A_{N-1}$  визначається за допомогою формули (5.9). Врахувавши всі попередні припущення та (5.10) отримуємо

$$A_N = \sum_{j=1}^M \prod_{i=1}^S A_{N-1} = \sum_{j=1}^M \prod_{i=1}^S M \frac{1-S^{N-2}}{1-S} = M \cdot \left( M \frac{1-S^{N-2}}{1-S} \right)^S = M \frac{1-S^{N-1}}{1-S}$$

Отже вірність формули (5.10) доведено.

Кількість правил, що використанні для отримання всіх значень функції належності на рівні визначається наступним чином

$$\hat{A}_n = \sum_{i=1}^n I \cdot K \cdot A_i = I \cdot K \cdot \left( \sum_{i=1}^n M \frac{1-S^{i-1}}{1-S} \right) \quad (5.11)$$

де  $I$  – кількість лінгвістичних змінних на рівні  $i$ ,  
 $K$  – кількість термів для кожної змінної.

Для визначення обчислювальної складності алгоритму необхідно визначити кількість операцій, які виконуються під час роботи алгоритму. Оскільки під час використання кожного правила знаходиться мінімум функцій належності, то обчислювальна складність для кожного правила дорівнює кількості членів мінус 1, тобто  $(S - 1)$ . Врахувавши (5.11) визначимо кількість операцій, які необхідно зробити для обчислення всіх значень функцій належності кожного терму на рівні  $n$ .

$$B_n = I \cdot K \cdot (S - 1) \cdot \left( \sum_{i=1}^n M^{\frac{1-S^{i-1}}{1-S}} \right) \quad (5.12)$$

Для того, щоб визначити остаточне рішення при проведенні задачі класифікації потрібно знайти максимальне значення функції належності для кожного терму лінгвістичної змінної, для якої проводиться класифікація та провести дефазифікацію, тобто у випадку, що розглядається, визначити той терм, який має найбільше значення функції належності.

Враховуючи припущення, що всі терми модельованої лінгвістичної змінної знаходяться на рівні  $N$ , то кількість операцій, які необхідно зробити для визначення кінцевого терму визначається наступним чином

$$\hat{B}_N = I \cdot K \cdot (S - 1) \cdot \left( \sum_{i=1}^N M^{\frac{1-S^{i-1}}{1-S}} \right) + \left( K \cdot M^{\frac{1-S^{N-1}}{1-S}} - 1 \right) \quad (5.13)$$

Верхню досягну межу можна досягти припустивши, що всі лінгвістичні змінні рівня використовуються при визначенні кожного терму наступного рівня, тобто  $S = I$ . Тоді загальна кількість правил для переходу з одного рівня на інший дорівнює  $K^I$ , а кількість різних правил, за допомогою яких можна вивести терм наступного рівня буде дорівнювати кількості можливих варіантів створення правила поділеної на загальну кількість термів на рівні  $M = \frac{K^I}{I \cdot K}$ , тобто  $M = I^{-1} \cdot K^{I-1}$ . В цьому випадку (5.13) переписеться наступним чином:

$$\hat{B}_N = (I - 1) \cdot \sum_{i=1}^N \left( I^{\frac{i-1}{1-I}} \cdot K^{I^{i-1}} \right) + \left( I^{\frac{N-1}{1-I}} K^{I^{N-1}-1} - 1 \right) \quad (5.14)$$

Твердження доведено.

Твердження 5.2 Алгоритм нечіткого виведення на основі деревоподібної БНЗ має обчислювальну складність  $O(I \cdot K(3 - N + (N - 2) \cdot K^{I-1}) + I^{-1} \cdot K^I - 1)$ , де  $I$  - загальна кількість лінгвістичних змінних на рівні,  $K$  - розмір терм-множин лінгвістичних змінних,  $N$  - кількість рівнів.

### Доведення

Для знаходження обчислювальної складності виведення на основі деревоподібної БНЗ будемо використовувати БНЗ вигляду (5.1). Також зробимо припущення як і при доведенні твердження 2.1. Тобто всі терми кожної лінгвістичної змінної знаходяться тільки на одному рівні, всі терми лінгвістичної змінної, для якої вирішується задача класифікації є кінцевими, максимальна кількість ЛЗ, що знаходяться в лівій частині правила для визначення терму ЛЗ, що знаходиться в правій частині правила дорівнює  $S$ , максимальна кількість термів, що відповідають одній ЛЗ дорівнює  $K$ .

Тоді для визначення термів на першому рівні виконується  $I \cdot K$  операцій, які полягають у визначенні функцій належності.

Для обчислення значення функції належності кожного терму наступного рівня необхідно використати  $M$  правил, в кожному з яких виконується  $(S - 1)$  операцій, та застосувати правило композиції, тобто у випадку, що розглядається, знайти максимум з отриманих функцій належності. Для цього необхідно виконати  $(M - 1)$  операцій. При такому виведенні кількість правил, які необхідні для виведення терма наступного рівня, не зростає зі зростанням рівня, оскільки кожен терм в лівій частині правила може мати лише одне значення функції належності отримане на попередньому рівні. Кількість операцій для визначення значень функцій належності всіх термів наступного рівня визначається формулою:

$$\tilde{A} = I \cdot K \cdot (M \cdot (S - 1) + (M - 1)) = I \cdot K \cdot (M \cdot S - 1) \quad (5.15)$$

Додамо, що на  $N$ -му рівні обчислюються лише значення функцій належності для термів шуканої змінної, а потім серед них визначається найбільший. Тоді загальна кількість операцій при виведенні на основі деревоподібної БЗ обчислюється наступним чином (5.16):

$$\hat{B}_N = I \cdot K \cdot (1 + (N - 2) \cdot (M \cdot (S - 1) + (M - 1))) + K \cdot (M - 1) + (K - 1) \quad (5.16)$$

Спростивши формулу (5.16) отримуємо:

$$\hat{B}_N = I \cdot K \cdot (3 - N + (N - 2) \cdot M \cdot S) + K \cdot M - 1 \quad (5.17)$$

Для обчислення верхньої досяжної межі припустимо, що при визначенні кожного терму наступного рівня використовуються всі лінгвістичні змінні рівня, тобто  $S = I$ . Тоді  $M = I^{-1} \cdot K^{I-1}$ . В цьому випадку (5.17) перепишеться наступним чином (5.18):

$$\hat{B}_N = I \cdot K \cdot (3 - N + (N - 2) \cdot K^{I-1}) + I^{-1} \cdot K^I - 1 \quad (5.18)$$

Твердження доведено.

## 5.2. Модифікація підсистеми нечіткого виведення

Сформулюємо постановку задачі модифікації підсистеми нечіткого виведення таким чином, щоб ПНВ дозволяла отримувати не тільки значення результуючої змінної, але й визначати найвпливовіші вхідні параметри.

Дано:

$$\text{ПНВ} : \{x_i \mid i = \overline{1, v}\} \longrightarrow y \quad (5.19)$$

Отримати:

$$\text{ПНВ} : \{x_i \mid i = \overline{1, v}\} \longrightarrow \{y, x_y\} \quad (5.20)$$

де  $\{x_i \mid i = \overline{1, v}\}$  – множина вхідних змінних,

$y$  – результуюча змінна,

$x_y$  – вхідна змінна, що визначила значення результуючої змінної.

Рішення цієї задачі полягає в модифікації схеми нечіткого виведення на основі бази нечітких знань та структури підсистеми нечіткого виведення (рисунок 5.4).

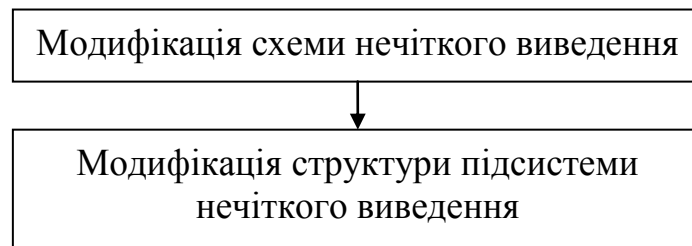


Рис. 5.4. Послідовність кроків розв'язання задачі модифікації ПНВ

## 5.3. Модифікація схеми нечіткого виведення

Модифікація схеми нечіткого виведення на основі БНЗ полягає у додатковому визначенні шляху до найвпливовішого параметру через знаходження на кожному кроці виведення того терму з лівої частини правила, що визначив відповідне значення функції належності терму лінгвістичної змінної в правій частині (рисунок 5.5).

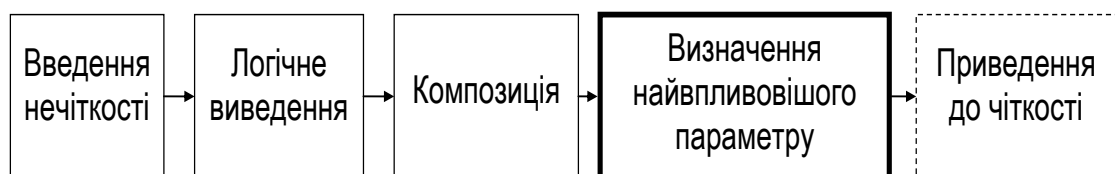


Рис. 5.5. Модифікована схема нечіткого виведення

Для реалізації запропонованого методу пропонується додатково пов'язати з кожною лінгвістичною змінною два параметри, які будуть вказувати на те, який саме терм та якої лінгвістичної змінної вплинув на кожен терм поточної лінгвістичної змінної  $(X_{nl}^z)_p^s$ ,  $p$  - номер вершини на

$(n-1)$ - му рівні, а  $s$ - номер її терма. Однак, оскільки може бути декілька термів вершин, що мають однакові значення, то  $p$  і  $s \in$  масиви  $p[a], s[a]$ .

Покажемо, що при використанні модифікованої схеми нечіткого виведення на основі деревоподібної БНЗ будуть отримані такі самі результати, як і на основі невпорядкованої. Для цього сформулюємо наступне твердження.

Твердження 5.3 При модифікованому нечіткому виведенні на основі деревоподібної БНЗ будуть отримані такі ж результуючі значення та найвпливовіші вхідні параметри, що і при виведенні на основі невпорядкованої БНЗ.

Доведення

Для доведення твердження слід довести, що при виведенні на основі деревоподібної БНЗ і невпорядкованої БНЗ будуть отримані однакові значення результуючої лінгвістичної змінної і значення функцій належності термів лінгвістичних змінних поточного рівня будуть визначатись однаковими термами лінгвістичних змінних попереднього рівня.

Доведення буде проводитись по індукції.

На першому рівні значення ФН термів ЛЗ будуть мати лише одне значення для обох БЗ.

Оскільки БНЗ буде мати не менше 2-х рівнів, то покажемо, що умова твердження виконується для рівня  $n = 2$ .

При виведенні на основі невпорядкованої БЗ, скориставшись (5.9), отримуємо для кожного терму  $A_{2l}^z = M_{2l}^z$  способів виведення, тобто  $M_{2l}^z$  значень функції належності. Максимальне значення функції належності  $z$ -го терму  $l$ -ї лінгвістичної змінної знайдемо врахувавши те, що на першому рівні кількість значень функцій належності для кожного терму дорівнює 1, тобто існує лише один набір значень функцій належності термів лівої частини правила  $j$ . Тоді  $\mu_{X_{2l}^z}$ , врахувавши впорядкованість по рівням та (5.8), обчислюється за формулою:

$$\mu_{X_{2l}^z} = \bigvee_{j=1}^{M_{2l}^z} \left( \mu_{X_{2q_1}^{k_{2q_1}^{lj}}} \wedge \mu_{X_{2q_2}^{k_{2q_2}^{lj}}} \wedge \dots \wedge \mu_{X_{2q_s}^{k_{2q_s}^{lj}}} \right) \quad (5.21)$$

$$S = S_{2li}^z$$

При визначенні значення  $z$ -го терму  $l$ -ї лінгвістичної змінної другого рівня на основі деревоподібної БНЗ скористаємось (5.4), врахувавши впорядкованість по рівням і те що в цьому випадку  $M_{lz} = M_{2l}^z$ :

$$\mu_{X_{2l}^z} = \bigvee_{j=1}^{M_{lz}} \left( \mu_{X_{2l}^z} \right)_j = \bigvee_{j=1}^{M_{2l}^z} \left( \mu_{X_{2l}^z} \right)_j = \bigvee_{j=1}^{M_{2l}^z} \left( \mu_{X_{2q_1}^{k_{2q_1}^{lj}}} \wedge \mu_{X_{2q_2}^{k_{2q_2}^{lj}}} \wedge \dots \wedge \mu_{X_{2q_s}^{k_{2q_s}^{lj}}} \right) \quad (5.22)$$

$$S = S_{2li}^z$$



де  $(\mu_{X_l^z})_j$  означає значення функції належності терму  $X_l^z$ , яке отримано при обчисленні  $j$ -го правила.

Порівнюючи формули можна зробити висновок, що у випадку  $n = 2$  результуючі значення функцій належності термів однакові, при використанні як неупорядкованої так і деревоподібної БНЗ.

Терми лінгвістичних змінних попереднього рівня, які визначили функції належності термів поточного рівня, врахувавши що під операціями кон'юнкції та диз'юнкції розуміється  $\min$  та  $\max$  відповідно, та виходячи з формул в обох випадках визначаються за формулою:

$$X_{1q_i}^{k^{lj}} = \arg \min_{s=1, S} \arg \max_{j=1, M_{2l}^z} \left( \mu_{X_{2l}^z} \left( X_{1q_1}^{k^{lj}}, X_{1q_2}^{k^{lj}}, \dots, X_{1q_S}^{k^{lj}} \right) \right) \quad (5.23)$$

$$S = S_{2lj}^z$$

Робимо індуктивне припущення, що умови твердження вірні для рівня  $(n-1)$ . Доведемо, що для рівня  $n$  вони також вірні.

У випадку неупорядкованої БНЗ максимальне значення функції належності  $z$ -го терму  $l$ -ї лінгвістичної змінної визначається наступним чином:

$$\mu_{X_{nl}^z} = \bigvee_{j=1}^{M_{nl}^z} \left( \mu_{X_{nl}^z} \right)_j = \bigvee_{j=1}^{M_{nl}^z} \bigvee_{h=1}^{\prod_{i=1}^S A} \left( \mu_{X_{(n-1),q_1}^{k^{lj}}} \wedge \mu_{X_{(n-1),q_2}^{k^{lj}}} \wedge \dots \wedge \mu_{X_{(n-1),q_S}^{k^{lj}}} \right)_h \quad (5.24)$$

$$S = S_{nlj}^z, A = A_{(n-1),i}^{k^{lj}}$$

Скориставшись дистрибутивністю операцій диз'юнкції і кон'юнкції спростимо вираз:

$$\mu_{X_{nl}^z} = \bigvee_{j=1}^{M_{nl}^z} \left( \left( \bigvee_{h=1}^{\prod_{i=1}^S A} \left( \mu_{X_{(n-1),q_1}^{k^{lj}}} \right)_h \right) \wedge \left( \bigvee_{h=1}^{\prod_{i=1}^S A} \left( \mu_{X_{(n-1),q_2}^{k^{lj}}} \right)_h \right) \wedge \dots \wedge \left( \bigvee_{h=1}^{\prod_{i=1}^S A} \left( \mu_{X_{(n-1),q_S}^{k^{lj}}} \right)_h \right) \right) \quad (5.25)$$

$$S = S_{nlj}^z, A = A_{(n-1),i}^{k^{lj}}$$

Оскільки для рівня  $(n-1)$  твердження виконується, тобто максимальне значення функції належності для кожного терму отримане на основі

непорядкованої БЗ співпадало з отриманим на основі деревоподібної, то (5.25) переписеться наступним чином:

$$\mu_{X_{nl}^z} = \bigvee_{j=1}^{M_{nl}^z} \left( \mu_{X_{(n-1),q_1}^{k^{lj}}} \wedge \mu_{X_{(n-1),q_2}^{k^{lj}}} \wedge \dots \wedge \mu_{X_{(n-1),q_S}^{k^{lj}}} \right) \quad (5.26)$$

$$S = S_{nlj}^z$$

де  $\mu_{X_{(n-1),q_r}^{k^{lj}}}$  ( $r = \overline{1, S}$ ,  $S = S_{nlj}^z$ ) – максимальне значення функції

належності  $k_{(n-1),q_r}^{lj}$  терму лінгвістичної змінної  $X_{(n-1),q_r}$ .

Як бачимо формула співпадає з тим, що було б отримано за допомогою деревоподібної БЗ за формулою, врахувавши те, що  $M_{lz} = M_{nl}^z$ .

Терми лінгвістичних змінних  $(n-1)$ -го рівня, які визначили значення функцій належності термів лінгвістичних змінних  $n$ -го рівня врахувавши будуть обчислюватись однаково для непорядкованої та деревоподібної БЗ за формулою:

$$X_{(n-1),q_i}^{k^{lj}} = \arg \min_{s=1, S} \arg \max_{j=1, M_{nl}^z} \left( \mu_{X_{nl}^z} \left( X_{(n-1),q_1}^{k^{lj}}, X_{(n-1),q_2}^{k^{lj}}, \dots, X_{(n-1),q_S}^{k^{lj}} \right) \right) \quad (5.27)$$

$$S = S_{nlj}^z$$

Твердження доведено.

#### 5.4. Модифікація структури підсистеми нечіткого виведення

Другим кроком в рішенні задачі модифікації ПНВ є модифікація структури ПНВ таким чином, що вона враховувала модифікацію схеми нечіткого виведення. Для цього структуру ПНВ слід змінити так як показано на рисунку 5.6.

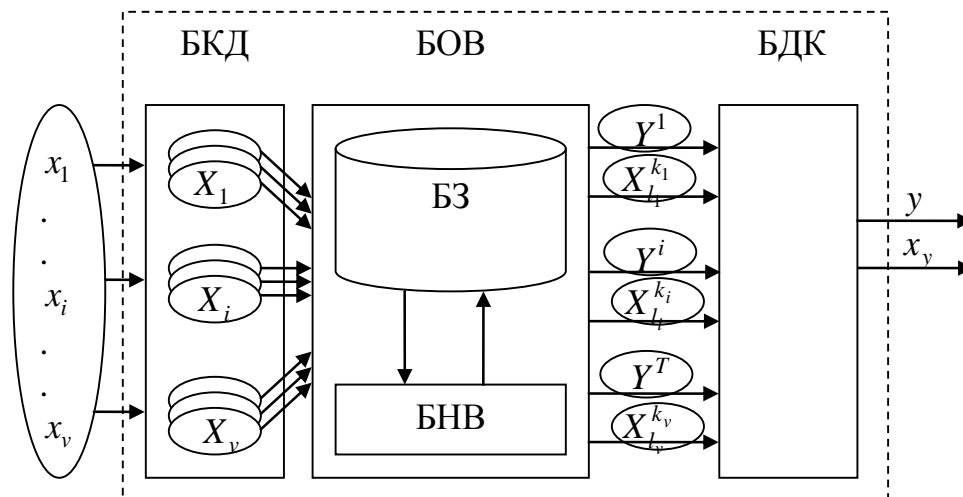


Рис. 5.6. Модифікована підсистема нечіткого виведення

Як і в розглянутій вище стандартній ПНВ на вхід БКД подається кількісна інформація  $\{x_i | i = \overline{1, v}\}$ , якій на виході ставиться у відповідність сукупність множин.

Далі ці множини  $\{X_i | i = \overline{1, v}\}$  подаються на вхід БОВ. БОВ передає ці дані БНВ, який, виходячи проводить нечітке виведення на основі БЗ. На виході ПНВ отримуємо множину пар  $\{(Y^j, X_{l_j}^{k_j}) | j = \overline{1, T}\}$  (5.7).

$$\text{БОВ} : \{X_i | i = \overline{1, v}\} \xrightarrow{\text{модифіковане нечітке логічне виведення}} \{(Y^j, X_{l_j}^{k_j}) | j = \overline{1, T}\} \quad (5.7)$$

де  $Y^j$  – нечітка множина, що відповідає значенню результуючої лінгвістичної змінної  $Y$ ,

$X_{l_j}^{k_j}$  – нечітка множина, яка відповідає значенню лінгвістичної змінної  $X_{l_j}$ , та значення функції належності якої  $\mu_{X_{l_j}^{k_j}}$  визначило значення функції належності  $\mu_{Y^j}$ .

З БОВ інформація передається на вхід дефаззифікатора БДК, який перетворює отриманні нечіткі дані, в формат, який відповідає формату даних отриманих для модельованої змінної  $y$  та вхідної змінної  $x_y$ , яка вплинула на значення змінної  $y$ :

$$\text{БДК} : \{(Y^j, X_{l_j}^{k_j}) | j = \overline{1, T}\} \xrightarrow{\text{дефаззифікація}} (y, x_y) \quad (5.29)$$

Тоді обробка інформації з використанням підсистеми нечіткого виведення буде здійснюватись за формулою .

Алгоритм, який відповідає модифікованому методу нечіткого виведення на основі деревоподібної БНЗ наведено нижче:

1. Визначити значення функцій належності термів вхідних змінних, за допомогою визначених функцій фаззифікації, якщо дані подаються в чисельному вигляді, або експертно, якщо вхідні дані вказані як такі, що визначаються експертно .

$$\mu_{X_{1i}^c}(x_i), \quad i = \overline{1, I_1}, c = \overline{1, K_{1i}}$$

2. Провести нечітке виведення та взяття композиції. В результаті для кожного терму всіх лінгвістичних змінних рівня обчислюються функції належності

$$\mu_{X_{2l}^z} \left( X_{1q_1}^{k_{1q_1}^{lj}}, X_{1q_2}^{k_{1q_2}^{lj}}, \dots, X_{1q_s}^{k_{1q_s}^{lj}} \right) \quad l = \overline{1, I_2}, z = \overline{1, K_{2l}}, \quad S = S_{nlj}^z, \quad \text{які}$$

відповідають вектору значень вхідних змінних.

3. Визначити значення  $p[a]$ ,  $s[a]$

$$X_{(n-1), p}^{s*} = \arg \min_{s=1, \overline{S}} \arg \max_{j=1, \overline{M_{nl}^z}} \left( \mu_{X_{nl}^z} \left( X_{(n-1), q_1}^{k_{(n-1), q_1}^{lj}}, X_{(n-1), q_2}^{k_{(n-1), q_2}^{lj}}, \dots, X_{(n-1), q_s}^{k_{(n-1), q_s}^{lj}} \right) \right) \quad (5.30)$$

$$S = S_{nlj}^z$$

$$p[a] = p^*, s[a] = s^*, a = \overline{1, a_{(n-1),l}^z} \quad (5.31)$$

де  $a_{(n-1),l}^z$  - кількість термів лінгвістичних змінних попереднього рівня, значення функції належності яких співпадає з значенням функції належності терма оцінюваної лінгвістичної змінної.

4. Повторювати пп. 2, 3 доти, доки не буде отримано значення функції належності для всіх термів результуючої змінної  $\mu_{X_{Nw}^g}$ ,  $g = \overline{1, K_{Nw}}$ , та визначені всі терми вершин попереднього рівня, що впливають.

5. Провести дефаззіфікацію шляхом визначення терму результуючої змінної з найбільшим значенням функції належності  $X_{Nw}^{g*}$ .

6. Побудувати шлях до найвпливовішого параметру, та визначити найвпливовіший параметр.

### 5.5. Представлення верхньої частини бази нечітких знань як класичної логічної

Для додаткового спрощення виведення в роботі запропоновані обмеження, за умови яких верхню частину деревоподібної БНЗ можна використовувати як класичну логічну БЗ.

Твердження 5.4 Якщо вірні наступні умови:

- 1) деревоподібна БНЗ побудована для вирішення задачі класифікації;
- 2) належність класу визначається шляхом нечіткого виведення Мамдани для задач класифікації;
- 3) часткові БНЗ, що визначають залежність ЛЗ вище другого рівня від ЛЗ попереднього рівня є повними;
- 4) значення ФН термів ЛЗ другого рівня, які визначають поточні значення цих ЛЗ, більше ніж значення ФН інших термів ЛЗ другого рівня

То

1) верхню частину деревоподібної БНЗ можна розглядати як класичну логічну БЗ, тобто розв'язки задачі класифікації при виведенні по класичній логічній БЗ і БНЗ будуть збігатися;

2) значення ФН результуючого терму модельованої ЛЗ, який буде отримано як результат виведення, дорівнюватиме мінімуму серед значень ФН термів ЛЗ другого рівня, які визначають значення цих ЛЗ.

Доведення

Перепишемо четвертий критерій. Він означає, що для ФН термів ЛЗ 2-го рівня вірна умова  $\mu_{X_{2l_s}^{z_1}} > \mu_{X_{2l_r}^{z_k}}, \forall s, r, k \cup k \neq 1$ , де  $\mu_{X_{2l_s}^{z_1}}$  - ФН  $z_1$ -го терму  $l_s$ -ї ЛЗ, що визначає значення ЛЗ, її клас належності.

Відзначимо, що оскільки БНЗ має деревоподібну структуру, то кожна її частина також буде мати деревоподібну структуру.

Для доведення твердження необхідно також показати, що для кожного рівня вірна умова, що значення ФН термів ЛЗ рівня, які визначають поточну

належність до певного класу, більше ніж значення ФН інших термів ЛЗ рівня.

Доведення цієї умови та всього твердження буде проводитися по індукції.

Покажемо вірність твердження для частин загальної БНЗ, а саме для деревоподібних БНЗ, які складаються з 3-х рівнів.

*База нечітких знань.* У випадку трирівневої БНЗ, верхня частина цієї БЗ складається лише за часткової БНЗ, яка визначає залежність ЛЗ третього рівня від ЛЗ другого рівня. Оскільки така часткова БНЗ повна, то існує правило, у якому задіяні всі  $z_1$ -і терми ЛЗ 2-го рівня, через які визначається значення ЛЗ третього рівня. В цьому випадку значення ФН такого терму дорівнює мінімуму зі значень ФН термів ЛЗ лівої частини, тобто  $\min_{s=1,p} \{ \mu_{X_{2l_s}^{z_1}} \}$ .

Кожне інше правило в частковій БНЗ буде мати в своєму складі хоча б один терм, з ФН меншою ніж  $\min_{s=1,p} \{ \mu_{X_{2l_s}^{z_1}} \}$ , тому значення ФН терму в правій частині буде менше ніж  $\min_{s=1,p} \{ \mu_{X_{2l_s}^{z_1}} \}$ . Відзначимо, що значення ФН всіх інших термів ЛЗ третього рівня буде менше ніж  $\min_{s=1,p} \{ \mu_{X_{2l_s}^{z_1}} \}$ .

Тому можна стверджувати, що значення ФН термів ЛЗ третього рівня, які визначають поточну належність до певних класів, більше ніж значення ФН інших термів ЛЗ третього рівня та дорівнює мінімуму ФН термів ЛЗ другого рівня, що визначають поточну належність до певних класів.

Оскільки рішення про значення ЛЗ, тобто класу, до якого відноситься поточна ситуація, визначається через знаходження максимуму ФН, то це буде як раз той терм, що буде визначатися правилом, в якому задіяні всі  $z_1$ -і терми лінгвістичних змінних. Це означає, що результуючий клас буде визначатися тим правилом, у якому перебувають терми, які були обрані як результуючі класи на попередньому рівні, а значення його ФН дорівнюватиме  $\min_{s=1,p} \{ \mu_{X_{2l_s}^{z_1}} \}$ .

*Класична логічна БЗ.* У випадку класичної БЗ буде спрацьовувати лише те правило, у якому є терми ЛЗ 2-го рівня, належність до яких як до класів була встановлена на попередньому рівні. Отже, як результат буде отриманий той же клас, що й при виведення на основі БНЗ.

Робимо індуктивне припущення, що твердження вірне для частин загальної БНЗ, а саме деревоподібних БНЗ з кількістю рівнів  $(n-1)$ , та значення ФН термів ЛЗ  $(n-1)$ -го рівня, які визначають поточну належність до певних класу, більше ніж значення ФН інших термів ЛЗ  $(n-1)$ -го рівня і дорівнюють мінімуму ФН термів ЛЗ другого рівня, що визначають поточну належність до певних класів, та з якими ЛЗ  $(n-1)$ -го рівня зв'язані.

Покажемо, що твердження вірне для деревоподібної БНЗ з кількістю рівнів  $n$ .

*База нечітких знань.* Деревоподібна БНЗ з кількістю рівнів  $n$

складається з деревоподібних БНЗ з кількістю рівнів  $(n-1)$  та часткової БНЗ, яка визначає залежність ЛЗ  $n$ -го рівня від ЛЗ  $(n-1)$ -го рівня. Оскільки така часткова БНЗ повна, то існує правило, у якому задіяні всі  $z_1$ -і терми ЛЗ  $(n-1)$ -го рівня, через які визначається значення ЛЗ  $n$ -го рівня. В цьому випадку значення ФН такого терму дорівнює мінімуму зі значень ФН термів ЛЗ лівої частини, тобто  $\min_{s=1, p_{n-1}} \{ \mu_{X_{(n-1)l_s}^{z_1}} \}$ .

Кожне інше правило в частковій БНЗ буде мати в своєму складі хоча б один терм, з ФН меншою ніж  $\min_{s=1, p_{n-1}} \{ \mu_{X_{(n-1)l_s}^{z_1}} \}$ , тому значення ФН терму в правій частині буде менше ніж  $\min_{s=1, p_{n-1}} \{ \mu_{X_{(n-1)l_s}^{z_1}} \}$ . Значення функцій належності всіх інших термів ЛЗ  $n$ -го рівня буде менше ніж  $\min_{s=1, p_{n-1}} \{ \mu_{X_{(n-1)l_s}^{z_1}} \}$ .

Використовуючи індуктивне припущення отримуємо, що  $\min_{s=1, p_{n-1}} \{ \mu_{X_{(n-1)l_s}^{z_1}} \} = \min_{s=1, p} \{ \mu_{X_{2l_s}^{z_1}} \}$ .

Оскільки рішення про значення ЛЗ, тобто класу, до якого відноситься поточна ситуація, визначається через знаходження максимуму ФН, то це буде як раз той терм, що буде визначатися правилом, в якому задіяні всі  $z_1$ -і терми ЛЗ  $(n-1)$ -го рівня. Це означає, що результуючий клас буде визначатися тим правилом, у якому перебувають терми, які були обрані як результуючі класи на попередньому рівні. А значення ФН цього терму дорівнюватиме мінімуму серед значень ФН термів ЛЗ другого рівня, які визначають належність до певного класу.

*Класична логічна БЗ.* Для визначення ЛЗ  $n$ -го рівня буде використовуватись правило, яке буде поєднувати значення ЛЗ  $(n-1)$ -го рівня, які отримані в результаті логічного виведення. А оскільки твердження вірне для частин загальної БНЗ, а саме деревоподібних БНЗ з  $(n-1)$  рівнями, то як результат буде отриманий той же клас, що й при виведенні на основі деревоподібної БНЗ.

Твердження доведене.

Зазначимо, що за умови виконання твердження 5.4 зменшується обчислювальна складність та час виведення за рахунок використання схеми логічного виведення, але все одно можна встановити значення ФН результуючого терму ЛЗ верхнього рівня, через знаходження мінімуму серед значень ФН термів ЛЗ другого рівня, які визначають належність до певного класу.

Твердження 5.5 Якщо вірні умови:

- 1) належність класу визначається шляхом нечіткого виведення Мамдані для задач класифікації;
- 2) для сусідніх термів всіх ЛЗ першого рівня ФН визначені таким чином, що вони перетинаються на однаковій висоті;
- 3) для кожної ЛЗ встановлено ранжування термів за важливістю;

4) у випадках коли значення відповідної до входу ФН знаходиться на перетині функцій, то робиться зменшення значення ФН терму з меншою важливістю;

5) часткові БЗ, що визначають залежність ЛЗ другого рівня від ЛЗ першого рівня є повними

То виконується умова 4 Твердження 5.4.

Доведення

З 2-ї та 3-ї умови теореми випливає, що кожна вхідна змінна може відповідати лише одному класу, тобто не може бути двох термів ЛЗ першого рівня із однаковими значеннями ФН. Оскільки всі ФН для різних термів перетинаються на однаковій висоті (умова 1), то

$\mu_{X_{1i_g}^{j_1}} > \mu_{X_{1i_h}^{j_1}}, \forall g, h, d \cup d \neq 1$ , де  $\mu_{X_{1i_g}^{j_1}}$  - ФН  $j_1$ -го терму  $i_g$ -ї ЛЗ, що має ФН з найбільшим для даної ЛЗ значенням.

З умови 4 слідує, що існує таке правило, в якому задіяні всі  $j_1$ -і терми лінгвістичних змінних першого рівня, що визначають змінну другого рівня, тобто терми з максимальним значенням ФН. Значення ФН терма ЛЗ у правій частині дорівнює мінімуму зі значень ФН термів ЛЗ лівої частини, тобто

$$\min_{g=1,q} \left\{ \mu_{X_{1i_g}^{j_1}} \right\}.$$

Кожне інше правило в БЗ буде мати в своєму складі хоча б один терм, з ФН меншою ніж  $\min_{g=1,q} \left\{ \mu_{X_{1i_g}^{j_1}} \right\}$ , тому значення ФН терму в правій частині буде

менше ніж  $\min_{g=1,q} \left\{ \mu_{X_{1i_g}^{j_1}} \right\}$ . Звідси слідує, що значення ФН терму ЛЗ другого

рівня, що визначається правилом, в якому задіяні всі  $j_1$ -і терми лінгвістичних змінних першого рівня визначається наступним чином

$$\mu_{X_{2i_s}^{z_1}} = \min_{g=1,q} \left\{ \mu_{X_{1i_g}^{j_1}} \right\} \quad (2.32)$$

Використавши умову 1 та (2.32) отримуємо  $\mu_{X_{2i_s}^{z_1}} > \mu_{X_{2i_r}^{z_k}}, \forall s, r, k \cup k \neq 1$

Твердження доведено.

Однак повнота часткових БЗ нижнього рівня ускладнює інтерпретацію результату та робить базу менш прозорою, тому в наступних підрозділах запропоновані методи для спрощення БЗ, що визначають лінгвістичні змінні другого рівня. Зазначимо, що за умови неповноти цих баз знань твердження 5.5 не буде виконуватись. Це накладає додаткові вимоги до процесу навчання та формування цих баз знань, а саме необхідність виконання умови 4 твердження 5.4.

## 5.6. Метод обробки інформації на основі деревоподібних баз нечітких знань зі змішаною схемою виведення

Метод обробки інформації полягає в послідовному виконанні двох кроків (рисунок 5.7). По-перше, підготовки БНЗ, яка полягає в структуруванні та настроюванні БНЗ. По-друге, використання БНЗ для обробки інформації, яке полягає в використанні змішаної схеми виведення на основі деревоподібної БНЗ. Послідовність дій при структуруванні і настроюванні БНЗ показана на рисунку 5.8.

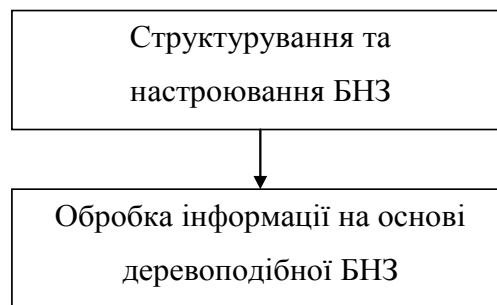


Рис. 5.7. Кроки використання методу обробки інформації

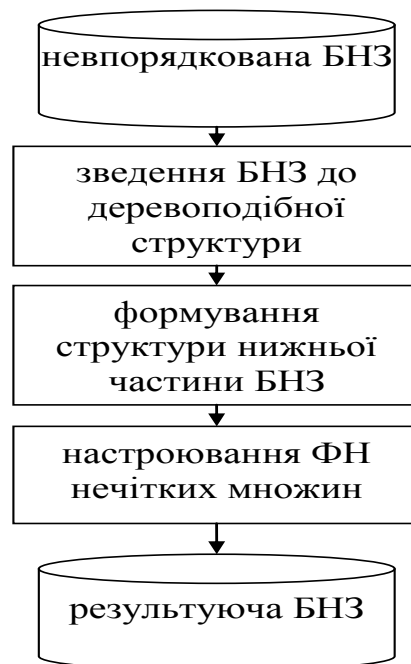


Рис. 5.8. Етапи структурування та настроювання БНЗ

Виходячи із сформованої неспорядкованої БНЗ та використовуючи підхід до зведення БНЗ до деревоподібної структури отримуємо деревоподібну БНЗ, що складається з двох частин нижньої і верхньої. Під час зведення проводиться перевірка того факту, що верхня частина деревоподібної БНЗ є повною, тобто перевіряється повнота кожної часткової БЗ, яка визначає залежність ЛЗ поточного рівня від ЛЗ попереднього рівня.

Як було відмічено вище навчальні вибірки, для яких відомі значення виходів, можуть бути запропоновані лише для частин предметної області, що



знаходяться на нижньому рівні ієрархії, тому автоматичні механізми формування та настроювання можливо застосовувати лише для нижньої частини деревоподібної БНЗ. Застосування таких механізмів полягає в формуванні часткових БНЗ, які визначають залежність ЛЗ другого рівня від ЛЗ першого рівня та настроюванні параметрів ФН нечітких множин, що відповідають термам ЛЗ першого рівня. Такі механізми добре розроблені для нейромереж, зокрема для нечітких нейромереж. Тому в роботі запропоновано на етапі формування і настроювання розглядати нижню частину деревоподібної БНЗ як нечітку нейромережу аналогічну NEFCLASS. Це дозволило використати розроблені для нечітких нейромереж механізми навчання. Суть етапів формування і настроювання розкрито в наступному розділі.

Після проведення цих етапів і об'єднання отриманої нижньої та верхньої частини отримуємо деревоподібну БНЗ, яка готова для використання.

Така БНЗ складає основу удосконаленої нечіткої логічної моделі об'єкту. Відзначимо, що удосконалення нечіткої логічної моделі полягає в зведенні БНЗ до деревоподібної структури та введенні обмежень, за умови яких частину БНЗ можна використовувати як класичну логічну БЗ. Ця нечітка модель використовується при вирішенні задачі обробки інформації на основі деревоподібних БНЗ зі змішаною схемою виведення, яка сформульована нижче.

Задача обробки інформації на основі деревоподібних БНЗ зі змішаною схемою виведення.

Дано:

1. Математична модель об'єкту

1)  $\{x_i \mid i = \overline{1, v}\}$  – множина вхідних змінних;

2)  $\{X_{li} \mid i = \overline{1, v}\}$  – множина ЛЗ, що відповідають  $\{x_i \mid i = \overline{1, v}\}$ ;

3)  $\{X_{li}^z \mid z = \overline{1, Z_{li}}\}$ ,  $i = \overline{1, v}$  – терм-множини ЛЗ 1-го рівня;

4)  $\{\mu_{X_{li}^z} \mid i = \overline{1, v}, z = \overline{1, Z_{li}}\}$  – множина ФН, відповідних до термів ЛЗ 1-го

рівня, за якими проводиться фаззифікація  $\{x_i \mid i = \overline{1, v}\}$ ;

5)  $\{X_{ri} \mid r = \overline{2, N-1}, i = \overline{1, M_r}\}$  – множина ЛЗ  $r$ -го рівня;

6)  $\{X_{ri}^z \mid z = \overline{1, Z_{ri}}\}$ ,  $r = \overline{2, N-1}, i = \overline{1, M_r}$  – терм-множини ЛЗ  $r$ -го рівня;

7)  $\{\mu_{X_{ri}^z} \mid r = \overline{2, N-1}, i = \overline{1, M_r}, z = \overline{1, Z_{ri}}\}$  – множина ФН, відповідних до

термів ЛЗ  $r$ -го рівня (5.6);

8)  $y$  – результуюча змінна;

9)  $Y$  – ЛЗ, що відповідає  $y$ ;

10)  $\{Y_i^z \mid z = \overline{1, T}\}$  – терм-множина ЛЗ  $Y$ ;

11)  $BZ = \bigcup_{n,g} \widehat{R}_{(n+1),g}$  – деревоподібна БНЗ;

12)  $\widehat{R}_{(n+1),g}$  – часткова БНЗ (5.2).

II. Вектор фіксованих значень вхідних змінних об'єкта  $x = \langle x_1^*, x_2^*, \dots, x_v^* \rangle$

Знайти:

Результуюче значення  $y^*$  та вхідний параметр  $x_y^*$ , який визначив таке значення  $y^*$

У загальному випадку пропонується розв'язувати задачу обробки інформації на основі деревоподібної БНЗ за допомогою модифікованого нечіткого виведення. Але у випадку виконання умов наведених в твердженні 5.4 верхню частину деревоподібної БНЗ можна розглядати як класичну логічну.

Відзначимо, що перші три умови твердження 5.4 перевіряються на етапі структурування та настроювання, а четверта умова перевіряється під час проведення обробки інформації.

Якщо ці умови виконуються, то для розв'язання задачі обробки інформації можна використовувати інший підхід, суть якого полягає у використанні змішаної схеми виведення, тобто застосуванні схеми нечіткого виведення на основі нижньої частини БНЗ та використанні схеми логічного виведення на основі верхньої частини БНЗ при виконанні твердження 5.4, яка в цьому випадку розглядається як класична логічна. У випадку невиконання твердження 5.4 використовується модифіковане нечітке виведення для верхньої частини БНЗ.

Зазначимо, що за умови використання логічного виведення на основі верхньої частини деревоподібної БНЗ все одно можна визначити значення ФН результуючого терму ЛЗ верхнього рівня, яка, як доведено в твердженні 5.4, визначається через знаходження мінімуму серед значень ФН термів ЛЗ другого рівня, які визначають належність до певного класу.

Алгоритм обробки інформації на основі деревоподібних БНЗ зі змішаною схемою виведення описано нижче (рисунки 5.9).

1. Отримати вхідні дані.

2. Визначити значення ФН термів ЛЗ першого рівня, за допомогою визначених функцій фазифікації, якщо дані подаються в чисельному вигляді, або експертно, якщо вхідні дані вказані як такі, що визначаються експертно.

3. Визначити значення ФН термів ЛЗ другого рівня за допомогою нечіткого виведення на основі нижньої частини деревоподібної БНЗ.

4. Якщо умова 4 твердження 5.4 вірна, а саме значення функцій належності термів лінгвістичних змінних другого рівня, які визначають поточну належність до певного класу, більше ніж значення функцій належності інших термів лінгвістичних змінних другого рівня, то перехід на п. 5, в протилежному випадку п. 7.

5. Визначити значення результуючої лінгвістичної змінної за допомогою логічного виведення на основі верхньої частини деревоподібної БНЗ.

6. Визначити результуюче значення ФН терму модельованої ЛЗ

обчисленням мінімуму серед значень ФН термів ЛЗ другого рівня, які визначають належність до певного класу. Перехід на п. 8.

7. Визначити значення модельованої лінгвістичної змінної за допомогою нечіткого виведення на основі верхньої частини деревоподібної БНЗ.

8. Вивести значення модельованої ЛЗ та значення результуючої ФН.

Запропонований метод обробки інформації дозволяє зменшити обчислювальну складність без втрати значення ФН терму результуючої ЛЗ за рахунок, по-перше, зведення БНЗ до деревоподібної структури, а, по-друге, використання змішаної схеми виведення на основі отриманої БНЗ.



Рис. 5.9. Алгоритм обробки інформації на основі деревоподібних БНЗ зі змішаною схемою виведення

## РОЗДІЛ 6. МЕТОДИКА СТВОРЕННЯ ДЕРЕВОПОДІБНОЇ БАЗИ НЕЧІТКИХ ЗНАНЬ

Специфіка предметної області, а саме її ієрархічність та неможливість визначення стану безпосередньо, а лише через стани частин впливає на формування та настроювання БНЗ і всієї підсистеми нечіткого виведення. Як було зазначено в попередньому розділі застосування автоматизованих механізмів побудови та настроювання можливо лише для нижньої частини деревоподібної бази нечітких знань, тобто для тих часткових БНЗ, що визначають залежність ЛЗ другого рівня від ЛЗ першого рівня.

Це обумовлено тим, що адекватні вибірки можуть бути запропоновані лише для частин предметної області, які знаходяться на нижньому рівні ієрархії. А вже починаючи з наступного рівня важко сказати, виходячи із вхідних даних, до якого класу належить дана ситуації визначена вхідним вектором, тому що кількість вхідних даних збільшується, а результуючу змінну не можна виміряти якимось іншим незалежним способом. Підхід до побудови та настроювання нижньої частини деревоподібної бази нечітких знань наведено нижче.

### 6.1. Підхід до побудови та настроювання бази нечітких знань

Постановка задачі побудови та настроювання нижньої частини деревоподібної БНЗ(рисунок 6.1) наведена нижче.

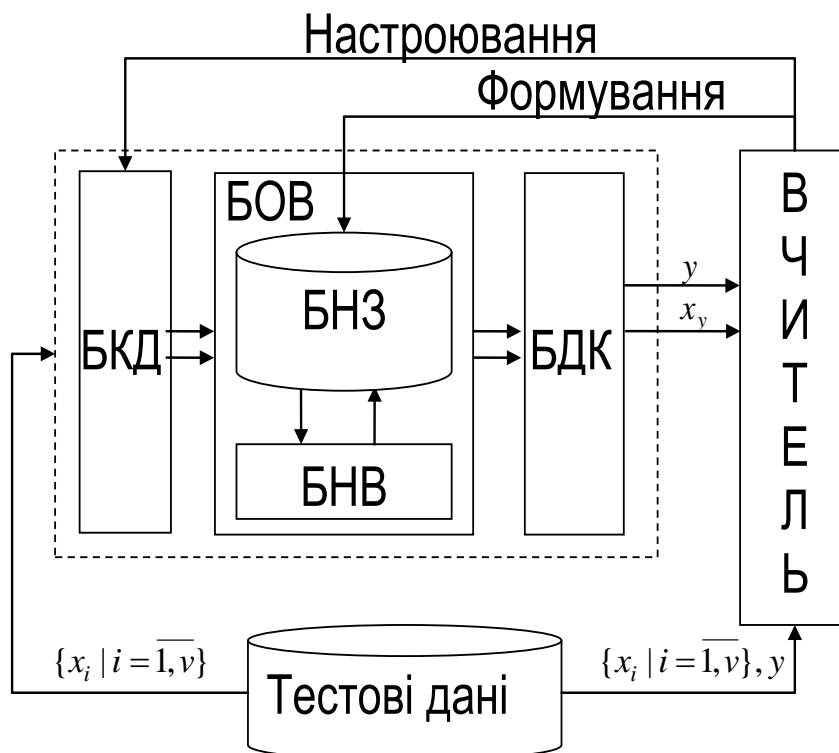


Рис. 6.1. Формування та настроювання БНЗ

Постановка задачі побудови та настроювання нижньої частини БНЗ.

Дано:

- 1)  $\{x_i | i = \overline{1, v}\}$  – множина вхідних змінних;
- 2)  $\{X_{1i} | i = \overline{1, v}\}$  – множина ЛЗ, що відповідають  $\{x_i | i = \overline{1, v}\}$ ;
- 3)  $\{X_{1i}^z | i = \overline{1, v}, z = \overline{1, Z_{1i}}\}$  – терм-множини ЛЗ 1-го рівня;
- 4)  $\{\mu_{x_{1i}^z} | i = \overline{1, v}, z = \overline{1, Z_{1i}}\}$  – множина ФН, відповідних до термів ЛЗ 1-го рівня, за якими буде проводитись фаззифікація;
- 5)  $\{X_{2i} | i = \overline{1, M_2}\}$  – множина ЛЗ 2-го рівня;
- 6)  $\{X_{2i}^z | z = \overline{1, Z_{2i}}\}$  – терм-множина ЛЗ  $X_{2i}$ ;
- 7)  $\{(x^l, X_2^l) | l = \overline{1, R_{Study}}\}$  – вибірка для побудови та настроювання, де  $x^l = \langle x_1^l, x_2^l, \dots, x_v^l \rangle$  –  $l$ -й набір значень вхідних змінних,  $X^l = \langle X_1^l, X_2^l, \dots, X_{N_2}^l \rangle$  –  $l$ -й набір значень ЛЗ 2-го рівня;
- 8)  $\{(x^l, X_2^l) | l = \overline{1, R_{Test}}\}$  – вибірка для тестування

Знайти:

1)  $\widehat{R}_2 = \bigcup_l \widehat{R}_{2l}$  – базу нечітких знань, що визначає залежність ЛЗ 2-го рівня від ЛЗ 1-го рівня;

2) Настроїти  $\{\mu_{x_{1i}^z} | i = \overline{1, v}, z = \overline{1, Z_{1i}}\}$ .

Задачу побудови та настроювання пропонується вирішувати з використанням підходів, що розроблені для нечіткої нейромережі NEFCLASS з деякою модифікацією. Модифікація буде полягати в тому, що система NEFCLASS це є тришаровий нечіткий перцептрон, в той час як структура нечіткої нейронної мережі, що відповідає базі знань другого рівня буде складатися з чотирьох шарів. Для цього часткові БНЗ другого рівня представляються як нечіткі нейромережі.

Виділимо два основних етапи побудови та настроювання БНЗ (рисунок 6.2), які в подальшому розглядаються з точки зору навчання нечіткої нейромережі.

Перший етап полягає в побудові структури нечіткої нейронної мережі, а другий в настроюванні функцій належності через навчання нейронної мережі.

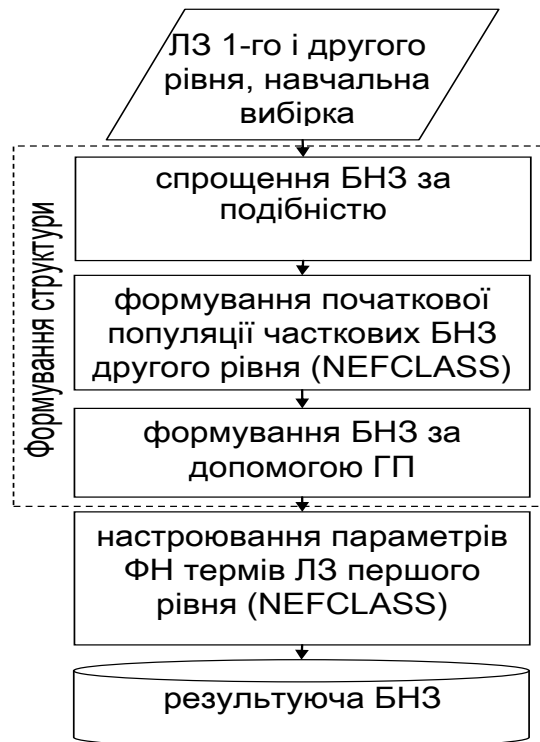


Рис. 6.2. Етапи побудови та настроювання ФНЗ

У випадках, коли не можна відокремити часткові бази знань другого рівня, тобто частина ЛЗ першого рівня використовується для визначення термів декількох ЛЗ другого рівня, то навчання таких ФНЗ проводиться наступним чином. Етап формування структури проводиться окремо для кожної часткової ФНЗ. У випадку коли однакові терми ЛЗ першого рівня визначають терми різних ЛЗ другого рівня настроювання параметрів ФН проводиться для мережі, яку буде отримано при об'єднанні мереж, що відповідатимуть частковим ФНЗ для визначення різних ЛЗ.

Настроювання часткових ФНЗ для визначення ЛЗ другого рівня полягає в послідовному проведенні етапів, наведених на рисунку 6.2. Тобто спочатку проводиться спрощення ФНЗ, після цього формується початкова ФНЗ, потім застосовується ГП, а після цього проводиться навчання нейронної мережі.

Однак, іноді кращих результатів можна досягти при циклічному застосуванні розглянутих нижче підходів та їх змішаному використанні.

## 6.2. Представлення бази нечітких знань у вигляді нейромережі

Задачу настроювання ФНЗ можна вирішити, якщо розглядати кожну часткову ФНЗ, що визначає залежність лінгвістичної змінної другого рівня від лінгвістичних змінних першого рівня, як чотиришарову нечітку нейронну мережу (рисунок 6.3).

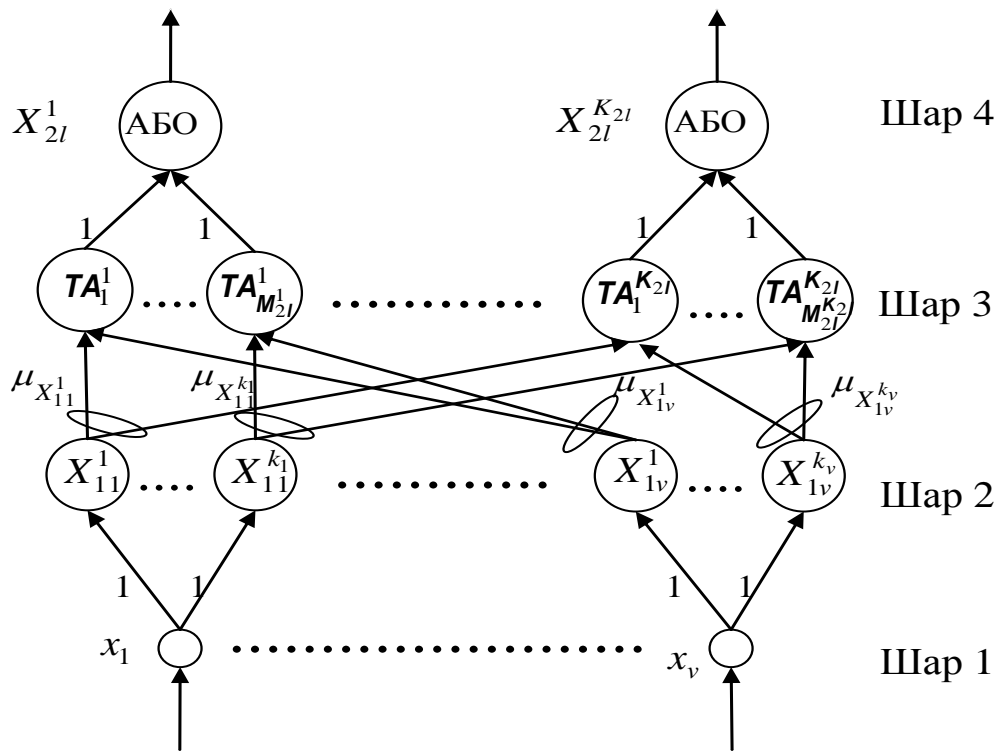


Рис. 6.3. Структура нечіткої нейронної мережі, що відповідає частковій БНЗ для визначення залежності ЛЗ другого рівня від ЛЗ першого рівня

Структура мережі визначається наступним чином:

Шар 1 – входні нейрони, що відповідають входам об'єкту;

Шар 2 – нейрони першого схованого шару, що відповідають термам, які використовуються в частковій БНЗ, та належать до терм-множин ЛЗ першого рівня;

Шар 3 – нейрони другого схованого шару, які визначають мінімальне значення серед значень, що до них надійшли;

Шар 4 – нейрони вихідного шару, які визначають максимальне значення серед значень, що до них надійшли.

Вага зв'язків між нейронами рівнів:

одиночку – вага між нейронами 1-го та 2-го шарів;

функціями належності входу до нечіткого терму – вага між нейронами 2-го і 3-го шару;

одиночку – вага між нейронами 3-го та 4-го шару.

Будемо розглядати обчислення результуючої змінної за даними виміру входніх змінних як послідовну процедуру в багатошарових нейромережах, де входні сигнали поширюються в прямому напрямку (feedforward), але якщо дійсні значення виходів відрізняються від бажаних, то помилка поширюється у зворотному напрямку з урахуванням величин, розрахованих під час прямого ходу.

В якості функцій належності термів ЛЗ, що відповідають нейронам 2-го шару будуть використовуватися функції належності трикутного виду (6.8), які є універсальним апроксиматором:

$$\mu: \mathfrak{R} \rightarrow [0,1]$$

$$\mu(x) = \begin{cases} \frac{x-a}{b-a}, & x \in [a,b), \\ \frac{c-x}{c-b}, & x \in [b,c], \\ 0, & x \notin [a,c]. \end{cases} \quad (6.8)$$

Кожен нейрон 3-го шару  $TA_j^z$  ( $j = \overline{1, M_{2l}^z}$ ,  $z = \overline{1, K_{2l}}$ ) та пов'язані з ним нейрони другого шару можна розглядати як рядок кон'юнкції часткової БНЗ для визначення залежності ЛЗ другого рівня  $X_{2l}$  від ЛЗ першого рівня, що визначає  $z$ -й терм цієї змінної, та переписати у вигляді правил «ЯКЩО..., ТО...» наступним чином (6.2):

$$\begin{aligned} & \text{ЯКЩО} \left( X_{1q_1} = X_{1q_1}^{k_{1q_1}^{lj}} \right) TA \left( X_{1q_2} = X_{1q_2}^{k_{1q_2}^{lj}} \right) TA \dots \\ & TA \left( X_{1q_s} = X_{1q_s}^{k_{1q_s}^{lj}} \right) TO X_{2l} = X_{2l}^z \end{aligned} \quad (6.2)$$

$j = \overline{1, M_{2l}^z}$ ,  $z = \overline{1, K_{2l}}$ ,  $S = S_{2,l}^z$

де  $q_s$  – номер ЛЗ на рівні 2, а  $s = \overline{1, S}$ ,  $S = S_{2,l}^z$  – номер ЛЗ у  $j$ -му правилі ( $j = \overline{1, M_{2l}^z}$ ) для визначення  $z$ -го терму ( $z = \overline{1, K_{2l}}$ )  $l$ -тої ЛЗ 2-го рівня.

Значення функції активації цього нейрону буде визначатись за формулою (6.8) враховуючи те, що  $\mu_{TA_j^z} = \mu_{X_{2l}^z}$ .

Запишемо алгоритм поширення сигналу, виходячи з того, що нечітка нейронна мережа, яка розглядається є мережею прямого поширення.

- 1) Вхідний вектор поступає на 1-й шар, в якому знаходяться нейрони, що відповідають змінним предметної області.
- 2) Передаємо значення відповідного елемента вектора до нейронів, які відповідають термам лінгвістичної змінної, зіставленої з вхідною змінною.
- 3) Обчислюємо значення вагових коефіцієнтів між другим шаром-шаром термів та третім шаром, тобто значення функцій належності термів

$$\mu_{X_{1q_r}^{k_{1q_r}^{lj}}} \quad (6.1)$$

4) Обчислюємо значення нейронів третього шару за формулою (6.8) та передаємо його на наступний шар.

5) Обчислюємо нейрони, що відповідають значенням класу правил,



тобто базі нечітких знань, що визначає значення функції належності терму лінгвістичної змінної  $X_{2l}^z$ :

$$\mu_{X_{2l}^z} = \bigvee_{j=1, M_{2l}^z} (\mu_{X_{2l}^z})_j = \bigvee_{j=1, M_{2l}^z} \left( \mu_{X_{1q_1}^{k_{1zj}}} \wedge \mu_{X_{1q_2}^{k_{2zj}}} \wedge \dots \wedge \mu_{X_{1q_S}^{k_{Szj}}} \right)$$

$$S = S_{2lj}^z$$

б) Визначаємо результуюче значення на виході мережі, яке представляє собою клас до якого необхідно віднести даний вектор:

$$\mu_{X_{2l}^z}^* = \bigvee_{z=1}^{n_{2l}} \mu_{X_{2l}^z} = \max_{z=1, n_{2l}} \mu_{X_{2l}^z}$$

### 6.3. Формування структури бази нечітких знань

Етап формування структури нижньої частини бази нечітких знань буде полягати в послідовному використанні описаних нижче підходів.

#### Спрощення бази нечітких знань за подібністю

Метод спрощення бази нечітких знань за подібністю полягає в використанні міри подібності для визначення надлишковості нечітких множин в базі знань. Міра подібності базується на теоретикомножинних операціях перетину та об'єднання та визначається наступним чином (6.3)

$$S(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (6.3)$$

де  $A, B$  – нечіткі множини,

$|\cdot|$  – означає потужність множини.

Якщо  $S(A, B) = 1$ , то функції належності, що відповідають нечітким множинам  $A$  та  $B$  рівні, а  $S(A, B) = 0$  означає, що функції належності не пересікаються.

Подібні нечіткі множини об'єднуються у випадку, коли міра подібності перевищує заданий експертом поріг  $\theta \in [0, 1]$ . Злиття множин зменшує кількість різних нечітких множин, що відповідають термам лінгвістичних змінних, які використовуються при побудові моделі, і тим самим підвищує прозорість.

У випадку, коли всі терми лінгвістичної змінної мають такі нечіткі множини, які відповідають універсальній множині, або якщо злиття призводить до того, що терм-множина лінгвістичної змінної містить лише один терм, то такі лінгвістичні змінні (властивості) виключаються з моделі (рисунок 6.4).

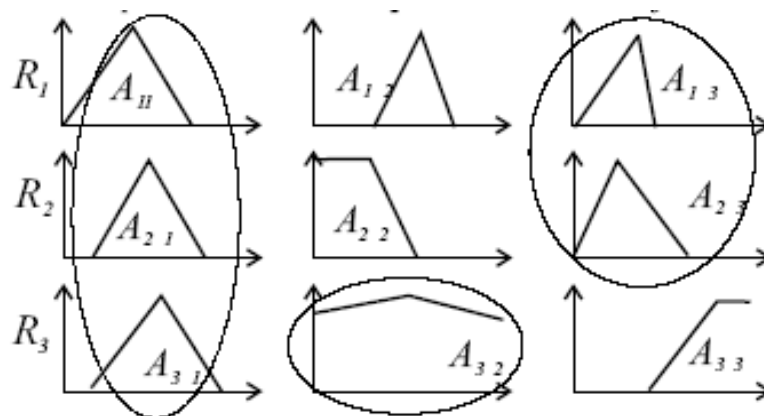


Рис. 6.4. Спрощення за подібністю

### Побудова початкової бази нечітких знань нижнього рівня

Для побудови структури бази нечітких знань будуть використовуватись підходи, що базуються на еволюційних обчисленнях, серед яких можна виділити генетичні алгоритми (ГА) та генетичне програмування (ГП). ГА добре показали себе для вирішення задач побудови нечітких правил з вихідних даних. Але описані в цих роботах підходи мають недолік, який полягає в тому, що отримані на виході правила містять всі лінгвістичні змінні. Це ускладнює як саму систему правил, так і її інтерпретацію, тобто за допомогою неї не можна визначити які саме параметри більш за все впливають на таке значення результуючої змінної.

В [5] запропоновано підхід, що базується на ГП та дозволяє будувати модель, яка може складатися з правил, до складу яких не обов'язково повинні входити всі лінгвістичні змінні. До того ж, в порівнянні з ГА, ГП має більш простіше та більш спрямоване відображення генотип-фенотип в застосуванні задачі формування структури [5].

Однак, всі вищезазначені підходи мають суттєвий недолік, вони починають навчання з довільного випадкового набору хромосом батьків, що сповільнює та ускладнює побудову БЗ.

В даній роботі пропонується для побудови початкової бази знань, яка буде відігравати роль набору хромосом використовувати модифікацію підходу до побудови системи NEFCLASS.

Кожна побудована під час навчання частина мережі, що відповідатиме вузлу правила та її складові частини, терми, будуть розглядатися як елементи початкового набору генів при використанні генетичного програмування для побудови часткової БЗ для визначення лінгвістичної змінної другого рівня.

На початку маємо нечітку нейронну мережу з  $v$  вхідними нейронами  $\{x_1, x_2, \dots, x_v\}$  і  $K_{2l}$  нейронами виходу  $\{X_{2l}^1, X_{2l}^2, \dots, X_{2l}^{K_{2l}}\}$ . Нехай в нас є навчальна множина зразків  $L = \{(p_s, t_s)\}$ , кожен з яких складається із вихідного зразка  $p \in R^v$  й бажаного зразка  $t \in \{0,1\}^{K_{2l}}$ .

Навчальний алгоритм, мета якого створити початкову структуру мережі, тобто створити нейрони правил складається з наступних етапів:

1. Вибрати наступний зразок  $(p, t) \in L$ .

2. Знайти ФН  $\mu_{X_{li}^j}(p^i)$  для кожного з нейронів  $X_{li}^j$  ( $j = \overline{1, k_i}$ ), що відповідають термам ЛЗ  $X_{li}$  ( $i = \overline{1, K_{2l}}$ ) пов'язаної з вхідним нейроном  $x_i$ , де  $p^i$  означає  $i$ -ту компоненту вектора  $p$ .

3. Для кожної ЛЗ знайти нейрон  $X_{li}^{j*}$ , який має максимальну ФН  $\mu_{X_{li}^{j*}}(p^i) = \max_{j \in \overline{1, k_i}} \{\mu_{X_{li}^j}(p^i)\}$

4. Якщо не існує вузла ТА такого, що з'єднується з нейронами другого шару  $X_{li}^{j*}$ ,  $i = \overline{1, K_{2l}}$ , то п. 5, інакше п.6

5. Створити вузол ТА, що з'єднується з нейронами другого шару  $X_{li}^{j*}$ ,  $i = \overline{1, K_{2l}}$ , і з'єднати його з вихідним вузлом  $X_{2l}^z$ , якщо  $t_z = 1$ .

6. Якщо ще залишилися неопрацьовані зразки в  $L$ , то переходимо на крок 1, а інакше стоп.

7. Обробляємо зразки в  $L$  і накопичуємо активації кожного нейрона ТА для кожного класу зразків, які були поширені. Якщо нейрон ТА показує більше нагромадження активації для класу  $X_{2l}^j$ , чим для класу  $X_{2l}^i$ , що був специфікований для наслідку правила, тоді змінюємо наслідок ТА на  $X_{2l}^j$ , тобто з'єднуємо ТА з нейроном виходу  $X_{2l}^j$ . Продовжуємо обробку зразків в  $L$  далі й обчислюємо для кожного нейрона ТА:

$$V_{TA} = \sum_{p \in L} \mu_{TA}(p) \cdot e_p, \quad (6.4)$$

$$e_p = \begin{cases} 1, & \text{якщо } p \text{ класифіковано вірно} \\ -1, & \text{інакше} \end{cases}$$

8. Для кожного нейрона  $\{X_{2l}^1, X_{2l}^2, \dots, X_{2l}^{K_{2l}}\}$  залишаємо  $k$  нейронів ТА з найвищими значеннями  $V_{TA}$  й видаляємо інші нейрони правил із мережі.

### **Формування нижньої частини бази нечітких знань методами генетичного програмування**

Формування БНЗ для визначення кожної лінгвістичної змінної другого рівня буде проводитись окремо для кожного її терму. Це обумовлено тим, що етап формування структури є незалежним для кожного з термів ЛЗ. Часткова БНЗ для визначення лінгвістичної змінної буде отримана шляхом об'єднання сформованих частин БНЗ для визначення термів, які подаються у вигляді правил «ЯКЩО.., ТО..» (6.3).

ГП оперує множиною можливих правил (6.3) для визначення терму ЛЗ. Ці правила подаються у вигляді структур даних – хромосом. Набір хромосом називається популяція. Тоді кожен елемент популяції представляє нечітку модель для визначення терму та може бути поданий у вигляді графу «ТА/АБО» (рисунок 6.2).

Хромосоми оцінюються за допомогою функції відповідності на здатність їх розв'язувати проблему, тобто вірно класифікувати. Ця функція включає всі умови, які висувуються до розв'язку проблеми, тобто до БНЗ терму. Результат оцінювання, тобто ті хромосоми що залишилися, використовуються в процесі формування нової множини можливих розв'язків, нової популяції. Вибір відповідних хромосом називається селекцією. Під час неї вибираються хромосоми із найвищою функцією відповідності. Додатково використовуються операції схрещування та мутації, які застосовуються до хромосом однієї популяції. Ці операції модифікують структуру нечіткої моделі. Схрещування забезпечує обмін інформацією між хромосомами та застосовується до випадково обраної пари хромосом з поточної популяції з ймовірністю  $p_c$ . Мутація, з іншого боку, гарантує урізноманітнення, яке необхідно для еволюціонування. Вона застосовується з ймовірністю  $p_m$ . Це все призводить до створення нової популяції та процес повторюється знову.

В кінці кожної епохи проходить селекція хромосом, яка полягає в оцінюванні за допомогою функції відповідності  $Fit\_Fun$  (6.5) хромосом на здатність їх розв'язувати проблему, тобто вірно класифікувати, та формуванні нової популяції. Умова закінчення записується наступним чином  $1 - Fit\_Fun < \varepsilon$ , де  $\varepsilon$  – поріг, при досягненні якого можна говорити, що отримана часткова БНЗ задовольняє вимогам.

$$Fit\_Fun = \frac{\sum_{p \in L} e_p}{|L|}, \quad (6.5)$$

$$e_p = \begin{cases} 1, & (\mu_{x_{2l}^z}(p) \geq \alpha \text{ та } t_z = 1) \text{ або } (\mu_{x_{2l}^z}(p) < \alpha \text{ та } t_z = 0) \\ 0, & (\mu_{x_{2l}^z}(p) \geq \alpha \text{ та } t_z = 0) \text{ або } (\mu_{x_{2l}^z}(p) < \alpha \text{ та } t_z = 1) \end{cases}$$

де  $|L|$  – потужність множини  $L$ ;

$\alpha$  – заданий експертом поріг, при досягненні якого можна говорити, що даний вектор належить класу.

Якщо після проходження заданої кількості епох не знайдена хромосома, що задовольняє умові закінчення, то змінюються параметри і алгоритм виконується знову. У протилежному випадку хромосома переписується у вигляді правила "ЯКЩО..., ТО...", яке є частиною часткової нечіткої БЗ для визначення залежності ЛЗ другого рівня від ЛЗ першого рівня.

Розглянемо докладніше операції схрещування та мутації, які застосовуються в ГП по відношенню до побудови нечіткої бази знань для визначення терму ЛЗ. Роль операції схрещування полягає в дослідженні простору пошуку в процесі обміну хромосом з популяції своїми частинами. Дві випадково обраних хромосоми «розрізаються» та «з'єднуються» для «вироблення» нової хромосоми. В даній роботі пропонується використовувати дві операції схрещування. Одна з них здійснюється на рівні

входів вузла АБО. Тобто схрещування призводить до обміну одним або декількома вузлами ТА. Приклад застосування описаного механізму схрещування показано на рисунку 6.5.

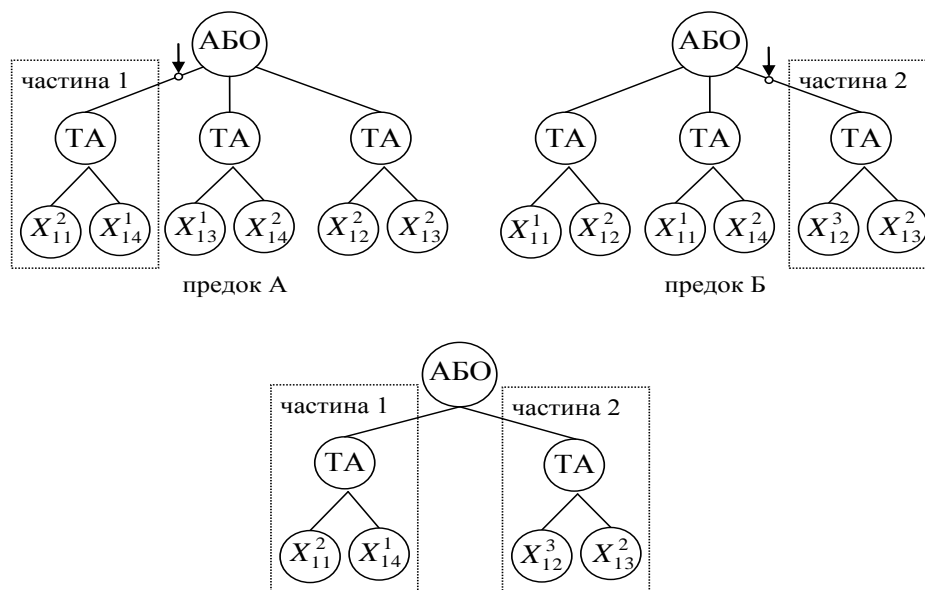


Рис. 6.5. Операція схрещування на рівні входів вузла АБО

Обираються дві випадкові точки в хромосомах-предках А і Б. Відсічені частини з обох предків об'єднуються та створюється нащадок.

Друга операція схрещування здійснюється на рівні входів вузла ТА. В цьому випадку процес обміну обмежується лише входами вузла ТА. Приклад операції схрещування на рівні вузла ТА показано на рисунку 6.6.

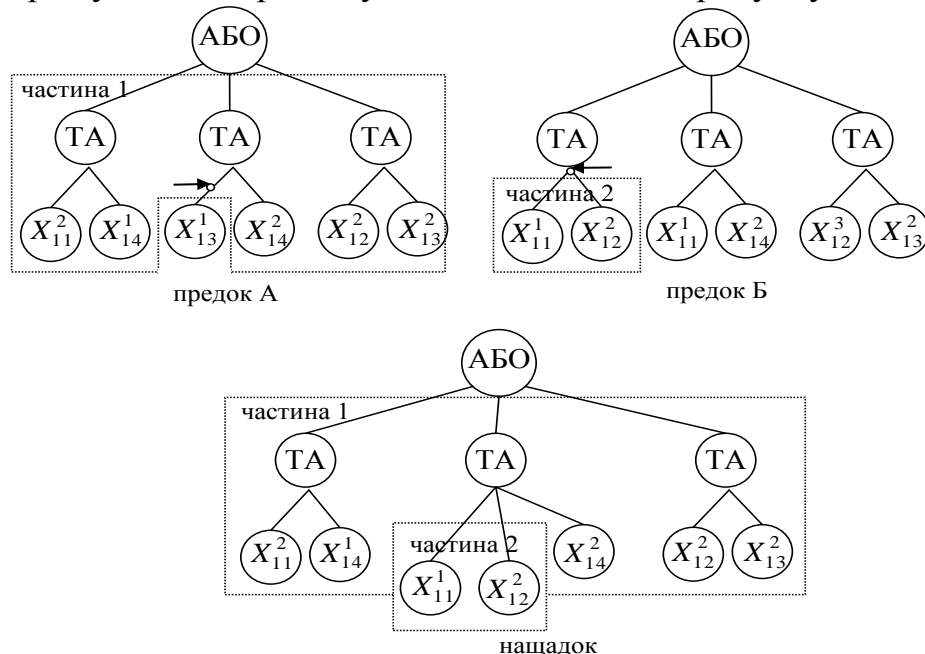


Рис. 6.6. Операція схрещування на рівні входів вузла ТА

Як видно з рисунку множина входів з випадково обраного вузла ТА заміщається множиною входів з випадково обраного вузла ТА іншого

елементу популяції.

Використовуючи два механізми схрещування ми вносимо зміни різного рівня в процес створення нащадків. Слід відмітити, що по своїй природі ГП в кожному новому випадку створює різні нащадки для одних й тих самих предків. Це може бути корисним в процесі пошуку оптимального рішення.

Метою використання операції мутації є внесення різноманітності в популяцію та сприяння знаходженню оптимального рішення в просторі пошуку. В процесі мутації будується нова хромосома внесенням невеликих змін в структуру хромосоми-предка. В даній роботі розглядається два типи мутації. В першому випадку мутація відбувається на рівні вузла ТА (рисунок 6.7), в другому, на рівні вузла АБО (рисунок 6.8).

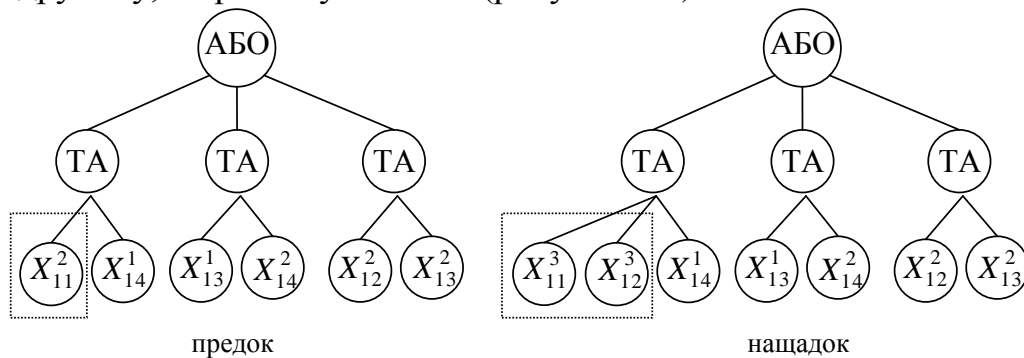


Рис. 6.7. Мутація на рівні вузла ТА

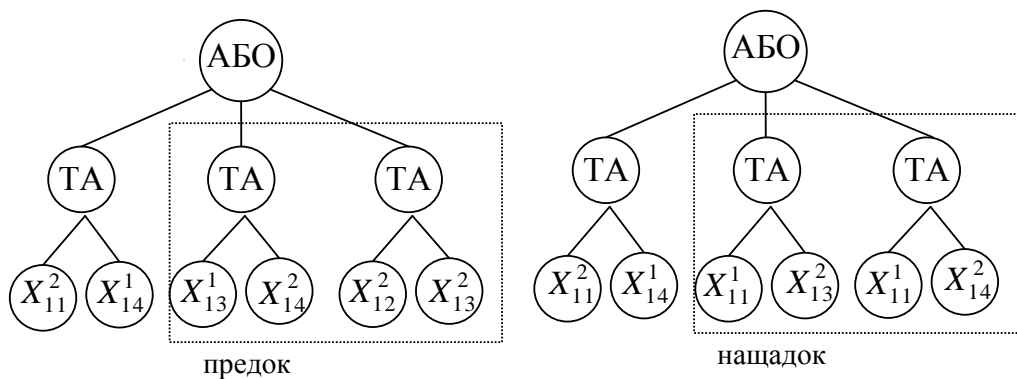


Рис. 6.8. Мутація на рівні вузла АБО

Важливим кроком в ГП є формування початкової популяції. Всі хромосоми, що відповідають нечітким моделям генеруються випадково з обмеженням на те, що вузол ТА не може поєднувати два терми однієї лінгвістичної змінної. Таке саме обмеження накладається на сформовані за допомогою операції мутації нащадки. В якості генів, тобто частин з яких формуються хромосоми початкової популяції та які використовуються під час операції мутації обираються правила та їх окремі частини з бази знань.

Після проведення формування структури проводиться настроювання параметрів ФН. Для цього, сформовані для визначення кожного терму графу ТА/АБО об'єднуються таким чином, щоб отримати нечітку нейрону мережу (рисунок 6.3).

#### 6.4. Настроювання функцій належності термів лінгвістичних змінних нижнього рівня

Під час етапу настроювання параметрів ФН проводиться настроювання вигляду функцій належності за допомогою алгоритму навчання нейронної мережі, який запропоновано та розглянуто нижче.

Як навчальна вибірка для off-line навчання використовується набір векторів  $L = (p, t)$ .

Тому що ми розглядаємо задачу класифікації стану об'єкту, то  $t$  являє собою не що інше, як клас, до якого повинно бути віднесено стан, обумовлений вектором  $p$ . Причому також відомим є який елемент вхідного вектора  $p$  визначає належність даному класу.

Як процедура дефазіфікації на виході системи для визначення класу, до якого належить стан обумовлений  $p$ , буде використовуватися процедура знаходження максимуму.

Розглянемо алгоритм навчання нечітких нейронних мереж. Цей алгоритм є алгоритмом навчання з учителем. Алгоритм навчання з учителем бази знань нижнього рівня адаптує його нечіткі множини. У процесі навчання через модель циклічно «пробігає» вся навчальна множина  $L$ , продовжуючи нижче перераховані кроки, поки не виконається один із критеріїв завершення.

Кроки:

1) Береться наступний вектор з навчальної вибірки  $L$  та поширюється через мережу.

2) Для кожного вихідного нейрона  $X_{2l}^z$  визначаємо значення  $\delta_{X_{2l}^z}$

$$\delta_{X_{2l}^z} = t_z - \mu_{X_{2l}^z}.$$

3) Для кожного нейрона  $TA_j^z$  з  $\mu_{TA_j^z} > 0$ :

а) Визначаємо значення  $\delta_{TA_j^z}$ , рівне

$$\delta_{TA_j^z} = \delta_{X_{2lj}^z} = \min_r \left( \mu_{X_{1qr}^{k_l z j}} \right) \cdot \left( 1 - \min_r \left( \mu_{X_{1qr}^{k_l z j}} \right) \right) \cdot (t_z - \mu_{X_{2l}^z})$$

б) Знаходимо таке  $X_{1q}^{k_{1q}^*}$ , що

$$\mu_{X_{1q}^{k_{1q}^*}} = \min_r \left( \mu_{X_{1qr}^{k_l z j}} \right)$$

в) Для нечітких множин, що відповідають терму  $X_{1q}^{k_{1q}^*}$  визначаємо  $\delta_a, \delta_b, \delta_c$ , використовуючи швидкість навчання  $\sigma > 0$ :

$$\delta_b = \sigma \cdot \delta_{TA_j^z} \cdot (c - a) \cdot \text{sgn}(x_{q^*} - b)$$

$$\delta_a = -\sigma \cdot \delta_{TA_j^z} \cdot (c - a) + \delta_b$$

$$\delta_c = \sigma \cdot \delta_{TA_j^z} \cdot (c - a) + \delta_b$$

і застосовуємо зміни до функції належності терму  $X_{1q^*}^{k^*}$ , яка визначається за формулою (6.1), якщо це не суперечить накладеним обмеженням.

Якщо епоха закінчена та критерій закінчення досягнуто, то зупинитись, в іншому разі продовжувати навчання з кроку 1.

Слід відмітити, що множник  $\min_r \left( \mu_{X_{1q_r}^{k^*lj}} \right) \cdot \left( 1 - \min_r \left( \mu_{X_{1q_r}^{k^*lj}} \right) \right)$  на кроці 3

а) алгоритму навчання показує, що зміни в нечіткій базі більше в тому випадку, коли вузол правила має активацію, тобто обчислене значення функції належності біля 0,5, та менше у випадку якщо значення біля 0 або 1. Завдяки цьому правило «повинно вирішити» буде воно визначати вхідний вектор або ні.

В якості критерію закінчення обирається відсоток не вірно класифікованих зразків.



## РОЗДІЛ 7. ЗАСОБИ ОБРОБКИ ОБРОБКИ ІНФОРМАЦІЇ У СИСТЕМАХ УПРАВЛІННЯ ЗНАННЯМИ

Для практичного впровадження розроблених методу, підходів та алгоритмів розроблено технологію створення програмних модулів обробки інформації. Ця технологія характеризується гнучкістю та мобільністю, здатністю до комплексної реалізації процесів життєвого циклу ПМОІ, залученню управлінського персоналу до розробки і розвитку ПМОІ в процесі експлуатації готових прикладних програм, адаптації до змін середовища.

### 7.1. Основні положення гнучкої технології створення програмних модулів обробки інформації

Гнучка технологією складається з інструментального комплексу, методичних матеріалів, системи нотацій, методів, критеріїв і правил оцінки результатів. Вона дозволяє зменшити витрати на створення ПМОІ, ефективно автоматизувати розробку, супровід та розвиток ПМОІ з врахуванням індивідуальних особливостей об'єкта керування (ОК) на базі сучасних ІТ.

Для реалізації такої технології була проведена інтеграція моделей, методів і засобів:

- автоматизації проектування програмних засобів (ПЗ) за рахунок опису його специфікацій в термінах об'єктного моделювання, застосування CASE- технологій;

- повторного використання ПЗ в рамках об'єктно-орієнтованого аналізу і проектування на базі репозитаріїв, які специфікуються в плані сервісів об'єктів, що ними надаються і засобів їх зв'язування з прикладними програмами, коли опис об'єкту інтерпретується та він стає доступним для використання іншими компонентами;

- інтеграції конструктивної логіки і прикладної теорії алгоритмів на тих кроках де це можливо, з огляду на властивості СУЗ, (дозволяє постановку проблеми звести до деякої формули визначеної логіко-математичної мови, доповненого засобами представлення визначників, розпізнавачів та інших операторів, які дають можливість рішення проблеми представити у вигляді знаходження або побудови об'єкту з визначеними властивостями або конструювання алгоритму розв'язання поставленої проблеми);

- реалізації рівня спілкування, зручного для людини і ефективного для комп'ютера, наприклад, використовуючи терміни предметної області (ПО), який дозволяє описувати проблеми та досвід управління, інформаційного обслуговування і проектування, а також моделювати навчання в системах знань, з безпосереднім залученням управлінського персоналу;

- теорії БД, яка є основою роботи з даними, знаннями, заготовками системи розробки та іншими складними об'єктами, коли користувач тільки ініціює операції над об'єктами, використовуючи зручні мовні засоби, а

реалізацію операцій, забезпечення цілісності, роботу з пам'яттю, відстеження об'єктів, що знаходяться в оперативній пам'яті, і інші сервісні функції бере на себе відповідний модуль ПМОІ та СКБД;

- організації інтелектуальних пакетів прикладних програм, що використовують для управління інформаційними процесами ПМОІ логіку бізнес-процесів, що описуються спеціальними мовними засобами (забезпечує рішення проблем управління шляхом підключення модулів, які реалізують окремі бізнес-операції і бізнес-процеси), надання можливості нарощування інтелектуальної складової системи під час її експлуатації за рахунок використання CASE- засобів для формування БЗ та її зв'язування з БД.

## 7.2. Інструментальні засоби гнучкої технології

Одними з головних принципів, що були враховані при розробці інструментальних засобів гнучкої технології, те щоб вони дозволяли процеси створення ПМОІ реалізувати як процеси ітеративно-послідовного виконання відображень специфікацій системи в кінцевий продукт, підтримувати як настроювання, так і самонастроювання системи. В такому випадку розробник і користувач можуть створювати систему шляхом проектування без перепрограмування. Тому інструментальні засоби гнучкої технології створення ПМОІ реалізовано за принципом оболонок інтелектуальних систем, виділяючи постійну (управляючу) і змінну (ту що настроюється) частини.

Структура інструментальних засобів гнучкої технології створення ПМОІ наведена на рисунку 7.1.



Рис. 7.1. Структурна схема інструментальних засобів гнучкої технології створення ПМОІ

В ній виділяються три великі частини:

- система підтримки, яка містить блок керування, блок алгебраїчної обробки, модуль нечіткого виведення, генератор запитів;
- систему проектування та розвитку, яка містить інструментальне середовище розробки, блок прив'язки до ПО, блок формування та настроювання БЗ, CASE-засоби обробки інформації в термінах ПО;
- інформаційні бази системи – базу знань, базу метаданих де міститься інформація про ПО та користувачів системи, базу звітів.

Розвиток розглядається як нарощування моделей і методів обробки даних та реалізується користувачем за допомогою інструментальних засобів та зручних мов специфікації системи.

Центральними проблемами при розробці інструментальних засобів є створення механізмів прив'язки до предметної області, формування моделей обробки інформації та звітів за запитом користувача. Проблеми реалізації системи керування даними пов'язані з роботою зі складними об'єктами, що вимагають досить нестандартної логіки обробки. Для вирішення цих проблем необхідно розробляти структуру таким чином, щоб вона з одного боку дозволяла вирішувати задачі обробки інформації, а саме проводити обробку інформації в термінах ПО, а з іншого боку була незалежною від ПО, тобто не залежала не тільки від даних, але і від структури таблиць БД та зв'язків між ними, а також не залежала від СКБД, тобто надавала можливість прив'язки та адаптації до ПО.

Поняття незалежності від даних вводиться в теорії побудови баз даних, де говориться про трирівневу архітектуру ANSI-SPARC, коли представлення користувача не залежить від змін в концептуальній схемі БД, якщо воно не стосується тих даних, що винесені для відображення у представлені.

На відміну від цього, у випадку ПМОІ, під незалежністю розуміється незалежність процедур обробки інформації від всієї структури даних, тобто незалежність від конкретної ПО, яка характеризується своєю БД. Виходячи з цього розроблено відповідний механізму прив'язки, який реалізовано у відповідному блоці інструментальних засобів гнучкої технології створення ПМОІ.

Також слід взяти до уваги необхідність надання можливості нарощування функціональності за допомогою побудови та використання нових моделей, використовуючи для цього терміни предметної області, а також надання можливості динамічного формування звітів за запитом користувача, механізм побудови яких теж треба розробити. Це викликає проблеми отримання відповідних даних з БД.

Розробка ПМОІ для конкретного об'єкту починається з заготівки (керуючої частини), і система поступово «вирощується» – формують початкову змінну частину і безперервно її розвивають. Досвід дозволить визначити та ефективно використати в процесі розробки множини базових заготовок обох частин для різних СУЗ.

Реалізація ПМОІ для конкретної ПО проходить через наступні етапи (рисунок 7.2)



Рис. 7.2. Етапи реалізації ПМОІ для конкретної ПО

Перед початком роботи слід налаштувати систему до ПО, що робиться за допомогою блоку прив'язки. Слід відзначити, що прив'язку необхідно робити лише один раз, за умови, що в подальшому структура таблиць та взаємозв'язки між ними не будуть змінюватися. В протилежному випадку слід скоригувати введену інформацію про ПО.

В процесі прив'язки адміністратор за допомогою програмного інтерфейсу, або вручну заповнює відповідну системну таблицю, а також вносить інформацію про користувачів системи. Ця інформація вноситься та зберігається в базі метаданих (БМД).

Під час наступного етапу комплекс інструментальних засобів автоматично проводить перевірку повноти прив'язки, тобто такої прив'язки яка дозволить вірно отримувати дані з БД.

Наступний етап полягає в формуванні БЗ. Під час нього експерти вносять інформацію про лінгвістичні змінні та правила що їх пов'язують до БЗ, виходячи з яких будуть створюватись відповідні моделі та проводитись аналіз вхідної інформації. Процес опитування експертів та формування початкової бази знань не розглядається в даній роботі та вважається, що база знань, яка сформована експертами є повною та несуперечливою.

На етапі модифікації і настроювання автоматично відбувається процес зведення БНЗ до деревоподібної структури, який полягає в застосуванні розглянутого раніше підходу до зведення до деревоподібної структури. Після цього проводиться настроювання нижньої частини нечіткої бази знань за допомогою підходу до формування та настроювання.

### 7.3. Прив'язка до предметної області

Як було сказано у попередньому розділі прив'язка ПМОІ до БД повинна здійснюватись для реалізації обробки інформації в термінах предметної області. Для проведення ж такої обробки в першу чергу слід

вибрати дані з БД.

Оскільки структура БД може бути довільною, то це ускладнює процес вибору даних, а саме формування відповідного запиту. Мова формування звітів є мовою специфікаційного типу відноситься до мов QBE.

**Мовні засоби гнучкої технології побудови програмних модулів обробки інформації.** Серед мовних засобів гнучкої технології як основну можна виділити мови побудови звітів операторного та специфікаційного типів.

Перша – це мова специфікаційного типу. Її можна віднести до класу мов QBE, з додатковими функціями, а саме визначення користувачем структури звіту в термінах предметної області й автоматичним визначенням зв'язків між таблицями, що буде гарантувати семантичну коректність побудованого запиту до БД. Друга – це мова операторного типу, яка дозволяє використовувати SQL та будувати складніші звіти, однак вона призначена для користувачів-програмістів.

Мова формування звітів специфікаційного типу орієнтована на користувача, що не програмує, та надає йому зручні засоби визначення змісту бажаного звіту в термінах предметної області. Ця мова сполучає зручний графічний інтерфейс для визначення умов на дані, що будуть міститись в звіті, структури звіту, типу відображення звітної інформації. Умова на дані в звіті формуються з термінів предметної області, констант, зв'язок логічних операцій (NOT, OR, AND) і операцій порівняння (=, ≠, ≥, >, ≤, <). Інструментарій обробки запитів побудованих цією мовою надає змогу отримання послідовності таблиць, через які проходить зв'язок, отримуючи на вході лише терміни предметної області.

Для опису граматики, що відповідає мові побудови звітів скористуємось формою Бекуса-Наура:

```
<Звіт>→<Назва_звіту><Тип_звіту><Зріз><Реквізити_звіту><Умови_на_реквізити >  
<Зріз>→<Вертикальний_зріз><Вертикальний_Реквізит_зрізу>|  
<Горизонтальний_зріз><Горизонтальний_Реквізит_зрізу>|  
<Горизонтальний_Вертикальний_зріз><Горизонтальний_Реквізит_зрізу><Вертикальний_Реквізит_зрізу>  
<Тип_звіту>→<Таблиця>|<Графіки>|<Таблиця><Графіки>  
< Реквізити_звіту>→<Реквізит>| < Реквізити_звіту><Реквізит>|  
<Графіки>|<Таблиця><Графіки>  
<Горизонтальний_Реквізит_зрізу>,<Вертикальний_Реквізит_зрізу>,  
<Реквізит_зрізу>→<Реквізит>  
<Реквізит>→<Термін_ПО>  
<Операція>→ = | ≠ | ≥ | > | ≤ | <  
<Логічна_зв'язка>→ NOT | OR | AND  
<Умови_на_реквізити>→<Умови_на_реквізити><Логічна_зв'язка><Умова_на_реквізит>  
<Умова_на_реквізит>→<Реквізит><Операція><Константа>|<Реквізит>
```

### <Операція><Реквізит>

Користувач формує запит на звіт в діалоговому режимі, на протязі сеансу роботи користувачу повідомляється вся необхідна інформація про хід побудови звіту і видається досить докладна діагностика при помилковому запиті або неможливості його реалізації. Формування умов не потребує спеціальної математичної підготовки, передбачена докладна діагностика. Схема отримання результуючого звіту по інформаційному запиту користувача сформульованому в графічному редакторі наведена на рисунку 7.3

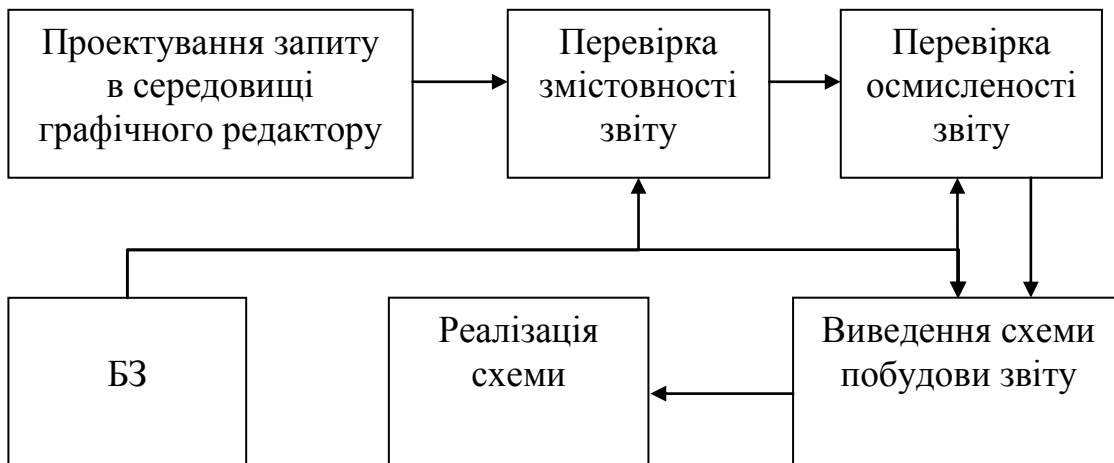


Рис. 7.3. Схема обробки запитів в мові формування звітів

Під час її інтерпретації додатково визначаються зв'язки між таблицями. В роботі обмежимося лише реляційними БД, в яких для маніпулювання даними використовується мова структурованих запитів SQL.

В мові SQL запит на вибірку визначається оператором SELECT. Тоді можна сказати, що проблема полягає в автоматичній побудові SQL-запитів SELECT, при умові що користувач задає лише терміни предметної області та відповідні обмеження на них. Але треба бути впевненим, що користувач вірно сформулював запит. Для цього необхідно щоб семантичний аналіз показав, що запит сформульований вірно.

Метою проведення семантичного аналізу є відхилення запитів, які некоректно сформульовані або містять суперечливі вимоги. Запит вважається некоректно сформульованим, якщо його виконання ніколи не зможе привести до створення результуючого набору записів. Частіше за все така ситуація виникає тоді, коли в предикаті відсутнє визначення необхідних операцій об'єднання.

Рішення цієї проблеми полягає в розробці підходу до формування такої службової інформації, та способу побудови запиту на основі цієї інформації, використовуючи які ПМОІ зможе автоматично будувати запити на вибірку.

Для надання можливості використання термінів предметної області при проведенні обробки даних необхідно зробити прив'язку ПМОІ до бази даних.

Для формування прив'язки використовується відповідна мова

прив'язки, яка призначена для прив'язки інструментарію до відповідної предметної області, тобто введення термінів, що відповідають предметній області та їх прив'язки до структури БД. Вона орієнтована на кваліфікованого користувача, а інструментарій її обробки надає можливість перевірки повноти прив'язки, тобто визначення факту зв'язності всіх таблиць в базі даних і існування лише одного шляху між двома таблицями, що надасть змогу побудови несуперечливого запиту.

Для опису граматики, що відповідає мові прив'язки скористаємось формою Бекуса-Наура:

```
<Прив'язка>→<Терміни_прив'язка><Зв'язки_таблиць>
<Терміни_прив'язка>→<Термін_ПО><Назва_табл><Назва_поля>
|<Терміни_прив'язка>; <Термін_ПО><Назва_табл><Назва_поля>
<Зв'язки>→<Назва_табл_1> <Назва_поля_1>=
=<Назва_табл_2> <Назва_поля_2>|<Зв'язки>; <Назва_табл_1><Назва_
_поля_1>=<Назва_табл_2> <Назва_поля_2>
```

При застосуванні мови прив'язки заповнюється відповідна таблиця, структура якої наведена нижче.

Наприклад, це можна зробити як показано в таблиці прив'язки термінів ПО до БД (ТПр).

*Таблиця 7.1*

**Таблиця прив'язки термінів ПО до бази даних**

термін предметної області	таблиця, в якій зберігаються дані	назва поля в таблиці, яке відповідає критерію

Але у випадку використання такої прив'язки не можна гарантувати вірність автоматизованої побудови SQL-запиту, оскільки в процесі його побудови необхідно знати зв'язки між таблицями, та правильно їх записати в запиті.

Для розуміння того, яку ще інформацію слід додати до таблиці 7.1 необхідно розглянути структуру запиту, та визначити ті параметри, які необхідні для його вірної побудови.

Загальний формат SQL-запиту на вибірку:

```
SELECT [column_ expression [AS new_name]] [...]
FROM table_name [alias] [...]
[WHERE condition]
[GROUP BY column_list] [HAVING condition_group]
[ORDER BY column_list]
```

де квадратні скобки визначають необов'язковий елемент;

багатокрапка (...) використовується для вказівки необов'язкової можливості повторення конструкції від нуля до декількох разів;

параметр *column\_expression* представляє собою ім'я стовпчика або вираз з декількох імен. Параметр *table\_name* це назва існуючих в базі даних таблиці або представлення, до яких необхідно отримати доступ.

Обробка елементів оператора SELECT виконується в наступній послідовності:

FROM – визначає імена таблиці або декількох таблиць, що використовуються у запиті;

WHERE – проводиться фільтрація рядків об'єкту у відповідності з заданими умовами та визначається зв'язок між таблицями;

GROUP BY – визначаються групи рядків, що мають однакові значення у вказаному стовпці;

HAVING – фільтруються групи рядків у відповідності з умовами;

SELECT – визначає, які стовпці повинні бути присутні у результируючих даних;

ORDER BY – визначається упорядкованість результатів виконання оператора.

Залишивши тільки ту частину, яка буде використовуватися в подальшому викладенні та дещо спростивши її, отримаємо SQL- запит наступного вигляду (7.1):

```
SELECT column_expression  
FROM table_name (7.1)  
[WHERE condition [AND column_condition]]
```

При побудові багатотабличного запиту необхідно в операторі FROM вказати всі таблиці, що використовуються в запиті, а також проміжні таблиці, через які проходить зв'язок між таблицями, що не пов'язані напряму, а в операторі WHERE необхідно вказати як саме зв'язані таблиці *condition*, а також умови на значення полів *column\_condition*, які входять в *column\_expression*. При виникненні помилок в побудові запиту можуть бути отримані помилкові дані. Наприклад, у випадку необхідності виведення даних з полів двох незв'язаних таблиць буде отримано декартовий добуток всіх значень поля з однієї таблиці на всі значення поля другої таблиці. Для того щоб обійти цю проблему і пропонується наведений нижче підхід.

Для його реалізації треба ще додатково до таблиці 7.1 ввести таблицю зв'язків (ТЗв), в якій перший стовпчик та перший рядок це назви таблиць, які є в базі даних та будуть використовуватися в запитах, а на перетині стовпчика з рядком, у випадку коли є прямий зв'язок між таблицями, записуються поля таблиці, що знаходиться у верхньому рядку, та які пов'язані з відповідними полями таблиці, яка знаходиться в першому стовпчику. Поля таблиць організують у вигляді масивів. У випадку коли зв'язку немає, поле залишається порожнім. На діагоналі знаходяться порожні поля.



Отже, під прив'язкою ПМОІ до БД буде розумітись заповнення ТПр (таблиця 7.1) та ТЗв. Повною прив'язкою назвемо таке заповнення таблиці ТПр та ТЗв, виходячи з якого можна побудувати вірний SQL-запит на вибірку.

Множиною імен таблиць буде називатись впорядкована множина потужністю  $N$ , елементами якої є назви таблиць в базі даних

$$TN = \{TableName_1, TableName_2, \dots, TableName_N\}.$$

Будемо позначати  $TZv_{ij}$  елемент таблиці зв'язків, який знаходиться на перетині рядка, що відповідає таблиці  $TableName_i$ , та стовпчику, що відповідає таблиці  $TableName_j$ .

Під відображенням зв'язності буде розумітись  $C: TN \times TN \rightarrow \{0,1\}$ , де  $C(TableName_i, TableName_j) = 1$  у випадку коли таблиця  $i$  зв'язана з таблицею  $j$ , та  $C(TableName_i, TableName_j) = 0$  у протилежному випадку.

Матрицею зв'язності  $\hat{C}$  буде називатись  $N \times N$  розмірна матриця, яка відповідає відображенню  $C$  та кожен елемент якої вказує на те чи існує прямий зв'язок між таблицями. Ця матриця будується на основі таблиці зв'язків, заміною непустиого поля на одиницю, а пустиого на нуль.

Формування множини імен таблиць і матриці зв'язності в ПМОІ може проводитись автоматично.

Головною вимогою до прив'язки є її повнота, тобто прив'язка повинна гарантувати, що буде отримано вірний запит.

Схему даних БД можна представити у вигляді зваженого графу  $G = (V, E)$ , де замість таблиць слід поставити вершини, а замість зв'язків між таблицями – ребра з одиничною вагою. Множина вершин  $V$  графу  $G$  буде дорівнювати множині назв таблиць  $TN$ , а матриця суміжності  $A$  графу  $G$  буде відповідати матриці зв'язків  $\hat{C}$  (7.2).

$$V = TN, A = \hat{C} \quad (7.2)$$

Тоді пошук зв'язку між таблицями БД можна інтерпретувати як знаходження мінімального підграфа  $H$  графу  $G$ , який буде містити всі вершини, що відповідають таблицям, які входять до запиту. Така задача носить назву задачі про мінімальне зв'язування.

Для отримання вірного та однозначного запиту необхідно існування однозначного зв'язку між таблицями. Адже, як було описано вище результатом виконання SQL-запиту, який побудовано по таблицях, що не зв'язані напряму та між якими не вказано зв'язок через інші таблиці, буде декартовий добуток значень одного поля на значення другого. До того ж, якщо зв'язок є неоднозначним, тобто може реалізовуватись через різні таблиці, то з теорії побудови реляційних баз даних та реляційної алгебри буде слідувати можливість побудови різних запитів, використовуючи які будуть отримані різні результуючі дані з БД. Якщо розглядати відповідний граф  $G$ , то це буде

означати існування всього одного ланцюгу між кожними двома вершинами та, як наслідок, єдиничність підграфу  $H$ .

Позначимо деякий ланцюг у графі  $G$ , врахувавши те що граф  $G$  є деревом, та ланцюг однозначно задається початковою і кінцевою вершинами через

$$P_{k_1 k_h} = (v_{k_1}, v_{k_2}, \dots, v_{k_h}) \quad (7.3)$$

З теорії графів відомо, що необхідною та достатньою умовою для існування між кожними двома вершинами тільки одного ланцюгу є зв'язність відповідного графу, а також відсутність циклів, тобто відповідний граф повинен бути деревом.

Виходячи з вищесказаного, слід відмітити, що для доведення повноти прив'язки слід визначити, чи є граф  $G$ , який відповідає структурі БД та визначається матрицею суміжності  $A = \hat{C}$ , деревом. Перевірка цього факту проводиться описаним нижче методом.

Метод перевірки того що прив'язка є повною можна розділити на три частини.

По-перше, виходячи з таблиці 7.1 та ТЗв, слід побудувати матрицю зв'язності  $\hat{C}$ , метод побудови якої описано вище, та визначити граф  $G$ , який буде відповідати структурі даних БД ПО. Як було показано раніше, матриця суміжності  $A$  буде дорівнювати матриці зв'язності  $A = \hat{C}$ .

По-друге, перевірити чи є граф  $G$  зв'язним. Перевірка зв'язності графу  $G$  проводиться виходячи з теореми, яка говорить що граф  $G$ , який визначається матрицею суміжності  $A$ , зв'язний тоді і тільки тоді, коли  $\hat{A}_{ij}^I = 1$  для всіх його вершин  $i$  і  $j$ . Де  $\hat{A}^I = I \vee \hat{A}$ ,  $I$  – мультиплікативна одинична матриця.

$$\hat{A} = A \vee A^{\otimes 2} \vee \dots \vee A^{\otimes n}, \quad A^{\otimes k} = A^{\otimes k-1} \otimes A, \quad A^{\otimes 1} = A, \quad k = \overline{1, n}.$$

$$B = C \vee D \text{ означає } B_{ij} = C_{ij} \vee D_{ij}$$

де  $B, C, D$ - матриці розмірністю  $m \times s$ .

$$F = C \otimes E \text{ означає}$$

$$F_{ij} = (C_{i1} \wedge E_{1j}) \vee (C_{i2} \wedge E_{2j}) \vee \dots \vee (C_{is} \wedge E_{sj}), \quad i = \overline{1, m}, \quad j = \overline{1, p}$$

де  $F, C, E$ -матриці, які мають розмірність  $m \times p$ ,  $m \times s$ ,  $s \times p$

відповідно.

Тоді, перевірка зв'язності буде здійснюватись через наступні кроки.

1. Знайти матрицю  $\hat{A}$ , використавши алгоритм Уоршолла.
2. Визначити матрицю  $\hat{A}^I = I \vee \hat{A}$ .
3. Перевірити необхідну та достатню умову зв'язності

$$\hat{A}_{ij}^I = 1, \quad i = \overline{1, n}, \quad j = \overline{1, n}.$$

По-третє, у випадку зв'язності графу перевірити чи є граф  $G$  деревом. Для цього слід скористатися теоремою, яка говорить про те, що зв'язний граф є деревом у випадку коли кількість його вершин більше на одиницю ніж кількість його ребер.

$$|V| = |E| + 1 \quad (7.4)$$

Та якщо граф  $G$  є деревом, то робиться висновок про те що прив'язка є повною.

В графі  $G$  напряму не задані ребра, зв'язки між вершинами графу визначаються через матрицю зв'язності  $A$ . Для того щоб скористатися згаданою вище теоремою слід визначити кількість ребер. Вона буде дорівнювати сумі елементів матриці  $A$ , які знаходяться вище головної діагоналі:

$$|E| = \sum_{i=2}^{N-1} \sum_{j=i+1}^N A_{ij} \quad (7.5)$$

#### 7.4. Автоматична побудова запиту до бази даних

Як було відмічено вище, основною проблемою при побудові SQL-запиту є знаходження зв'язку між таблицями бази даних, що не зв'язані між собою напряму, та, виходячи із знайденого зв'язку, формування аргументів операторів FROM та WHERE.

Постановка цієї задачі може бути записана наступним чином. Базуючись на ТПр, ТЗв,  $\hat{C}$  та  $TN = (TableName1, TableName2, \dots, TableNameN)$ , та виходячи з вхідних даних  $\hat{D}_{I_F} = \{(T_{ik_l}, cond_{ik_l}) \mid l = \overline{1, n_F}\}$ , якими є обрані користувачем терміни предметної області  $T_{ik_l}$  та задані обмеження на них  $cond_{ik_l}$ , побудувати SQL-запит на вибірку.

Метод рішення задачі динамічної побудови запиту полягає в послідовному розв'язку двох більш дрібних задач.

По-перше, визначення полів таблиць, що відповідають термінам предметної області та формування *column\_expression* і завдання умов на значення полів *column\_condition* в (7.1). По-друге, визначення зв'язку між таблицями та формування *table\_name* і *condition* в (7.1). Далі розглянуто більш докладно постановки та методи розв'язання цих задач.

Постановка першої задачі полягає в тому, що виходячи з ТПр та  $\hat{D}_{I_F} = \{(T_{ik_l}, cond_{ik_l}) \mid l = \overline{1, n_F}\}$  необхідно визначити *column\_expression* і *column\_condition* в (7.1).

Розв'язок цієї задачі полягає в співставленні кожному терміну ПО  $T_{ik_l}$  з  $\hat{D}_{I_F}$  відповідної назви поля  $ColumnName_{t_l}$  та таблиці  $TableName_{t_l}$  з БД, яке знаходиться в ТПр. Після цього *column\_expression* формується простим переліком через кому назв відповідних полів, а *column\_condition* переліком умов на значення визначених раніше полів, які записуються визначеним в мові SQL способом (7.6).

$$\begin{aligned}
& \text{TableName}_{t_1}.\text{ColumnName}_{t_1r_1}, \\
\text{column\_expression} &= \text{TableName}_{t_2}.\text{ColumnName}_{t_2r_2}, \dots, \\
& \text{TableName}_{t_{n_F}}.\text{ColumnName}_{t_{n_F}r_{n_F}} \\
& \text{cond}_{ik_1}(\text{TableName}_{t_1}.\text{ColumnName}_{t_1r_1}) \text{ AND} \\
\text{column\_condition} &= \text{cond}_{ik_2}(\text{TableName}_{t_2}.\text{ColumnName}_{t_2r_2}) \text{ AND} \dots \\
& \text{AND cond}_{ik_{n_F}}(\text{TableName}_{t_{n_F}}.\text{ColumnName}_{t_{n_F}r_{n_F}})
\end{aligned} \tag{7.6}$$

Наступним кроком формується множина назв таблиць  $Table\_N = \{ \text{TableName}_{t_l} \mid l = \overline{1, n_F} \}$ , поля яких відповідають обраним користувачем термінам ПО.  $Table\_N$  буде використовуватись при розв'язку наступної задачі.

Постановку другої задачі можна записати наступним чином. Виходячи з  $Table\_N$ ,  $\hat{C}$  та  $TN$  визначити зв'язок між таблицями та сформувати  $table\_name$  і  $condition$  в (7.1).

Як відмічалось раніше, пошук зв'язку між таблицями БД можна інтерпретувати як розв'язок задачі про мінімальне зв'язування в теорії графів. В цьому випадку постановку другої задачі можна переписати наступним чином: для графу  $G = (V, E)$  необхідно визначити частковий підграф  $H = (V', E')$ , який буде містити вказані в деякій множині  $Q = Table\_N$  вершини з найменшою сумарною вагою ребер. Та, використовуючи підграф  $H$ ,  $TN$  та ТЗв, визначити  $table\_name$  і  $condition$  в (7.1). Слід відмітити, що  $|Q| > 1$ .

В загальному випадку першу частину цієї задачі пропонується розв'язувати шляхом вирішення задачі про найкоротший ланцюг для вершин множини  $Q$ , вказуючи як джерело по черзі кожну вершину. Тоді для кожної верши  $v_v \in V$  будуть обраховані  $|Q|$  чисел – відстаней від  $v_Q \in Q$ . Просумувавши їх для всіх вершин визначається вершина з найменшою сумою. Через неї вершини з  $Q$  з'єднуються найкоротшими шляхами, що й дає розв'язок поставленої задачі.

Однак, оскільки граф  $G$  є деревом можна запропонувати простіший підхід для визначення підграфу  $H$ . Ідея такого підходу полягає в обранні довільної вершини  $v_i \in Q$  в якості кореня, та введення відображення нумерації  $pred$  вершин графу  $G$ , таким чином, що кожній не кореневій вершині ставиться у відповідність інформація про номер вершини, яка знаходиться вище та з якою вона пов'язана (7.7). Таку вершину будемо називати батьківською. Оскільки  $G$  дерево, то задане таким чином відображення  $pred$  однозначне.

$$\begin{aligned}
& pred : V \setminus v_{i_1} \rightarrow [1; N] \\
& \forall v_j \in V \setminus v_{i_1}, \exists! pred(v_j) = i : (v_i, v_j) \in E
\end{aligned}
\tag{7.7}$$

Процес перенумерації, тобто визначення відображення  $pred$ , проводиться по вершинах графу  $G$  до тих пір доки не перенумеровані всі вершини з множини  $Q$ . Тобто множина вершин, для яких задана нумерація  $V_{num}$ , та на яких буде визначено відображення  $pred$  в загальному випадку є підмножиною множини вершин  $V$  графу  $G$  (7.8). Множину номерів вершин, які складають множину припустимих значень відображення  $pred$  позначимо через  $U$ .

$$\begin{aligned}
& V_{num} \subseteq V \\
& pred : V_{num} \rightarrow U \\
& \forall v_j \in V_{num}, \exists! pred(v_j) = i : (v_i, v_j) \in E
\end{aligned}
\tag{7.8}$$

Далі вводиться відображення розмітки вершин  $mark$ , що визначено на множині вершин  $V_{num} \subseteq V$  та дає одиницю, якщо вершина  $v_s \in V_{num}$  є частиною ланцюгу  $P_{i_k i_1}$ , який з'єднує вершину  $v_{i_k} \in Q$  та кореневу вершину  $v_{i_1}$ , та нуль в протилежному випадку (7.9). У випадку коли  $mark(v_s) = 1$  будемо казати, що вершина  $v_s \in V_{num}$  позначена, в протилежному випадку – не позначена.

$$\begin{aligned}
& mark : V_{num} \rightarrow \{0;1\} \\
& \begin{cases} mark(v_s) = 1, v_s \in P_{i_k i_1} \\ mark(v_s) = 0, v_s \notin P_{i_k i_1} \end{cases}
\end{aligned}
\tag{7.9}$$

Процес визначення відображення розмітки проводиться по всіх вершинах  $v_{i_k} \in Q$ , які ще не були позначені та проводиться вгору до першої на шляху позначеної вершини. Всі інші вершини вважаються не позначеними.

Тоді граф, який складається з всіх позначених вершин графу  $G$  і буде шуканим підграфом  $H$ . Описаний спосіб реалізується двома послідовно виконуваними алгоритмами: алгоритмом нумерації графу  $G$  та алгоритмом розмітки графу  $G$  і знаходження  $H$ . Після цього, виходячи з  $H$ ,  $TN$  та  $TЗв$  можна сформуванати  $table\_name$  і  $condition$ . Алгоритм формування  $table\_name$  і  $condition$  буде містити кроки, які аналогічні тим, що використовуються в алгоритмі розмітки, тому доречним буде об'єднати ці алгоритми в один. Цей алгоритм буде мати назву алгоритм розмітки та формування  $table\_name$  і  $condition$ .

Додатково до введених позначень в цих алгоритмах будуть використовуватись наступні. Матриця  $W$ , яка перед початком роботи

алгоритму дорівнює матриці суміжності  $A$  графу  $G$ ,  $W(i)$  буде означати  $i$ -ту строку матриці  $W$ , а  $w_{ij}$  – елемент матриці  $W$ , який знаходиться в  $i$ -му рядку та  $j$ -му стовпчику. Множина  $Q_1$ , яка буде складатись з тих вершин множини  $Q$ , які на даному кроці проведення перенумерації та розмітки не були помічені. Змінні  $temp$  та  $temp1$ , які будуть використовуватись для тимчасового зберігання номеру вершини. Множина вершин  $V_{num}^k$  з області визначення відображення  $pred$ , які були визначені на  $k$ -й ітерації, а множина припустимих значень –  $U^k$ . Тоді область визначення та множина припустимих значень відображення  $pred$  визначається через

$$\begin{aligned} V_{num} &= \bigcup_k V_{num}^k \\ U &= \bigcup_k U^k \end{aligned} \quad (7.10)$$

*Алгоритм нумерації*

1. Встановити  $Q_1 = Q$ ,  $W = A$ .
2. Обрати кореневу вершину  $v_{i_1} \in Q$ . З множини  $Q_1$  виключити вершину  $v_{i_1}$ . Визначити  $V_{num}^1 = \{v_{i_1}\}$ ,  $k = 2$ .

3. Для кожної вершини  $v_i \in V_{num}^{k-1}$  визначити

$$\begin{aligned} (V_{num}^k)_i &= \{v_j \mid w_{ij} = 1, 1 \leq j \leq N\} \\ (U^k)_i &= \{i \mid \forall v_j \in (V_{num}^k)_i, pred(v_j) = i\} \\ pred : (V_{num}^k)_i &\rightarrow (U^k)_i \end{aligned} \quad (7.11)$$

Зрозуміло, що множина  $(U^k)_i$  буде однокрапковою та складатися лише з номера вершини  $i$ , тому вираз  $pred : (V_{num}^k)_i \rightarrow (U^k)_i$  задає відображення однозначно.

4. Обнулити  $i$ -тий рядок матриці  $W$

$$W(i) = 0.$$

5. Визначити множини  $V_{num}^k$  та  $U^k$

$$\begin{aligned} V_{num}^k &= \bigcup_i (V_{num}^k)_i \\ U^k &= \bigcup_i (U^k)_i \end{aligned} \quad (7.12)$$

6. Виключити з множини  $Q_1$  вершини, що попали до множини  $V_{num}^k$  та належать множині  $Q$

$$Q_1 = Q_1 \setminus V_{num}^k \quad (7.13)$$

7. Якщо  $Q_1 = \emptyset$ , то п. 8, в протилежному випадку п. 9.

8. Визначити множини  $V_{num}$  та  $U$  (7.10) та кінець алгоритму.

9. Призначити  $k = k + 1$  та перехід до п. 3.  
 Блок-схема алгоритму нумерації наведена на рис. 7.4.

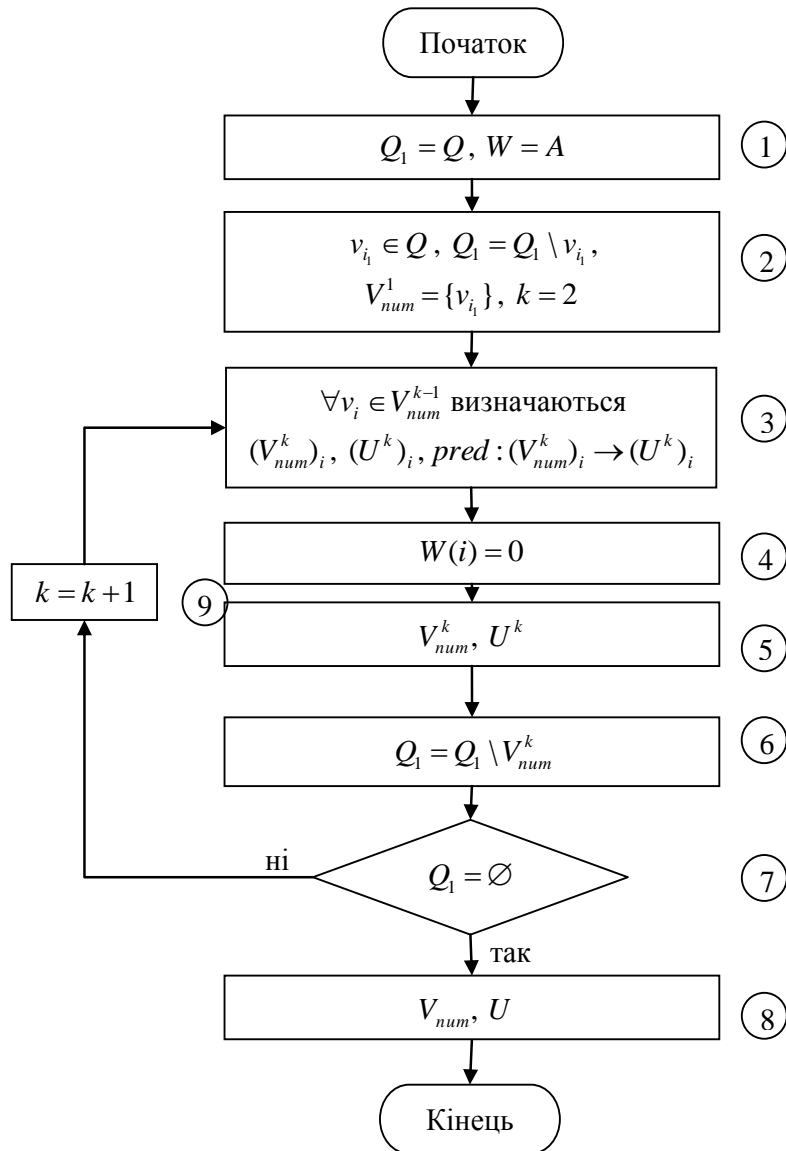


Рис.7.4. Блок-схема алгоритму нумерації

Перед початком викладення алгоритму розмітки графу  $G$  і знаходження підграфу  $H$ , а також формування *table\_name i condition* слід відзначити, що оскільки граф  $G$  відповідає всій схемі даних БД, то підграф  $H$  відповідає частині схеми даних БД. Це означає, що оскільки  $TN = V$ , а  $V' \subseteq V$ , то  $V' \subseteq TN$ . Позначимо  $TN' = V'$  множина імен таблиць БД, через які необхідно побудувати зв'язок в SQL-запиті, для його вірної реалізації. Множина ребер  $E'$  буде відповідати множині зв'язків між таблицями з  $TN'$ .

*Алгоритм розмітки та формування table\_name i condition*

1. Всі вершини з множини  $V_{num}$  призначити як не позначені та призначити пусте значення *condition*

$$mark(v_i) = 0, \forall v_i \in V_{num}, condition = " \quad (7.14)$$

2. Позначити кореневу вершину  $v_{i_1} \in Q$ , сформувати множину позначених вершин  $V_{mark} \subseteq V_{num}$  та визначити  $Q_1$

$$\begin{aligned} mark(v_{i_1}) &= 1 \\ V_{mark} &= \{v_{i_1}\} \\ Q_1 &= Q \setminus v_{i_1} \end{aligned} \quad (7.15)$$

3. Обрати вершину  $v_{i_k} \in Q_1$ , та призначити  $temp = i_k$ .

4. Позначити вершину  $v_{temp}$ , визначити  $V_{mark}$  та  $Q_1$ .

$$\begin{aligned} mark(v_{temp}) &= 1 \\ V_{mark} &= V_{mark} \cup \{v_{temp}\} \\ Q_1 &= Q_1 \setminus v_{temp} \end{aligned} \quad (7.16)$$

5. Визначити батьківську вершину для  $v_{temp}$  та визначити  $temp1$

$$temp1 = pred(v_{temp}) \quad (7.17)$$

6. Якщо  $condition = ''$ , то п. 8, інакше п.7.

7. Визначити  $condition$

$$condition = condition + ' ' + 'AND' + ' ' \quad (7.18)$$

8. Визначити  $condition$

$$\begin{aligned} condition &= condition + \\ &+ '(TableName_{temp}.TЗв_{temp,temp1} = TableName_{temp1}.TЗв_{temp1,temp})' \end{aligned} \quad (7.19)$$

(У випадку, коли таблиці зв'язані по декільком полям, пункти 7 і 8 повторюються для всіх відповідних полів, тобто для всіх елементів масиву, який знаходиться в полі ТЗВ).

9. Призначити  $temp1 = temp$ .

10. Якщо  $mark(v_{temp}) = 1$ , то п. 11, в протилежному випадку п. 4.

11. Якщо  $Q_1 \neq \emptyset$ , то п. 3, в протилежному випадку п. 12.

12. Визначити граф  $H(V', E')$

$$\begin{aligned} V' &= V_{mark} \\ E' &= \{(v_i, v_j) \mid v_i, v_j \in V_{mark}, (v_i, v_j) \in E\} \end{aligned} \quad (7.20)$$

13. Сформувати  $table\_name$ , який буде представляти з себе простий перелік через кому назв таблиць з  $TN' = V'$  і кінець алгоритму.

Блок-схема алгоритму розмітки графу формування  $table\_name$  і  $condition$  наведена на рисунку 7.5.

Використавши отримані в (7.6)  $column\_expression$  і  $column\_condition$ , а також сформовані в результаті виконання алгоритму  $table\_name$  і  $condition$  можемо записати SQL-запит у вигляді (7.1).



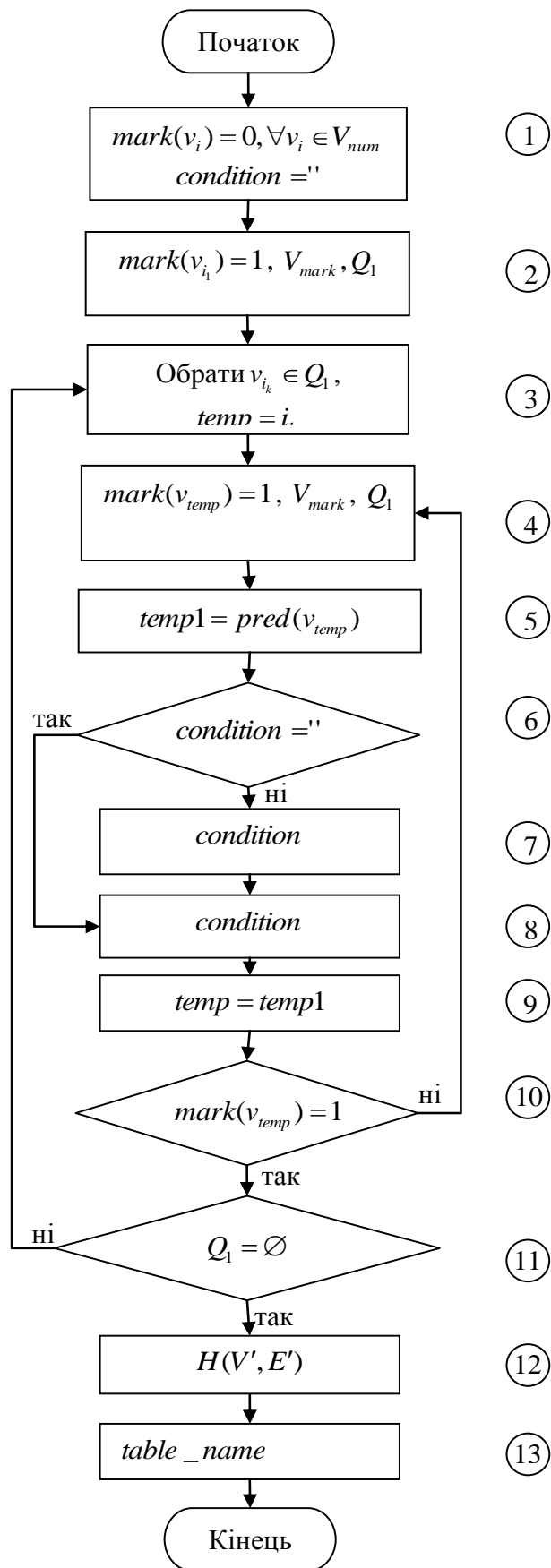


Рис. 7.5. Блок-схема алгоритму розмітки графу і знаходження

## **РОЗДІЛ 8. МЕТОДИКА ПРОВЕДЕННЯ КОНТЕНТ-АНАЛІЗУ ЛЕКСИКИ ХУДОЖНЬОГО ТВОРУ НА ОСНОВІ ВИКОРИСТАННЯ ФУНКЦІЙ СИСТЕМИ КОНСПЕКТ (ЗА ТЕКСТАМИ ЗБІРКИ Т. Г. ШЕВЧЕНКА «КОБЗАР»)**

*«Огненне слово його наскрізь проймало серце  
не тільки тих, кому близьке було народне горе,  
а й тих, кому й байдуже було до того.  
Всі дивувалися красі та силі тієї простої мови,  
якою Шевченко виливав свої вірші.  
Увесь світ став прислухатися до його мови,  
а на Україні вірші приймали  
як благовісне, пророче слово» (Панас Мирний).*

### **8.1. Програма обробки лінгвістичних даних “Конспект”**

Серед різних варіантів практичного використання сховищ текстової інформації превалює потреба виділення інформації, що характеризується тематичною єдністю. Тематичний пошук вимагає копіткої роботи з текстами, що зберігаються в бібліотеках, архівах, Інтернеті, текстових базах даних. Трудність цієї роботи полягає, зокрема, в тому, що частіше доводиться відбирати не цілі тексти, а фрагменти текстів які відносяться до теми. Зміст багатьох текстів є переплетенням ряду тем. Виникає проблема пошуку усередині текстових документів фрагментів, які релевантні заданій темі.

Існуючі методи інформаційного пошуку, зокрема в мережі Інтернет, видають багато непотрібної “смітцевої” інформації, фільтрація якої займає дуже багато часу. Вихід з положення полягає у використанні семантичних критеріїв, що забезпечують відбір найбільш істотних характеристик понять, щодо яких збирається інформація.

Функції обробки лінгвістичних даних системи “Конспект” забезпечують наступне:

- лінгвістичний аналіз тексту до рівня поверхневого синтактико-семантичного аналізу;
- виділення термінів предметної області з релевантних текстів;
- виділення і стисле конспектування фрагментів природно-мовних текстів, що відносяться до заданої теми, яка задається ключовим словом або словосполученням.
- генерація за наслідками семантичного аналізу заданого числа вторинних ключів, використання яких в циклічному режимі дозволяє поглибити розкриття теми у сформованих конспектах.
- використання стислих тематичних конспектів для вибору з множини текстових документів таких, які найбільшою мірою релевантні заданій темі.

В якості початкового етапу лінгвістичного аналізу виконується *графемно-морфологічний* аналіз.

Призначенням *графемного аналізу* є побудова моделі графемної структури вхідного тексту, в якій виділені і зв'язані відношеннями (де це можливо) такі змістовні одиниці тексту, як фрагмент, речення та лексема. Наприклад, для лексем: виділяються класи лексем, які відрізняються своєю графемною структурою і виконують різні функції в тексті. Крім того, аналізуються закономірності сполучуваності деяких лексичних одиниць, які уже на цьому етапі дозволяють об'єднувати декілька лексем в одну на тій основі, що вони виконують в тексті єдину функцію.

В результаті графемного аналізу текст перетворюється у впорядкований список складників ПМТ.

Під *морфологічним аналізом* розуміють повну обробку послідовності (без будь-якого зв'язку з контекстом) словоформ, сформованих на етапі графемного аналізу. В результаті такої обробки кожна словоформа доповнюється граматичною інформацією – ланцюжком символів, що визначають усі властивості словоформи, які необхідні для подальшого аналізу. Морфологічна інформація сукупності словоформ, в свою чергу, використовується на наступному етапі створення програмної моделі ЛП – під час синтаксичного аналізу, в результаті якого встановлюються зв'язки між усіма словоформами тексту та між частинами складних речень.

Вхідними даними процедури морфологічного аналізу є графемна структура тексту, база знань морфології ПМ та лексикографічна база даних, яка включає словники лексем та словозмінну і словотвірну моделі вхідної мови. Вона включає окремі таблиці для всіх частин мови. До кожної лексеми в таблиці приєднується, окрім традиційних морфологічних характеристик, набори синтаксичних і семантичних ознак.

Крім того, існують окремі таблиці відмінкових закінчень (для змінних частин мови) для формування словоформ лексеми. Всі лексичні одиниці в таблиці відповідним чином проіндексовані та мають однакове інтерпретаційне значення на всіх етапах лінгвістичного аналізу.

Існує зворотний зв'язок між семантичним і синтаксичним аналізом, тому доцільно об'єднати ці дві процедури в одному блоці – *синтактико-семантичного аналізу*. Кінцевим завданням процедури синтаксичного аналізу є представлення кожного речення заданого ПМТ у вигляді синтаксичного дерева (лексем речення з семантико-синтаксичними відношеннями між ними).

Зв'язування слів у реченні відбувається поступово, від словосполучення до групи зв'язаних слів і, зрештою, до об'єднання всіх груп у реченні в одну структуру. Для встановлення зв'язку між окремими словами використовуються природномовні засоби вираження семантичних та синтаксичних відношень. У флективних мовах такими засобами є змінні частини повнозначних слів та службові слова.

Такі сегменти словосполучень, які кодують відношення між повнозначними словами, називаються *синтаксичними визначниками* [1]. Оскільки одному синтаксичному визначнику може відповідати декілька синтаксичних відношень, для однозначності визначення зв'язків між словами вводиться поняття *кореляторів* [1], які додатково враховують семантичні ознаки слів у словосполученні.

Вихідними даними для блоку синтаксичного аналізу є:

- результат попередніх етапів аналізу ПМТ (графемного і морфологічного);
- словник основ (містить основи слів та їх семантичні ознаки);
- список всіх можливих флексій слів;
- база даних з визначниками (містить синтаксичні визначники та списки кореляторів для кожного з них);
- база даних з кореляторами (кожен корелятор складається з назви відношення та списку пар семантичних ознак основ, між якими це відношення може існувати).

Незмінна та флективна частини слова та його семантична ознака для всіх слів ПМТ передаються з попереднього етапу аналізу (етап морфологічного аналізу). Синтаксичні характеристики в таблицях лексикографічної бази даних, наприклад для дієслів, включають лексико-семантичні валентності дієслів, що визначають ознаки найбільш вірогідного оточення та ознаки семантичних інтерпретацій, які на основі розпізнаних синтаксичних правил визначають семантичні відношення між словами у словосполученні. Якщо речення складається з декількох синтагм, то завдяки інтерпретації певної частини бази знань синтаксису відбувається їх об'єднання у єдине синтаксичне дерево з відповідними семантико-синтаксичними відношеннями.

Нижче описані основні етапи роботи блоку *синтактико-семантичного аналізу* природномовного тексту.

#### *Перший етап.*

Проводиться класифікація слів у реченні за семантичними ознаками відповідних основ у словнику. Зв'язування слів у реченні починається зі словосполучення, що визначає головне відношення (відношення між підметом і присудком) у цьому реченні. У випадку, коли таке словосполучення встановити неможливо, речення аналізується зліва направо, починаючи з перших повнозначних слів.

Для вибраного словосполучення формується синтаксичний визначник, який складається зі службових слів та флективних частин повнозначних слів словосполучення. Наприклад, для виразу «права та свободи» таким визначником буде конструкція типу «-а та -и». Якщо сформований визначник існує в базі даних визначників, йому буде відповідати список кореляторів. Тоді, враховуючи семантичні ознаки слів у словосполученні, у базі даних з кореляторами знаходиться потрібний корелятор, що і встановить зв'язок між словами. Однозначність визначення такого зв'язку забезпечується тим, що

для окремого визначника множини пар семантичних ознак для кореляторів з його списку не перетинаються.

Далі до словосполучення поступово приєднуються прилеглі повнозначні слова речення, шляхом встановлення зв'язку між новим словом та одним із слів опрацьованої частини речення. Таким чином створюється група зв'язаних слів. Важливим є вибір слова з групи зв'язаних слів, яке буде пов'язуватись з наступними словами. Це має бути або головне, або останнє слово групи. У випадку, коли неможливо встановити зв'язок між новим словом та словами групи, створюється нова група зв'язаних слів. В кінці аналізу всі створені групи поєднуються, за можливістю, в одну, яка відображає структуру зв'язків між всіма словами речення.

Неможливість встановлення зв'язку між окремими групами в реченні та їх об'єднання свідчить або про складне речення, частини якого не пов'язані (або пов'язані неявно) між собою, або про некоректні зв'язки між словами у групах. Для уникнення проблеми некоректного зв'язування слів необхідно проводити додатковий аналіз можливих зв'язків кожного наступного слова із словами групи і вибирати найбільш вірогідний, або ж розглядати всі можливі варіанти зв'язків (що недоцільно, враховуючи зростання кількості таких варіантів для кожного наступного слова).

#### *Другий етап.*

На другому етапі вирішується задача встановлення анафоричних зв'язків або заміни займенників у тексті на відповідні поняття (як правило іменники) [2]. Заміщенню підлягають деякі особові, відносні, вказівні, присвійні та зворотні займенники. Алгоритм заміщення будується на основі аналізу закономірностей (відображених в базі знань синтаксису ПМ) вживання займенників в природній мові і відрізняється в залежності від типу та граматичних характеристик займенника.

Із практики відомо, що поняттям, якому відповідає займенник, є узгоджене з ним за граматичними характеристиками найближче повнозначне слово, яке стоїть попереду, або слово, що входить до ядра даного чи попередніх речень. Тому вихідними даними для проведення заміни займенників є граматичні характеристики слів (отримані в результаті морфологічного аналізу) та синтаксично-семантичні зв'язки між ними (результат роботи синтаксичного блоку). Заміщення займенників дозволяє отримати додаткові зв'язки між відповідними поняттями у реченні та у всьому тексті, тому після цього проводиться корекція синтактико-семантичних зв'язків у окремих реченнях (із займенниками).

Результатом роботи блоку синтактико-семантичного аналізу є поняттява (логіко-семантична) структура ПМО.

#### *Виділення термінів предметної області з ПМТ у програмі “Конспект”*

Задача виділення термінів з тексту виникає при побудові онтологій, формуванні та поповненні словників предметної області, рубрикаторів та будь-якій інтелектуальній обробці текстових документів.

Для побудови понятійного апарату з текстів ПдО необхідно використовувати модель іменникових словосполучень, що виражаються схемою: іменник + слово, що погоджується з іменником. У цій моделі іменник є головним словом, а слово, що погоджується, - залежним і може виражатися як прикметником, так і іменником. [3]. У загальному випадку іменникові словосполучення можуть включати в свій склад наступні класи слів: іменники, прикметники, прийменники, сурядні союзи і прислівники. Кількість слів в іменникових словосполученнях коливається від двох до п'ятнадцяти і в середньому складає три слова. У роботі [3] приводиться 18 шаблонів іменникових словосполучень, які можна використовувати для виділення термінів ПдО. У російській мові синтаксична структура термінів ПдО більш ніж в 90 відсотків випадків має наступну конструкцію [4]:

1. Одиночні іменники, прикметники, і скорочення;
2. Іменникова група: іменник + іменник в родовому відмінку;
3. Іменникова група: прикметник + іменник
4. Іменникова група: прикметник + прикметник + іменник
5. Іменникова група: іменник + прикметник + іменник в родовому відмінку.

Експерименти по виділенню термінів показали, що в українській мові для ПдО ділова проза доцільно збільшити кількість слів в синтаксичній структурі іменникових словосполучень до 5. Розглянемо метод автоматичного виділення однослівних та багатослівних термінів, який, крім шаблонів, використовує результати синтактико-семантичного аналізу тексту [5]. Розпізнання поверхневих семантичних відношень здійснюється на попередньому етапі аналізу тексту.

Процедура виділення поділяється на два етапи. На першому відбувається безпосередній пошук у тексті слів та словосполучень – кандидатів у терміни. В якості однослівних термінів вибираються іменники та аббревіатури. Багатослівні терміни формуються за допомогою визначених типів відношень між словами речення, шляхом поступового приєднання слів до однослівного терміну-іменника. Для термінів іменникового типу використовуються такі типи відношень між словами: об'єктне, приналежність (між двома іменниками), визначальне (між прикметником та іменником), однорідні слова (між двома іменниками або двома прикметниками) та інші. Виділені групи слів перевіряються на відповідність визначеним шаблонам іменникових словосполучень, загальна кількість яких складає 35.

Шаблони термінів мають вигляд послідовності морфологічних характеристик слів (наприклад, іменник\_прикметник\_іменник – шаблон для терміну «онтологія предметної області»). Слова, що складають термін, в реченні не обов'язково стоять поруч, але обов'язково пов'язані між собою певним типом відношень. Це дозволяє, наприклад, з речення «Побудова онтології зазначеної предметної області.» виокремити термін «онтологія предметної області». Якість виділення таким чином термінів в значній мірі

залежить від якості роботи синтаксичного блоку, а саме, від коректності класифікації слів за семантичними ознаками та встановлених зв'язків між словами.

На другому етапі список кандидатів у терміни фільтрується: враховується значимість виділених словосполучень (наближеність у дереві розбору до підмета чи присудка речення) та частота, з якою вони зустрічаються у тексті. Відфільтрований список виводиться на екран користувачу.

### *Конспектування фрагментів ПМТ у системі “Конспект”*

Розглянемо алгоритм виділення і стислого конспектування фрагментів природномовних текстів, що відносяться до заданої теми, яка задається ключовим словом або словосполученням [5,6].

Постановкою задачі концептуального аналізу природномовних текстів є вилучення з тексту його семантичної комунікативної складової, в ролі якої виступає лінгвістична категорія – тема документу.

Тематичний аналіз текстів спирається на концептуальну систему онтології асоціацій та на зв'язки між словами робочого мовного словника та концептами. Зв'язки мають семантичні ваги. Онтологія асоціацій є мережевою структурою для представлення знань про довколишній світ, які можуть бути представленими в текстах природної мови. Вона використовує концепти в якості асоціативних інтерпретаторів одиниць мови. Структура онтології асоціацій є адекватною до поставленої задачі і конструктивною для її вирішення.

За результатами синтактико-семантичного аналізу у вихідному тексті відокремлюються речення, які вміщують “ядерні конструкції”. Термін “ядерні конструкції” використовується в трансформаційній граматиці для визначення простого базового твердження, шляхом трансформації якого формується речення в цілому. В даному випадку ядерною конструкцією є речення, яке складається з підмета, присудка та зв'язку, що їх об'єднує.

Таким чином, на даному етапі із вихідного тексту для подальшого аналізу відокремлюються повноскладові речення. Для кожного з них формується  $n$  - крокове розширення ядра, а саме – така частина речення, яка вміщує його ядро, а також слова, що зв'язані в дереві залежностей з елементами ядра шляхами, загальна довжина яких не перевищує  $n$  (параметр, що має бути апріорі заданим).

В результаті синтактико-семантичного аналізу формується масив  $n$  - крокових розширень повноскладових речень вихідного тексту. Даний масив також називається масивом значимих слів. Він є об'єктом наступного аналізу за допомогою алгоритму генерації ключових слів.

Первинний ключ (що задає тему) вказується користувачем безпосередньо перед початком роботи програми. Цей ключ може не відповідати головній темі тексту: він відповідає інтересам користувача щодо повідомлення. Тобто, користувач бажає вилучити з тексту ту інформацію, яка

представляє для нього інтерес. Його цікавить локальна тема повідомлення і він перевіряє текст, наскільки повно ця локальна тема є в ньому відображеною. Результатом перевірки – тематичного аналізу є протокол, тобто – послідовність вторинних ключів.

Початковий ключ задає напрям пошуку нових ключів, семантично пов'язаних з первинним.

Принцип пошуку чергового ключа є простим. Він базується на такому припущенні: попередній ключ та новий претендент знаходяться в  $n$  – крокових розширеннях їхніх ядер, а їх асоціативні ознаки (концепти з онтології асоціацій) перетинаються. Перше припущення означає, що новий ключ треба шукати в дереві розбору речення поблизу від головних членів речення. Друга теза потребує, щоб між знайденим ключем та претендентом існували тісні семантичні (асоціативні) зв'язки. Ці дві умови пошуку можна вважати логічними.

Таким чином встановлюється семантична взаємозалежність між ключовими словами та кореляція ключів з темою, що відшукується.

Знайдена послідовність ключів відслідковує тему. Але виникає необхідність перетворити множину ключів на семантично-зрозумілий переказ теми. З цією метою під керуванням ключів з тексту відбираються окремі фрази (у яких ключ входить до  $n$ -крокових розширень ядра повноскладових речень) вихідного тексту і концентруються в конспекті.

За рахунок вибору в якості чергового ключа знаменного слова, що є найбільш асоціативно близьким до попереднього ключа, розрізнені фрагменти поєднуються і кожна наступна фраза конспекту виявляється семантично поєднаною з попередньою. За об'ємом та зв'язністю викладення автоматичний конспект є близьким до традиційного уявлення про властивості конспекту.

Загальний вигляд головного вікна програми “Конспект” наведено на рисунку 8.1.



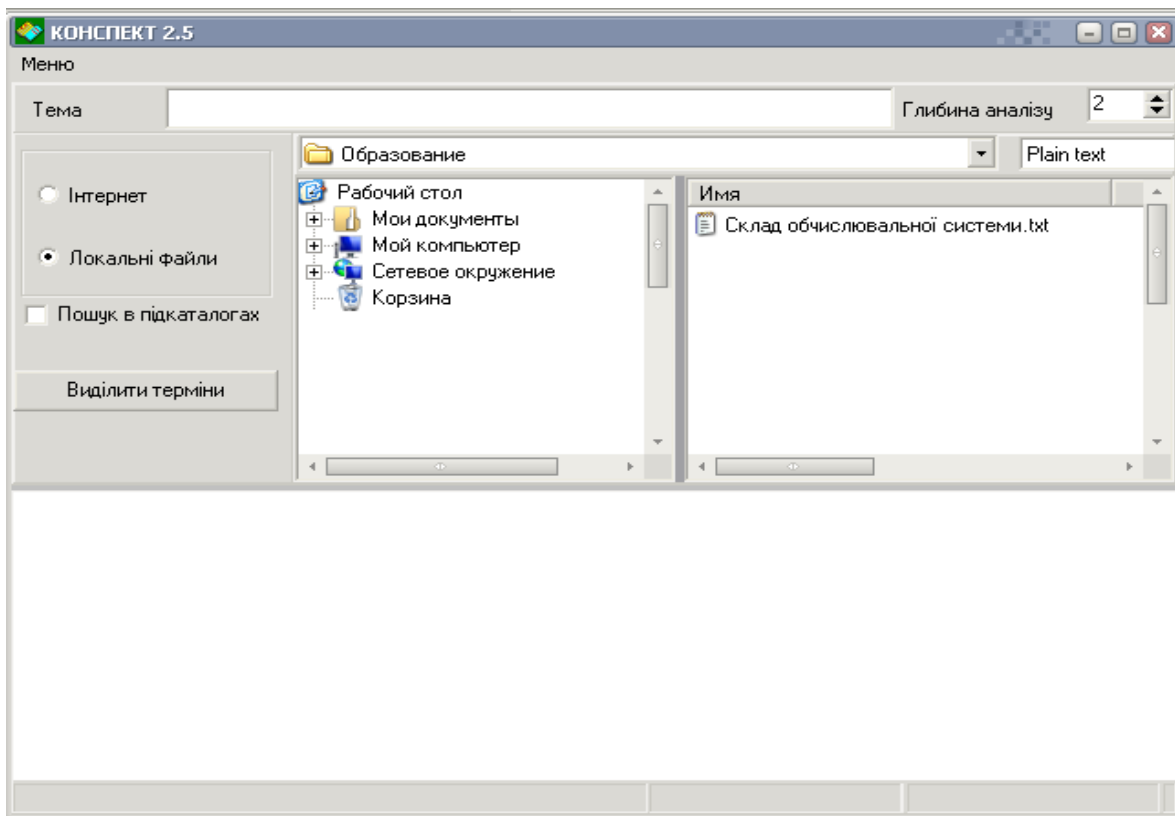


Рис. 8.1. Загальний вигляд головного вікна програми “Конспект”

## 8.2. Методика обробки текстів у середовищі системи Конспект

Науково-дослідницька робота з української мови сприяє розвитку такої визначальної особистісної якості, як креативність. Креативний школяр прагне знаходити самостійні рішення, перебуває в постійному пошуку оригінальних шляхів вирішення цікавої для нього проблеми, а якщо йдеться про мовознавчу площину, то такий учень ще й досконало володіє засобами літературної мови, має сформовані читацькі інтереси, є цікавим співрозмовником. Юні мовознавці стають професійними дослідниками рідного слова, цікавляться творчістю літераторів-класиків. Беззаперечно, сучасні інформаційні технології стають у нагоді як аналітичний інструментарій для виконання науково-дослідницької роботи, зокрема, з української філології.

Мовнотворчість Тараса Шевченка на сьогодні так само, як і в ХХ, і в ХІХ століттях, викликає інтерес і у відомих учених, і в юних дослідників; "Кобзар" читають і перечитують, кожного разу знаходячи у цій невичерпній скарбниці нові й не помічені досі перлини. Наведемо для прикладу уривок зі статті відомої української вченої – педагога-мовознавця – Любові Мацько: "Для поетичної мови Шевченка є характерним вживання лексеми "слово" саме в її узагальнених значеннях: як мови людської взагалі, як української мови, як високої мови, правдивої, благородної – сакральної.<...> Широка понятійна сфера (слово і мова, слово і розмова, слово і серце, слово і душа, слово і сльози, слово і правда, слово і істина, слово і вчення, наука)

конкретизується у різноманітних текстових реалізаціях. <...> У творчості Шевченка знаходимо широкий семантичний спектр і оригінальне трактування "вічних" світових понять-концептів, у колі яких хоче чи не хоче, а торує свій шлях кожна людина" [7].

Кількісно-якісна характеристика лексем текстів збірки "Кобзар", виконана за допомогою програми TextTermin, дозволяє досліднику встановити кількість слововживань та з'ясувати за допомогою знайдених автоматично контекстів семантичні відтінки, надані Т. Г. Шевченком окремим словам, побудувати між цими словами «дерево» відносин. До того ж, такі дослідження збагачують сучасне мовознавство, зокрема його сучасну галузь - комп'ютерну лінгвістику, дають матеріал для теоретичної і практичної стилістики, літературознавства та інших філологічних галузей.

Користуючись функціями системи **TextTermin**, юний науковець у цілком доступній формі отримує важливі для узагальнених об'єктивних висновків результати, підтвердить або спростує робочу гіпотезу свого дослідження.

Отже, організована у такий спосіб навчальна і дослідницька діяльність учнівської і студентської молоді сприяє створенню полісуб'єктного освітнього середовища.

Наведені нижче матеріали є орієнтиром для студентів, школярів та вчителів/керівників науково-дослідницької роботи, – у процесі поетапного вирішення будь-якої проблеми наукового філологічного дослідження.

### **ЗМІСТ МЕТОДИКИ**

#### ***Дослідження концепта "Україна" у текстах збірки Т. Г. Шевченка "Кобзар" за допомогою системи Коспект***

#### **Підготовчий етап**

#### **• З'ясування та обґрунтування позицій:**

##### ***Наукова галузь дослідження.***

*Дослідження концепта «Україна» здійснюється у площині стилістики української мови як галузі конкретного мовознавства. Під час роботи використовуються ресурси Інтернету і програма TextTermin.*

##### ***Актуальність дослідження.***

*Актуальність теми зумовлена важливістю творчої спадщини Т. Г. Шевченка - патріота України - для становлення і розвитку сучасної української мови, літератури і культури. Аналіз кількості вживань лексеми «Україна» у текстах творів збірки "Кобзар", а також дослідження різноманітних семантичних відтінків цього слова допомагає усвідомити мовні особливості індивідуального стилю автора та відображення у ньому ментальності українців.*

##### ***Теоретична і практична цінність та перспективи дослідження.***

*Теоретичне значення дослідження полягатиме в окресленні філософсько-педагогічних, психологічних та лінгвістичних засад концепта "Україна"; практична цінність - у встановленні семантичних відтінків поняття «Україна» на основі кількісно-якісного аналізу масиву текстів творів Тараса Шевченка; у доборі в реченнях прикладів вживання слова «Україна» та похідних від нього слів і виразів. Матеріали роботи зможуть використати не тільки школярі, але й мовознавці, літературознавці, викладачі-філологи, вчителі-словесники, студенти.*

• **З'ясування значення ключових слів теми.**

Для з'ясування семантики ключових слів теми використовуються словники, посібники, підручники, довідники, монографії, наукова періодика, матеріали, знайдені у відкритих джерелах Інтернету, які підтверджують висловлене. Наприклад: *концепт* – "зміст поняття, смислове значення імені".

• **Формулювання мети дослідження, якої можна досягти шляхом вирішення завдань (відповідно до мікротем у рамках теми) – кроки 1 - 4.**

Мета роботи – окреслити філософсько-педагогічні, психологічні, лінгвістичні засади концепта «Україна» та встановити семантичні відтінки цього слова та похідних від нього слів на основі кількісно-якісного аналізу масиву текстів творів Тараса Шевченка засобом програми *TextTermin*.

**Основний етап**  
**Вирішення завдань дослідження**

**КРОК 1.** Ознайомлення з Інструкцією для користувача програмою *TextTermin*.

**КРОК 2.** Пошук якісної електронної версії збірки "Кобзар" в мережі Інтернету (бажано у форматі *MicrosoftOfficeWord*).

**КРОК 3.** За допомогою інструкції переведення тексту "Кобзаря" у формат Блокнот. Збереження джерела у папці (рисунок 8.2):

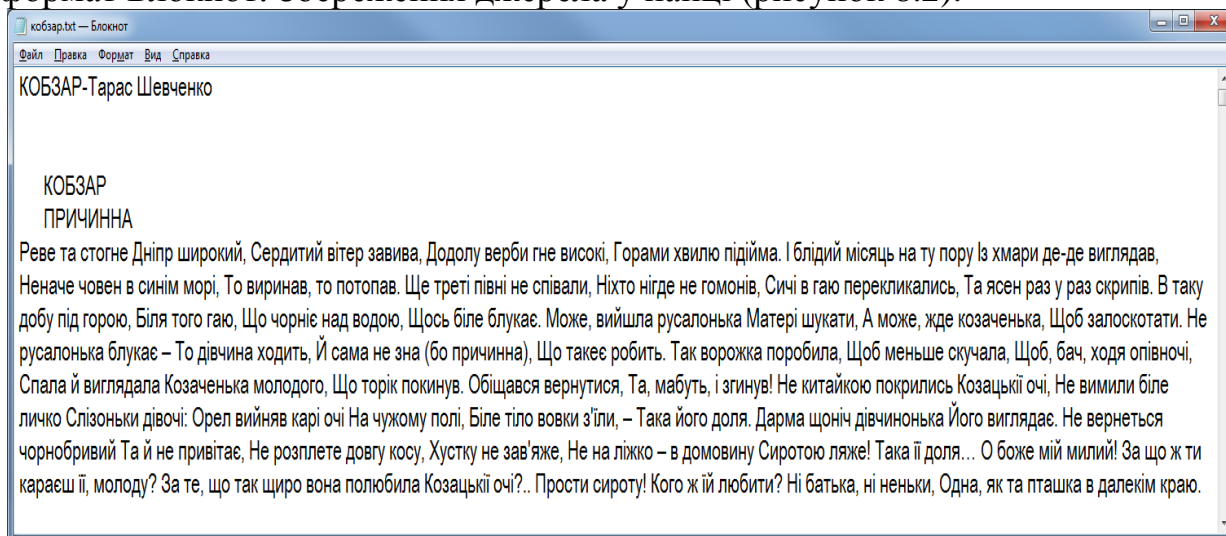


Рис. 8.2. Текст джерела

**КРОК 4.** Запуск програми TextTermin. Виділення джерела ("Кобзар"), лінгвістичний аналіз якого здійснить програма (рисунок 8.3).

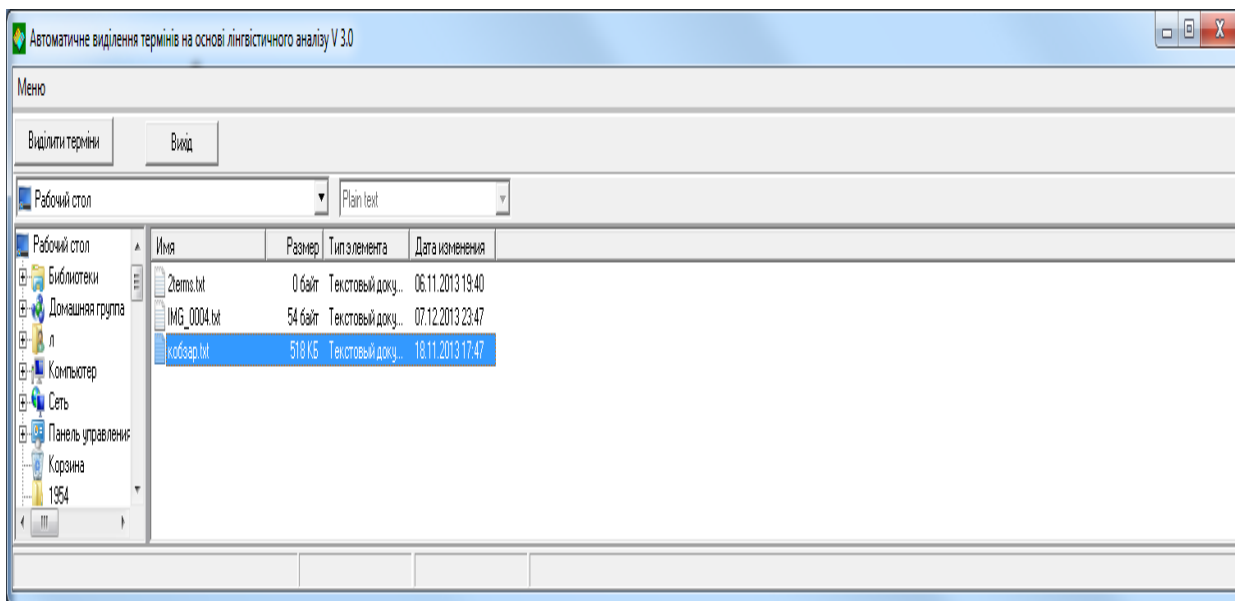


Рис. 8.3. Вибір джерела

**КРОК 5.** Ознайомлення з результатами лінгвістичного аналізу (рисунок 8.4).

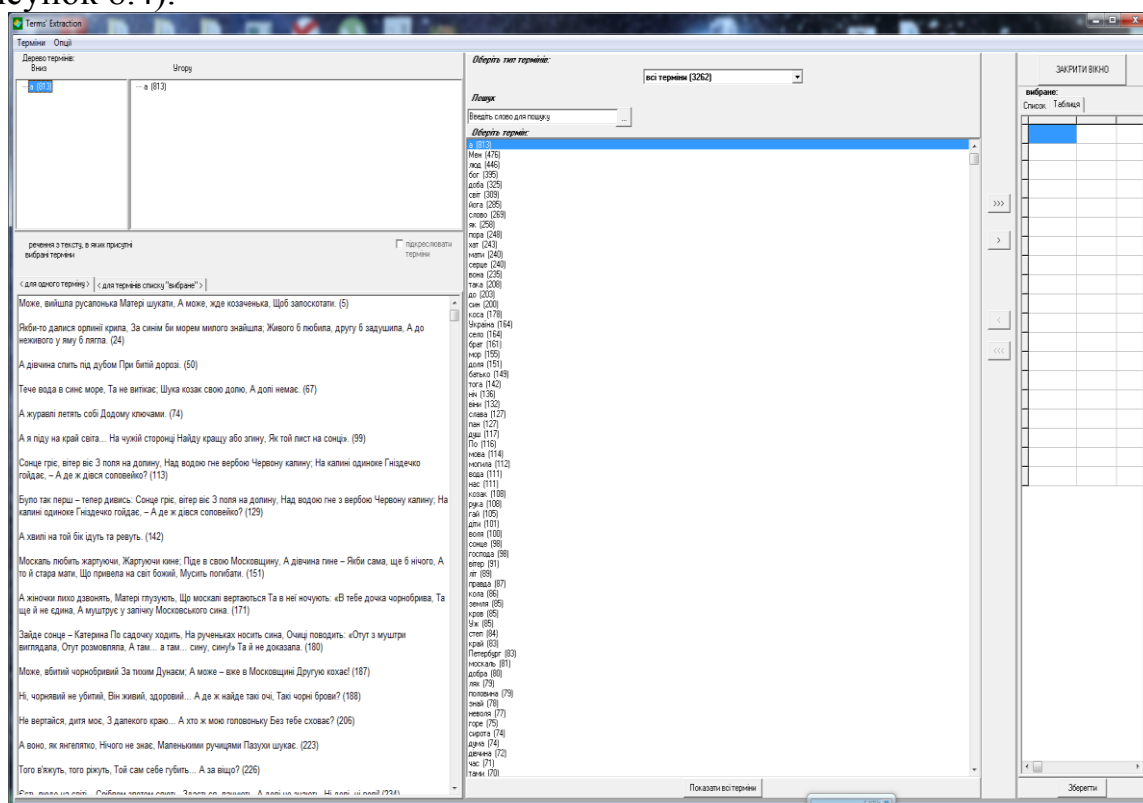


Рис. 8.4. Створення програмою списку термінів

Як бачимо з рисунку (екран монітора), загальна кількість лексем, виділених програмою, - 3262. Слово "Україна" повторюється 164 рази.

**КРОК 6.** Виділимо слово "Україна" у списку термінів (рисунок 8.5). За бажанням можна задати опцію "підкреслювати терміни". Тоді у контекстах зліва слово "Україна" буде підкреслюватися.

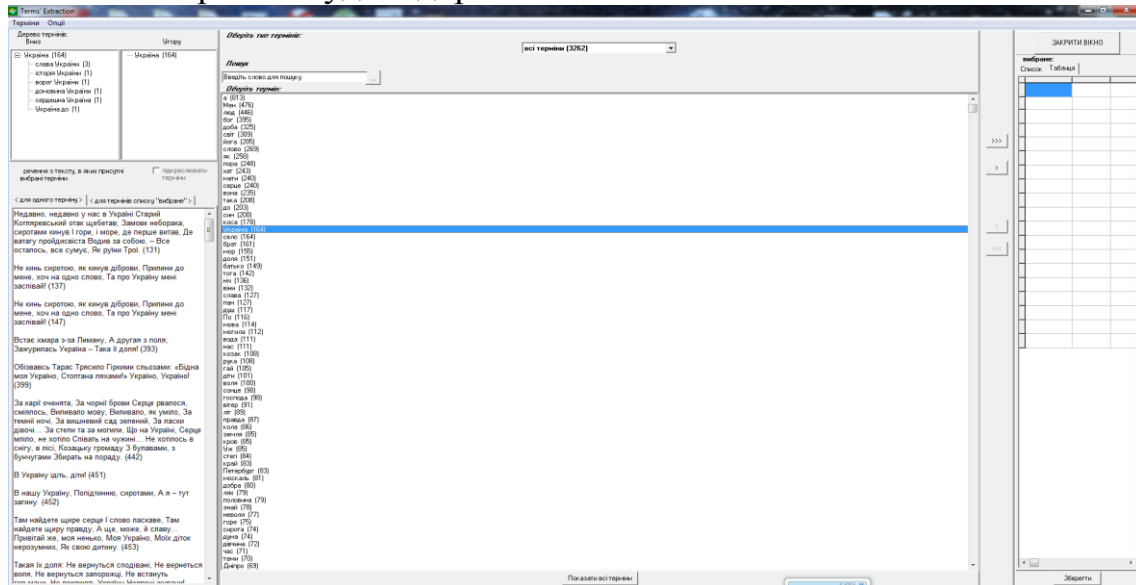


Рис. 8.5. Робота програми із заданим словом

**КРОК 7.** Для подальшого аналізу концепту "Україна" у творах збірки "Кобзар" бажано зберегти контексти з підкресленнями окремим документом (можна скопіювати контексти і вставити в окремий файл). Наприклад:

Недавно, недавно у нас в Україні Старий Котляревський отак шебетав; Замокв неборака, сиротами кинув І гори, і море, де перше витав, Де ватагу пройдисвіста Водив за собою, – Все осталося, все сумує, Як руїни Трої. (131)

Не кинь сиротою, як кинув діброви, Прилини до мене, хоч на одно слово, Та про Україну мені заспівай! (137)

Встає хмара з-за Лиману, А другая з поля; Зажурилась Україна – Така її доля! (393)

Обізвався Тарас Трясило Гіркими сльозами: «Бідна моя Україно, Стоптана ляхами!» Україно, Україно! (399)

За карії оченята, За чорнії брови Серце рвалося, сміялось, Виливало мову, Виливало, як уміло, За темнії ночі, За вишневий сад зелений, За ласки дівочі... За степи та за могили, Що на Україні, Серце мліло, не хотіло Співать на чужині... Не хотілось в снігу, в лісі, Козацьку громаду З булавами, з бунчугами Збирають на пораду. (442)

В Україну ідіть, діти! (451)

В нашу Україну, Попідтинню, сиротами, А я – тут загину. (452)

Там найдете щире серце І слово ласкаве, Там найдете щирю правду, А ще, може, й славу... Привітай же, моя ненько, Моя Україно, Моїх діток нерозумних, Як свою дитину. (453)

Аналіз кожного з контекстів допомагає, наприклад, з'ясувати семантичні відтінки слова "Україна": територія, населення, країна, ненька; зафіксувати перенос значення (Зажурилась Україна – метонімія); встановити граматичні відмінкові форми слова: в Україну, на Україні, про Україну, Україно тощо.

**КРОК 8.** Для подальшого створення «дерева» відносин між словом "Україна" та іншими словами у творах збірки "Кобзар" створюємо таблицю, у напівручному режимі копіюючи слова у списку термінів та вставляючи їх послідовно у рядки таблиці (рисунк 8.6).

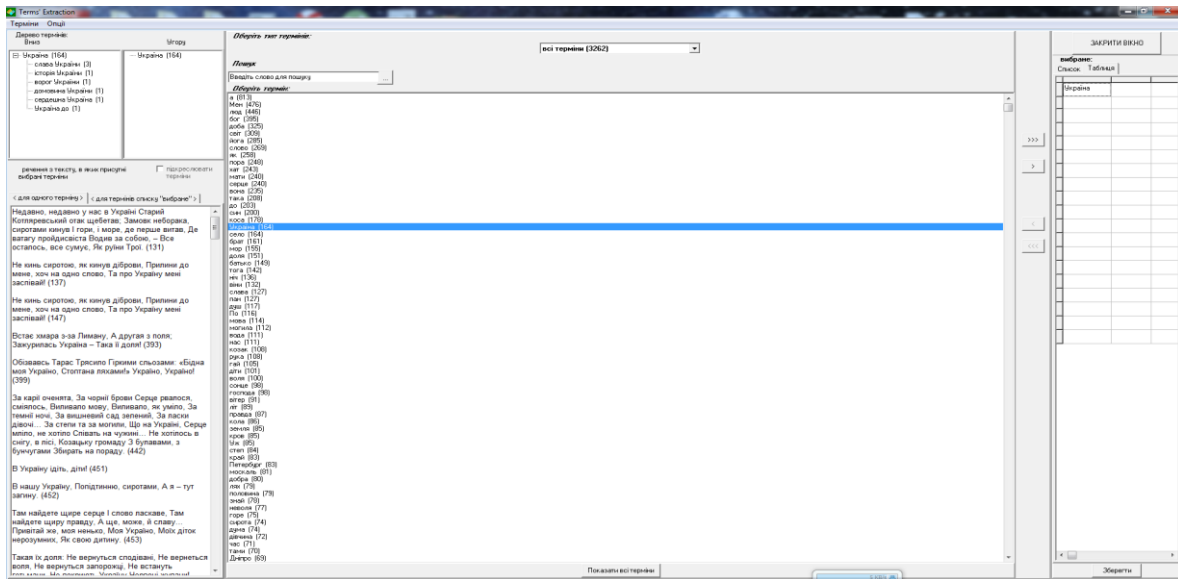


Рис. 8.6. Заповнення таблиці  
Збережена таблиця має вигляд (рисунк 8.7):

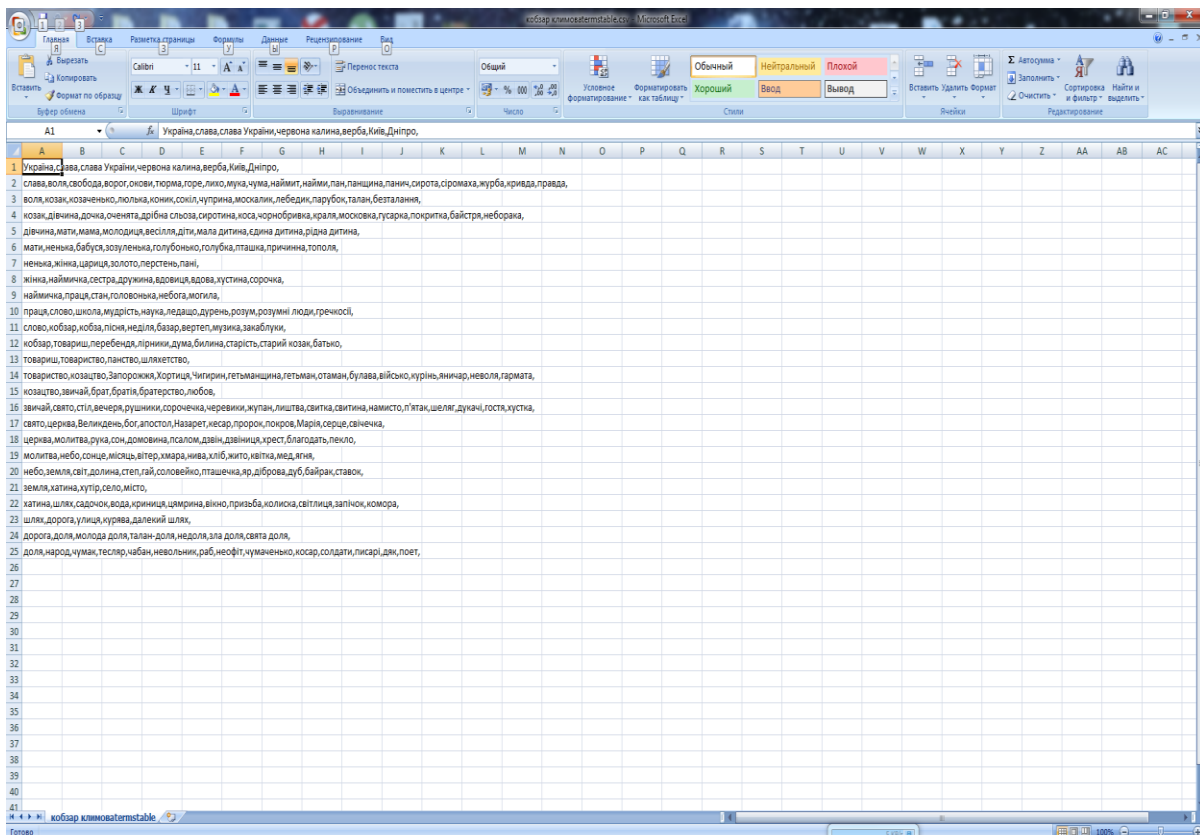


Рис. 8.7. Таблиця – основа для створення "дерева" відносин між словами

**КРОК 9. Оформлення результатів практичної частини дослідження і повного тексту науково-дослідницької роботи.**

*Отже, наведена як приклад технологія контент-аналізу лексики художнього твору (за текстами збірки Т. Г. Шевченка «Кобзар») – на прикладі дослідження концепту "Україна" учнями (студентами) доводить думку про те, що послідовне вирішення завдань дозволяє за допомогою функцій системи Коспект та відкритих джерел Інтернету досягти поставленої мети[8].*

## РОЗДІЛ 9. МЕТОДИКА ПОБУДОВИ ОНТОЛОГІЇ (НА ПРИКЛАДІ ОПИСУ ЖИТТЕДІЯЛЬНОСТІ ТАРАСА ГРИГОРОВИЧА ШЕВЧЕНКО)

### 9.1. Системаобробки предметних знань “Конфор”

У метапроцедурах атрибутивного аналізу використовуються методи гіпотетичного моделювання, зокрема індуктивне виведення і виведення за аналогією. В результаті атрибутивного аналізу формується аналітична імпліцитна інформація, необхідна для вибору основних елементів поведінкового акту: цілей, дій і актантів дій. При рішенні задач різних типів в мережі понять можуть формуватися мережеві структури, що найбільш адекватно відображають особливості вирішуваних задач. Вимогам атрибутивного аналізу повною мірою відповідають пірамідальні мережі [1, 2].

Пірамідальною мережею  $Q$  називається ациклічний орієнтований граф  $Q=(U,E)$  ( $U$  – множина вершин,  $E$  – множина дуг), в якому нема вершин, які мають одну дугу, що заходить. Множина вершин пірамідальної мережі це множина  $U = R_Q \cup C_Q$ , де  $R_Q$  – множина рецепторів пірамідальної мережі,  $C_Q$  – множина концепторів пірамідальної мережі. Вершини, які не мають дуг що заходять, називають *рецепторами*, інші вершини – *концепторами*. Рецептори відповідають окремим значенням ознак з описів об'єктів, тобто між множиною  $\bigcup_{i=1}^n D_i$  та множиною рецепторів  $R_Q$  існує взаємооднозначна відповідність  $v: \bigcup_{i=1}^n D_i \rightarrow R_Q$ . Концептори відповідають комбінаціям значень ознак, що ідентифікують об'єкт в цілому або відповідним спільним частинам описів декількох об'єктів. Приклад пірамідальної мережі зображено на рисунку 9.1.

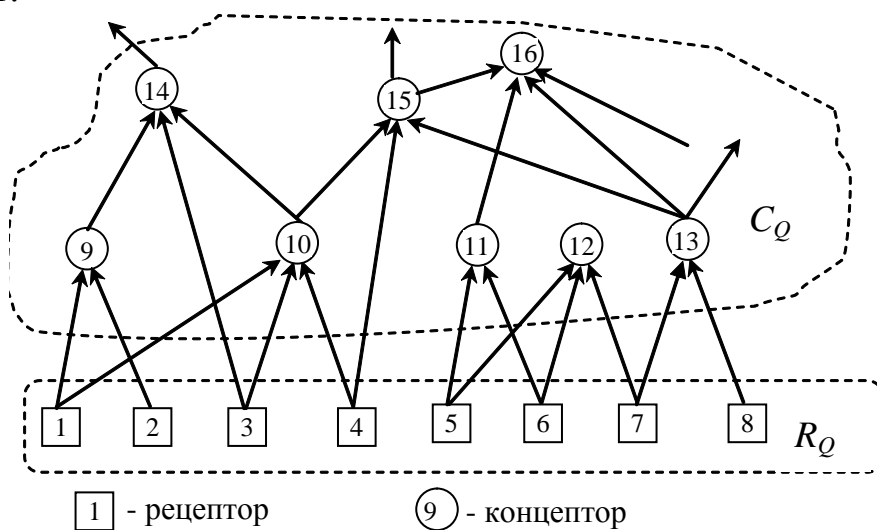


Рис. 9.1. Приклад семантичної мережі пірамідальної структури

Підграф пірамідальної мережі, що включає вершину  $a$  та всі вершини, з яких існує шлях до  $a$ , називається пірамідою вершини  $a$ . Вершини, що



входять до піраміди вершини  $a$  утворюють її субмножину – **Subset**( $a$ ). Множина вершин, до яких існують шляхи з вершини  $a$ , називається її супермножиною – **Superset**( $a$ ). Позначимо через **Osubset**( $a$ ) множину вершин з субмножини вершини  $a$ , що пов'язані дугами з вершиною  $a$ , а через **Osuperset**( $a$ ) – множину вершин з супермножини вершини  $a$ , що пов'язані дугами з вершиною  $a$ .

При побудові мережі вхідною інформацією служать набори значень ознак, що описують деякі об'єкти. У різних задачах це можуть бути імена властивостей, відносин, станів, дій, об'єктів або класів об'єктів. Концептори відповідають описам об'єктів в цілому і перетинам описів.

У початковому стані мережа полягає тільки з рецепторів. Концептори формуються в результаті роботи алгоритму побудови мережі. В пірамідальній мережі визначено процес збудження. При введенні ознакового опису об'єкту, рецептори, відповідні значенням ознак, що входять в опис, переводяться в стан збудження. Якщо група рецепторів була переведена в стан збудження, то збудження розповсюджується по мережі за наступним правилом: концептор  $a$  переводиться в стан збудження, якщо збудженими є всі вершини з множини **Osubset**( $a$ ). Процес збудження розповсюджується по мережі. Рецептори і концептори зберігають стан збудження протягом виконання всіх операцій побудови мережі.

#### *Правила побудови семантичних мереж пірамідальної структури*

В результаті роботи алгоритму побудови зростаючої пірамідальної мережі (ЗПМ) здійснюється виділення фрагментів описів, що повторюються в описах об'єктів з навчаючої вибірки, та формування відповідних концепторів. Новий об'єкт порівнюється з раніше переглянутими об'єктами, в результаті чого у мережі виникають концептори, що відповідають комбінаціям значень ознак, які є спільними для описів декількох об'єктів.

При обробці нового об'єкту  $X$  множина рецепторів  $R_x = \{v(d) | d \in D_x\}$  переводиться в стан збудження. Збудження розповсюджується по мережі. Якщо  $F_a$  - підмножина збуджених вершин  $O$ -субмножини вершини  $a$ ;  $G$ -множина збуджених вершин мережі, що не мають інших збуджених вершин в своїх супермножинах, то введення нових концепторів у мережу буде здійснюватись за наступними основними правилами.

#### *Правило А1.*

Якщо вершина  $a$  не збуджена та **Card**( $F_a$ )>1, то дуги, що з'єднують вершини з множини  $F_a$  з вершиною  $a$ , ліквідуються, та у мережу вводиться новий концептор, який з'єднується дугами, що виходять з вершин множини  $F_a$  та дугою, що заходить до вершини  $a$ . Нова вершина знаходиться у стані збудження. Виконання правила А1 показано на рисунку 9.2 (I, II). Мережа II виникає після збудження в мережі I рецепторів 2, 3, 4, 5.

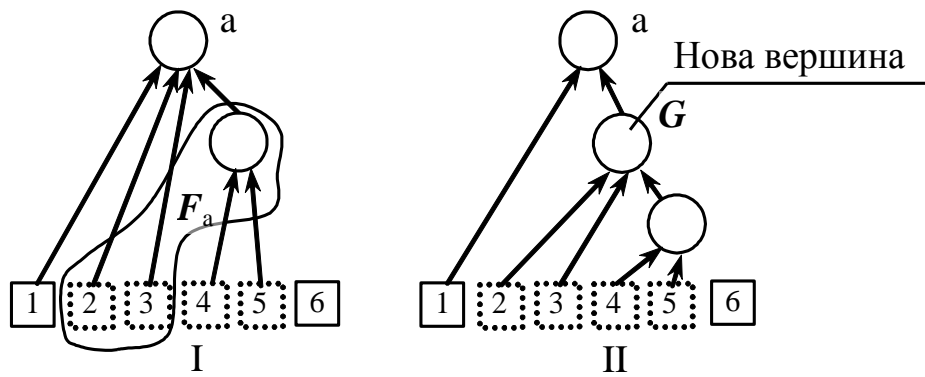


Рис. 9.2. Правило A1 побудови пірамідальної мережі

Як випливає з правила A1, умовою введення в мережу нової вершини є ситуація, коли деяка вершина мережі виявляється не повністю збудженою (збуджуються не всі, але не менше двох вершин її 0-субмножини). Нові вершини вводяться в 0-субмножину не повністю збуджених вершин. Вони представляють в мережі перетини описів об'єктів.

Після введення нових вершин до всіх ділянок мережі, де задовольняється правило A1, виконується правило A2.

*Правило A2.*

Якщо  $\text{Card}(G) > 1$ , то до мережі приєднується новий концептор, який з'єднується дугами, що виходять з усіх вершин множини  $G$ . Нова вершина знаходиться у збудженому стані.

Виконання правила A2 ілюструється рисунку 9.3 (I,II). Мережа II виникає після збудження в мережі I рецептори 2,3,4,5,6.

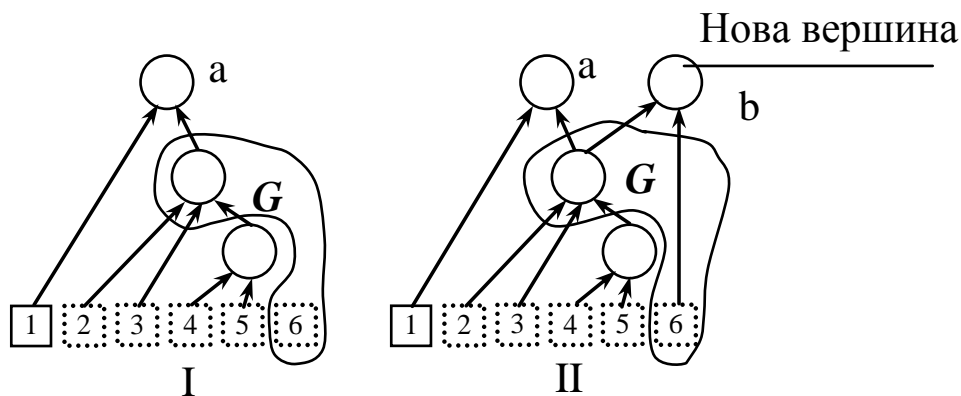


Рис. 9.3. Правило A2 побудови пірамідальної мережі

У правилі побудови ЗПМ - A1 основним відношенням, яке здійснює структурування, є відношення перетину між множиною рецепторів, збуджених при введенні опису об'єкту і деякою множиною рецепторів, що входять в піраміди ЗПМ. Правило A2 завершує побудову піраміди, що представляє опис введеного об'єкту в цілому.

Пірамідальні мережі зручні для виконання різних операцій асоціативного пошуку. Наприклад, можна вибрати всі об'єкти, що включають задане поєднання значень ознак, простежуючи шляхи, що виходять з вершини мережі, яка відповідає цьому поєднанню. Для вибірки всіх об'єктів,

описи яких перетинаються з описом заданого об'єкту, досить прослідкувати шляхи, що виходять з вершин піраміди цього об'єкту. Алгоритм побудови мережі забезпечує автоматичне встановлення асоціативної близькості між об'єктами за спільними елементами їх описів.

Важливою властивістю пірамідальних мереж є їх ієрархічність, що дозволяє природним чином відображати структуру складених об'єктів і родовидові зв'язки.

Концептори мережі відповідають поєднанням значень ознак, що визначають окремі об'єкти і кон'юнктивні класи об'єктів. При включенні збуджених вершин в піраміду об'єкту здійснюється прив'язка об'єкту до класів, визначення яких представлені цими вершинами. Таким чином, при побудові мережі формуються кон'юнктивні класи об'єктів, тобто здійснюється класифікація без вчителя. Властивість пірамідальної мережі здійснювати класифікацію має велике значення для моделювання середовищ і ситуацій.

Перехід від конвергованих представлень об'єктів (концепторів) до розгорнутих (наборів рецепторів) здійснюється шляхом перегляду пірамід у різних напрямках (зверху вниз і у зворотному напрямку).

Для того, щоб виконувати класифікацію, діагностику або прогнозування, потрібно знати залежності невідомих величин, визначуваних при рішенні цих задач, від відомих. Виникає необхідність в знанні закономірностей, що характеризують об'єкт дослідження.

Проблема автоматизації процесу виділення складних закономірностей одержала назву «виявлення знань» (knowledge discovery). Який тип знань є більш переважним при діагностиці і прогнозуванні? Можна використовувати одно-, двох- або тривимірні залежності, але для отримання надійнішого висновку краще мати модель, що відображає всі істотні зв'язки, властиві об'єктам, щодо яких розв'язуються ці завдання. Модель класів об'єктів, що використовується для вирішення завдань класифікації, діагностики і прогнозування, повинна включати всі найбільш важливі ознаки, що характеризують клас, і відображати характерні для цього класу логічні зв'язки між істотними ознаками. Тому ми, перш за все, зосередимо увагу на формування узагальнених логічних багатовимірних моделей класів об'єктів. Такого роду моделі, по суті, є поняттями, відповідними класам об'єктів.

У логіці поняття звичайно визначається як думка, що відображає суть об'єктів. Більшість понять, що вживаються, є результатом узагальнення уявлень про об'єкти деякого класу за істотними, специфічними для цього класу ознаками. До ознак відноситься все, що характеризує об'єкти і може бути використане в таких логічних операціях, як виділення, розпізнавання, ототожнення і т.п. Слід зазначити, що розділення ознак на істотні і неістотні в значній мірі умовно і залежить від завдань, для вирішення яких вони використовуються.

Поняття по відношенню до слова – його сенс, інформація, яку містить слово про можливих денотатах – реаліях, що позначаються словом. Наприклад, денотатом слова "книга" може бути деяка конкретна книга, про

яку йде мова в тексті. Поняття, пов'язане із словом "книга", характеризує множину всіх книг – можливих денотатів цього слова. Множина узагальнених в понятті об'єктів складає його об'єм. Поняття є лексичним правилом вживання слова. У системі знань поняття грають роль базових елементів, з яких складаються думки і інші логічні форми мислення. Перехід від ступеня пізнання до абстрактного мислення, по суті, означає перехід від віддзеркалення світу у формі відчуттів і уявлень до віддзеркалення його в поняттях. Будучи концентратом знання, підсумком деякого етапу пізнання, поняття служить найважливішим інструментом формування нових знань і рішення задач.

На основі наявного запасу понять здійснюються класифікація, узагальнення, структуризація сприйнятої інформації, включення її в систему знань. У цих процесах реалізуються дві основні функції понять – розпізнавання і генерація моделей елементів світу, в якому оперує носій знань. Генерація моделей є процесом формування моделей конкретних об'єктів шляхом введення в поняття інших понять і констант. Успіх рішення задач, що включають генерацію моделей, залежить від того, наскільки використовувані поняття правильно і повно характеризують відповідні класи елементів світу.

Тепер можна дати конструктивніше визначення поняття, придатніше при розгляді інформаційно-технічних аспектів проблем формування і обробки понять.

*Поняття* – елемент системи знань, що є узагальненою логічною ознаковою моделлю класу об'єктів, за допомогою якої реалізуються процеси розпізнавання і генерації моделей конкретних об'єктів. Множина узагальнених в понятті об'єктів складає його *об'єм*.

#### *Індуктивне формування понять в семантичних мережах пірамідальної структури*

Навчання ЗПМ полягає у формуванні в них структур, що представляють поняття.

Розглянемо завдання індуктивного формування понять для множини об'єктів  $V_1, V_2, \dots, V_n$ . Нехай  $L$  – множина об'єктів, які використовуються як вибірка для навчання. Мають місце співвідношення  $L \cap V_i \neq \emptyset$  і  $V_i \not\subset L (i = 1, 2, \dots, n)$ . Задані ознакові описи всіх елементів  $L$ . Для кожного об'єкта  $l \in L$  відомо відношення  $l \in V_i$ . Потрібно за допомогою аналізу  $L$  сформувати  $n$  понять с об'ємами  $V_1, V_2, \dots, V_n$ , достатніх для правильного розпізнавання всіх об'єктів  $l \in L$ .

Поняття, сформоване на основі вибірки для навчання, в загальному випадку є наближенням до дійсного поняття, причому ступінь близькості цих понять залежить від показності вибірки для навчання, тобто від того, наскільки повно в ній відображені особливості об'єму поняття.

При формуванні поняття, відповідного множини  $V_i$ , об'єкти навчаючої вибірки, що входять в  $V_i$ , розглядаються як приклади множини  $V_i$  (позитивні об'єкти), а об'єкти, що не входять в  $V_i$ , – як контрприкладів множини  $V_i$  (негативні об'єкти).

Після побудови пірамідальної мережі утворюється структура, яка є представленням описів об'єктів навчаючої вибірки. Кожна вершина мережі визначає кон'юнктивний клас об'єктів.

Таким чином, об'єкт належить до всіх класів об'єктів, які визначаються вершинами, що входять до його піраміди, і також кожний об'єкт, піраміда якого містить в собі деяку вершину, належить класу об'єктів, який вона представляє. Поєднання ознак, виділені на першому етапі, є «будівельним матеріалом», з якого формується логічна структура поняття на другому етапі. На основі побудованої пірамідальної мережі визначено процес індуктивного формування понять.

Суть процесу формування понять в пірамідальних мережах полягає в аналізі елементів побудованої мережі та доборі з них таких, які найчастіше зустрічаються в об'єктах навчаючої вибірки. Відібрані вершини помічаються як контрольні елементи поняття, що формується. Поняття класу  $B_i$  об'єктів представляється у мережі сукупністю контрольних елементів  $K_i$ . Позначимо всі контрольні вершини пірамідальної мережі  $K = \bigcup_{i=1}^n K_i$ . В разі необхідності поняття може бути представлено у вигляді логічного виразу.

Нехай є пірамідальна мережа, що представляє всі об'єкти навчаючої вибірки  $L$ . Для формування понять  $B_1, B_2, \dots, B_n$ , що відповідають множинам  $V_1, V_2, \dots, V_n$  послідовно проглядаються піраміди всіх об'єктів навчаючої вибірки. Вершини піраміди, що переглядається, під час її перегляду вважаються збудженими. Введено наступні характеристики вершин мережі:  $k^a$  - кількість рецепторів в піраміді вершини  $a$  (для рецепторів  $k=1$ ), тобто  $k^a = \text{Card}(R_a)$ ;  $\vec{m}^a = (m_1^a, m_2^a, \dots, m_s^a)$ , де  $m_i^a$  ( $i=1, 2, \dots, s$ ) - кількість об'єктів поняття  $B_i$ , до пірамід яких входить вершина  $a$ . При перегляді піраміди здійснюються перетворення, які описуються наступними правилами.

#### *Правило B1.*

Якщо в піраміді об'єкту  $X$  з об'єму поняття  $B_i$  вершина  $a$ , що має найбільше  $k$  з усіх вершин піраміди  $X$  з найбільшим  $m_i$ , не є контрольною вершиною поняття  $B_i$ , то вона відмічається як контрольна вершина поняття  $B_i$ . Формулювання правила враховує можливість існування серед збуджених вершин декількох вершин з однаковим  $m_i$ , таким, що перевищує  $m_i$  всіх інших збуджених вершин. Якщо в групі вершин з найбільшим  $m_i$  декілька вершин мають найбільше значення  $k$ , то в якості контрольної відмічається будь яка з них.

Дія правила B1 ілюструється рис.2.26. У ситуації, показаній на рисунку 9.4, при збудженні піраміди вершини 2 як контрольна вершина вибирається

вершина 6, оскільки вона має найбільше  $k$  зі всіх вершин, що мають найбільше  $m_i$  (6,13,14). Значення  $m_i$  наведені усередині символів вершин.

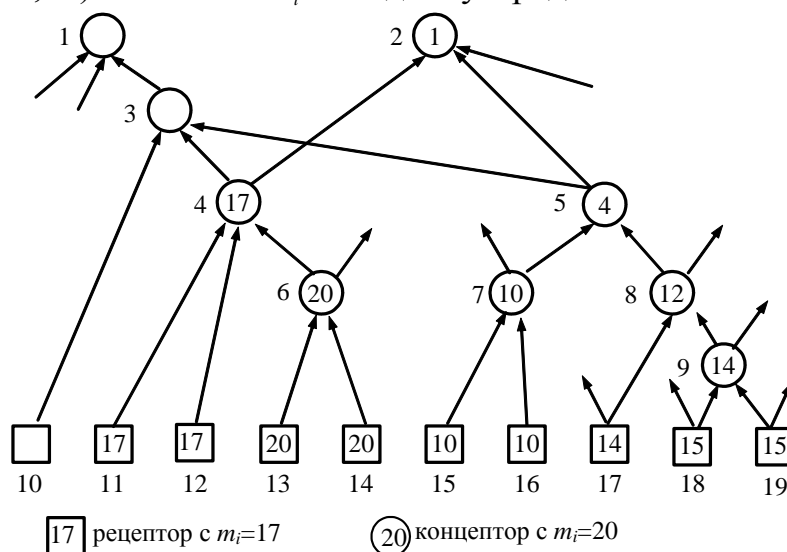


Рис. 9.4. Правило В1 вибору контрольних вершин пірамідальної мережі

### Правило В2.

Якщо в піраміді об'єкту  $X$  з об'єму поняття  $V_i$  є контрольні вершини інших понять, що не містять в своїх супермножинах збуджених контрольних вершин поняття контрольна вершина поняття  $V_i$ , в кожному з цих супермножин вершина, що має  $k$  найбільше з усіх збуджених вершин з найбільшим  $m_i$ , відзначається як контрольна вершина поняття  $V_i$ .

Якщо при перегляді всіх об'єктів навчаючої вибірки з'явилася хоча б одна нова контрольна вершина, то робиться новий перегляд всіх об'єктів навчаючої вибірки. Процес завершується, коли під час чергового перегляду навчаючої вибірки не виникло жодної нової контрольної вершини.

Відповідно до правила В2 збудження піраміди вершини 2 (рисунок 9.5, I) за умови, що вона представляє об'єкт з об'єму поняття  $V_i$ , приводить до виділення як контрольної вершини поняття  $V_i$  вершини 5 (рис. 2.27, II).

За допомогою контрольних вершин здійснюється виділення найбільш характерних (що мають найбільше  $m_i$ ) поєднань значень ознак, що належать об'єктам з об'єму поняття. Наприклад, виділення вершини 8 (рисунок 9.5) як контрольної вершину означає виділення поєднання значень ознак, відповідних рецепторам 17, 18, 19.

Якщо під час перегляду всіх об'єктів навчальної вибірки з'явилася хоча б одна нова контрольна вершина, здійснюється новий перегляд всіх об'єктів навчальної вибірки. Робота алгоритму закінчується, якщо при черговому перегляді навчальної вибірки не виникає ні однієї нової контрольної вершини. Сформоване поняття представляється в мережі ансамблем контрольних вершин.

Описаний метод забезпечує вирішення аналітичних проблем класифікації, діагностики та прогнозування на основі багатопараметричних моделей класів об'єктів. Модель відображає залежності досліджуваного класу

від сполучень значень ознак, тобто дає можливість врахувати ефект спільного впливу декількох ознак.

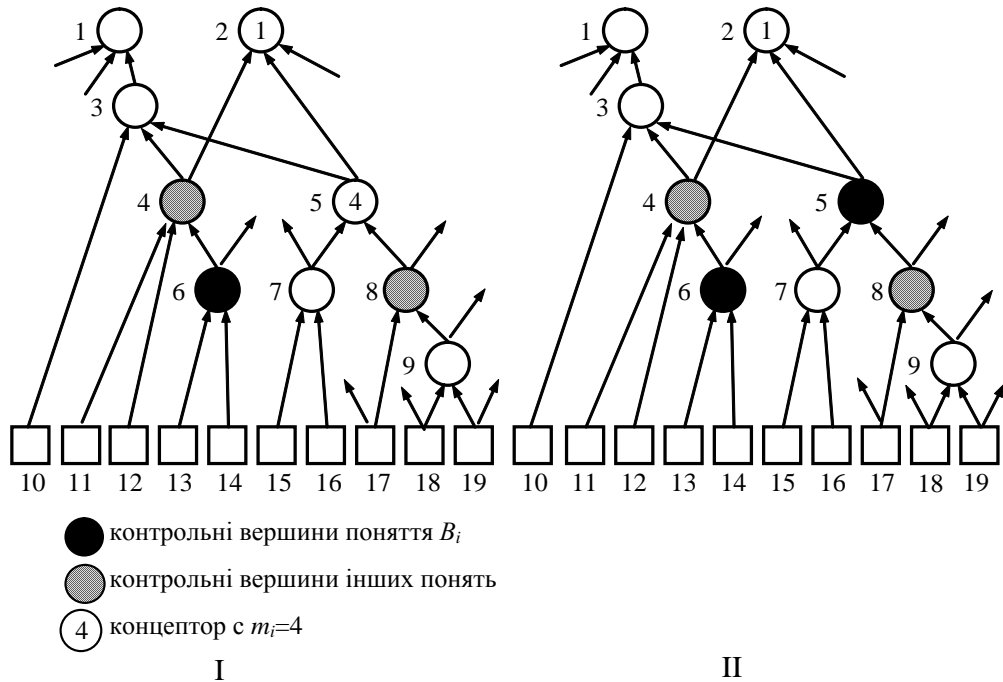


Рис. 9.5. Правило V2 вибору контрольних вершин пірамідальної мережі

На основі аналізу мережі можна побудувати поняття у формі логічного виразу. Наприклад, поняття, представлене на рисунку 9.5 заштрихованими контрольними вершинами, описується наступним виразом, де  $\vee$ ,  $\wedge$ ,  $\neg$  - логічні операції диз'юнкції, кон'юнкції та заперечення:

$$(11 \wedge 12) \wedge \neg (13 \wedge 14) \vee (17 \wedge 18 \wedge 19).$$

Важливою особливістю методу формування понять у пірамідальних мережах є можливість введення в поняття так званих виключних ознак, що не належать об'єктам досліджуваного класу. У результаті сформовані поняття мають більш компакту логічну структуру, що в принципі дає можливість підвищувати точність діагнозу чи прогнозу. У логічному виразі виключні ознаки представлені змінними з запереченням.

Після того, як поняття для деякого класу об'єктів сформовано, проблеми прогнозування або діагностики зводяться до проблеми класифікації опису прогнозованого об'єкту (стану, ситуації, процесу, події). Класифікація нових об'єктів виконується шляхом порівняння їх ознакових описів з поняттям, що характеризує клас прогнозованих об'єктів. Для цього використовується наступне правило розпізнавання.

*Розпізнавання на основі сформованих понять в семантичних мережах пірамідальної структури*

*Правило розпізнавання.* Об'єкт належить класу  $V_i$ , якщо в його піраміді є контрольні вершини поняття  $V_i$  немає жодної контрольної вершини будь-

якого іншого поняття такої, що не має збуджених контрольних вершин поняття  $B_i$  у своїй супермножині.

Якщо ця умова не виконується ні для одного з понять, об'єкт вважається невизначеним.

Крім цього, об'єкти можна класифікувати, обчислюючи значення логічних виразів, що представляють відповідні поняття. Змінним, які відповідають значенням ознак, що зустрічаються в описі об'єкту, який треба розпізнати, надається значення 1, іншим змінним – значення 0. Одиначне значення усього виразу означає, що об'єкт належить класу, що описується логічним виразом.

Доведено, що час виконання алгоритму формування понять завжди скінчений. Після виконання алгоритму вищеописане правило класифікації цілком розділяє навчаючу вибірку на підмножини  $L \cap V_i \neq \emptyset$  ( $i=1,2,\dots,n$ ).

Усі пошукові операції пірамідальної мережі обмежуються порівняно малою ділянкою мережі, що включає піраміду об'єкта і вершини, безпосередньо зв'язані з нею. У результаті з'являється принципова можливість вирішувати практичні аналітичні проблеми на основі великих об'ємів даних.

Існує аналогія між основними процесами, що мають місце в зростаючих пірамідальних мережах і нейронних мережах. Вирішальною перевагою зростаючої пірамідальної мережі є той факт, що її структура формується цілком автоматично у залежності від даних, що вводяться. У результаті досягається оптимізація представлення інформації за рахунок адаптації структури мережі до структурних особливостей даних. Причому, на відміну від нейронних мереж, ефект адаптації досягається без введення апріорної надмірності мережі. Процес навчання не залежить від визначеної конфігурації мережі. Виділені за допомогою зростаючих пірамідальних мереж узагальнені знання можуть бути явно представлені у вигляді правил чи понять, що неможливо зробити за допомогою нейронних мереж. Принципове рішення проблеми скорочення часу пошуку у ЗПМ полягає в перенесенні дій по встановленню подібності об'єктів в операції введення інформації в пам'ять. При цьому загальний вигравш в ефективності досягається за рахунок зменшення застосування операцій введення в порівнянні з пошуковими операціями. Застосування ЗПМ в різних областях науки і техніки підтвердило їх репутацію ефективного засобу структуризації великих об'ємів даних.

### *Реалізація зростаючих пірамідальних мереж в системі “Конфор”*

З метою виділення закономірностей в системі “Конфор” використовується наведений вище оригінальний метод індуктивного формування понять (CONcept FORmation) на основі пірамідальної мережі.

Система “Конфор” призначена для вирішення завдань виділення закономірностей, класифікації, діагностики і прогнозування. Названі завдання



складають ядро технології інтелектуального аналізу даних (Data Mining and Knowledge Discovery in Data Bases).

Основні функції, що виконуються за допомогою “Конфор”:

– виділення закономірностей (знань), що характеризують класи об'єктів або ситуацій;

– використання виділених закономірностей в цілях класифікації, діагностики, прогнозування.

Основні блоки системи “Конфор” й взаємозв'язок між ними представлені на рисунку 9.6.

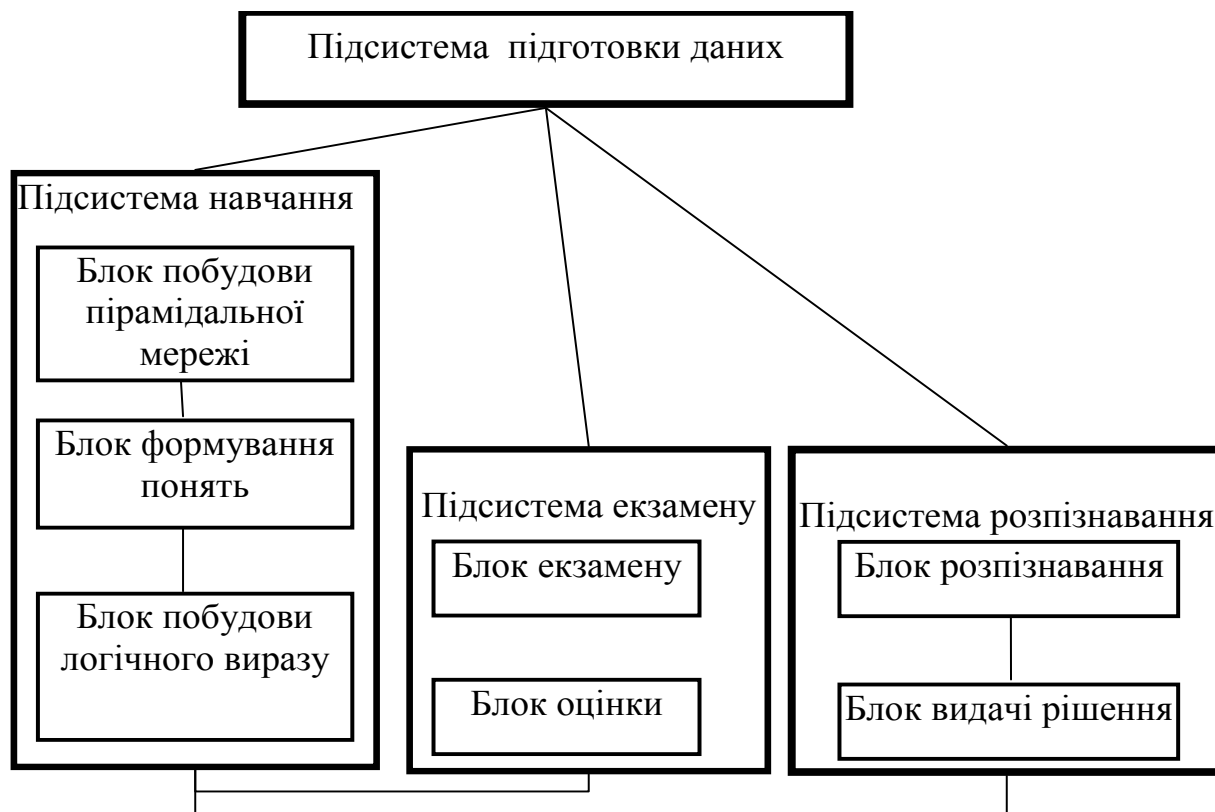


Рис. 9.6. Основні блоки системи “Конфор”

Система “Конфор” включає наступні підсистеми: підготовки даних, навчання, екзамену, розпізнавання.

1. Підсистема підготовки даних призначена для формування внутрішнього подання ознакових описів ситуацій, які використовуються у якості вихідних даних у підсистемах навчання, іспиту й розпізнавання.

2. Підсистема навчання, що здійснює виділення закономірностей (знань), які характеризують класи об'єктів.

У режимі навчання початковими даними для системи служать об'єкти вибірки для навчання, що належать як досліджуваному, так і іншим класам. Об'єкти вибірки для навчання задаються ознаковими описами. Ознаковий опис об'єкту повинен включати ім'я об'єкту, клас, якому об'єкт належить, і набір значень ознак, які характеризують об'єкт. Кожен клас об'єктів повинен бути представлений деякою кількістю прикладів, достатньої для того, щоб на

підставі їхнього аналізу була виділена закономірність, яка характеризує даний клас.

Процес навчання складається з наступних стадій:

- надання вхідних даних у вигляді пірамідальної мережі;
- формування понять на основі пірамідальної мережі;
- побудови логічних виразів, що відповідають сформованим поняттям.

Результатом роботи “Конфор”а в режимі навчання є узагальнена модель досліджуваного класу об’єктів, яка включає найбільш характерні властивості досліджуваних об’єктів.

Закономірність представляється у вигляді логічного виразу в термінах значень ознак, якими описуються початкові об’єкти. Таким чином, опис закономірності є наочним і таким, що легко інтерпретується для користувача. “Конфор” здатний виділяти скільки завгодно складну закономірність.

3. *Підсистема іспиту.* Вона призначена для тестування ступеня навченості системи “Конфор”.

Вхідними даними для підсистеми іспиту слугує екзаменаційна вибірка, що включає приклади об’єктів, які не ввійшли в навчальну вибірку, але щодо яких відомо, до якого класу вони належать. Приклади об’єктів задаються ознаковими описами, що включають ім’я об’єкту, клас, до якого об’єкт належить, і набір значень ознак, які характеризують об’єкт. Кожен об’єкт описується в термінах значень ознак, які були вибрані в режимі навчання.

У режимі іспиту система аналізує описи об’єктів екзаменаційної вибірки і розпізнає їх. Результати розпізнавання представляються у вигляді виразів наступного вигляду: “Об’єкт О віднесений до класу А, належить – В”; або “Об’єкт О не віднесений до якого-небудь класу, належить – В”, де О – ім’я об’єкту, А – ім’я класу, до якого система віднесла даний об’єкт, В – ім’я класу, якому об’єкт належить насправді.

В кінці процесу система оцінює результати іспиту і видає інформацію про кількість правильно і помилково розпізнаних об’єктів. За отриманими результатами можна оцінити якість навчання системи та прийняти рішення про можливість використання виділених закономірностей для розпізнавання.

4. *Підсистема розпізнавання,* що призначена для ідентифікації класу, до якого належить новий об’єкт.

Вхідними даними для підсистеми розпізнавання є ознаки опису нових об’єктів, що включають ім’я об’єкту й набір значень ознак, які характеризують його. Вихідні дані підсистеми – її відповідь щодо класу, до якого віднесений об’єкт, який розпізнається.

Розпізнавання виконується на підставі зіставлення опису нового об’єкту й понять, сформованих на етапі навчання. У режимі розпізнавання система аналізує описи нових об’єктів і видає відповіді двох типів: – “Об’єкт О віднесений до класу А” (де О – ім’я об’єкту, А – ім’я класу) – якщо системі вдалося визначити клас, якому належить новий об’єкт; – “Об’єкт О не віднесений до якого-небудь класу” – якщо системі не вдалося розпізнати об’єкт. Подібна ситуація можлива в двох випадках: новий об’єкт взагалі не

схожий на об'єкти вибірки для навчання, або він схожий на об'єкти вибірки для навчання, що належать різним класам. У другому випадку система видає додаткову інформацію: “Об'єкт  $O$  віднесений до класу  $A$  з упевненістю  $P_1$ ”, “Об'єкт  $O$  віднесений до класу  $B$  з упевненістю  $P_2$ ”, де  $P_1+P_2=1$ .

Залежно від специфіки завдання “Конфор” може бути використаний в режимі розпізнавання для вирішення основних аналітичних проблем, таких як класифікація, діагностика або прогнозування.

Метод ефективний при роботі з великими об'ємами даних і успішно пройшов випробування при рішенні задач виведення закономірностей, класифікації, діагностики і прогнозування в таких наочних областях, як матеріалознавство, хімія, медицина, економіка, геологія, технологія, астрономія, соціологія і т. д.

Система “Конфор” не є предметно-орієнтованою і може бути використана для аналізу даних в будь-якій предметній області за умови, що дані представлені як множина описів об'єктів у вигляді набору значень ознак.

## 9.2. Таксономічне відображення

Розглянемо процес формування онтологічного опису з застосування функцій системи Конфор на основі представлення таксономії системи знань, пов'язаної з дослідженням і аналізом фактів життя і діяльності Т. Г. Шевченка, у вигляді індуктивно поширюючоїся пірамідальної мережі. Це дозволить більш осмислене дослідити його творчість та забезпечить виділення головних факторів і причин тих чи інших подій, пов'язаних з біографією Кобзаря, а також їх можливих наслідків, вплив творчості на сучасників і майбутні покоління.

Таксономії [6-10] в системі знань забезпечують угруповання класів об'єктів предметної області. Таксономії формуються на основі встановлення відношень між поняттями і класами «частина – ціле». Розглянемо таксономії, вершини яких утворюються концептами, з яких в рамках обмежень решітки нечітких понять завжди можна сформулювати тавтологію [3]. Для цього на основі визначення кластера понять для решітки нечітких понять сформулюємо наступне правило формування тавтологій з концептів досліджуваних предметних областей: < концепт  $S_i$  >, що включає в себе < концепт  $S_s$  > забезпечує формування наступних таксономій (рисунок 9.7).

Застосуємо функцію приналежності як твірну для формування тавтологій з понять таксономій. Тоді множина всіх допустимих тавтологій складається з тверджень про приналежність понять до певної таксономії.

При обробці інформаційних джерел «Тарас Шевченко в Киргизстані», «Тарас Шевченко. БІОГРАФІЧНИЙ НАРИС», «Серце моє трудне», «Поетичне слово кобзаря», «Шевченко. Біографія» засобами системи КОНСПЕКТ на основі введеного правила були виділені поняття, об'єднані в класи таксономій розглянутої предметної області. Причому з концептів сформованих таксономій можуть бути сформульовані тавтології у вигляді

тверджень про події життєдіяльності. З множини всіх таксономій виділимо наступні[11-18]:

**Тарас Григорович Шевченко** (місця перебування в іпостасі).

- «Людина»
- «Поет»
- «Письменник»
- «Митець»
- «Громадський/Політичний діяч»

**Маршрути подорожей**

- «Перша подорож Україною»
- «Друга подорож Україною»
- «Заслання»
- «Третя подорож Україною»

**Вшанування пам'яті**

- «Музеї»
- «Пам'ятники»
- «Премії»

Причому окремо взяте поняття, в свою чергу, може бути суперпоняттям для наступних понять. Наприклад, поняття «Людина» є суперпоняттям для понять «Дитинство і юність», «Кохання Тараса Шевченка», «Смерть і поховання», а поняття «Письменник» – суперпоняття для понять «Драматичні твори», «Повісті», «Археологічні нотатки», «Записи народної творчості».

Для удосконалення існуючих таксономій в середовищі системи знань надається можливість додаткового пошуку за ключовими словами за допомогою пошукових машин Інтернет. Знайдені таким чином електронні інформаційні джерела можуть бути оброблені аналогічно для побудови нових таксономій, які в свою чергу, можуть доозначити або розширити існуючі таким чином, щоб конструкція, починаючи з поняття нижнього рівня і закінчуючи поняттям верхнього рівня таксономії, утворювала логічний ланцюжок.

Відображення понять таксономії в порталі знань дозволяє об'єднати поняття місця і часу з поняттями фактів і подій у невідомій досі комбінації, під новим кутом зору.

Завдяки об'єднанню різних типів баз даних в рамках системи знань, в таксономії атрибути об'єктів можуть бути представлені не тільки в табличному вигляді, але і в текстовому, а також у вигляді тематичних гіперпосилань на розподілені в мережі інформаційні ресурси.

Інформаційні та функціональні компоненти системи знань [1, 5-8] створюють умови для кожного його користувача певного мережного

середовища, у якому забезпечується висвітлення рис особистості Шевченка як людини, як митця, як письменника, як фігури мислителя. Дослідження та виявлення умов появи, існування та розвитку його таланту і його вплив на сутнісні віхи розвитку суспільства. Надання можливості досліджувати користувачам шляхи його життєдіяльності, виявляти вплив його особистості на оточення, на історичні події минулого, на сучасність та на майбутнє.

Онтологічні описи різних сторін життєдіяльності Тараса Григоровича Шевченка, що агрегуються у середовищі певної системи знань, дозволяють органічно з'єднати різні сфери прояву його особистості від народження і до поховання. Змістовно забезпечується висвітлення впливу його генія на розвиток світової культури, представлення та відображення динамічному та статичному вигляді усіх форм взаємодії його особистості зі світом та впливу його спадщини на світ [11-16, 18,20].

Слід також зауважити, онтологічний підхід забезпечує висвітлювання такої галузі знань як шевченкознавство. Онтологічні засоби забезпечують управління структурою цієї галузі знань, відображення таксономічних особливостей різних категорій життєдіяльності митця, аналітичне дослідження сучасного відношення до творчої спадщини Тараса Григоровича Шевченка.

Функціонально система знань забезпечує використання користувачами програмно-інформаційних засобів, які забезпечують формування та використання мережних систем знань при формуванні персоніфікованих електронних площадок та віртуальних навчально-дослідницьких лабораторій з вивчення та дослідження творчої спадщини Т.Г.Шевченка. Головною відмінністю цих засобів є використання механізмів і інструментів онтологічного управління результатами пізнавально-дослідницької діяльності відвідувачів порталу для формування мережної системи знань про життєдіяльність Т.Г.Шевченка.

До складу програмно-інформаційних компонентів системи знань входять:

- інформаційні, лінгвістичні та алгоритмічні моделі формалізації та представлення знань щодо життя і творчої спадщини Т.Г.Шевченка;
- інструментальний комплекс підтримки процесів онтологічного опису предметних галузей для формування мережних систем знань;
- програми опрацювання онтологічних даних, які підтримують процес семантичного контент-управління розподіленими інформаційними ресурсами навчально-освітнього порталу «Тарас Григорович Шевченко».

Це забезпечує систематизацію віртуальних навчально-дослідницьких середовищ, у яких підтримуються процеси навчально-пізнавальної та науково-дослідної роботи учнівської молоді під керівництвом викладачів-методистів та науковців-шевченкознавців.

Будуть створені системні рішення, які дозволять відвідувачам порталу мати дистанційний доступ до реальних електронних площадок лабораторій з шевченкознавства.

Портальне рішення з інтегрованими засобами побудови комп'ютерних онтологій забезпечує (рис.1):

- відбір, накопичення, актуалізацію інформації;
- проведення онтологічного та семантичного аналізу інформації, побудову тематичних глосаріїв, тезаурусів тощо;
- контекстно-орієнтовану тематичну класифікацію та каталогізацію інформації;
- наявність засобів онтологічного управління тайпінгів, які забезпечують встановлення ієрархічних структур сутностей на всіх рівнях мовно-онтологічного опису тематики предметних областей;
- наявність засобів побудови категоріального рівня мовно-онтологічних описів тематик предметних областей на основі використання засобів онтологічного інжинірингу;
- витяг з множини текстових документів знань, релевантних до заданої предметної області, їх системно-онтологічну структурування й формально-логічне подання;
- інтеграцію онтологічних описів, як основних компонентів методології міждисциплінарного представлення інформаційних статей;
- побудова тезаурусів термінів, описи яких включені до файлових електронних колекцій текстових документів;
- автоматична побудова ієрархії термінів заданої глибини. Відображення відповідних речень тексту при виборі терміну чи групи термінів;
- автоматичне створення списків пов'язаних слів – груп термінів, які найбільш характерні для документа чи групи документів та пов'язані між собою за змістом тексту;
- створення тематичних конспектів документів як за темами, визначеними користувачем, так і за автоматично виділеними темами з можливістю автоматичного розширення заданої теми за рахунок зв'язаних тем, які можуть автоматично визначатись при аналізі документа чи задаватись користувачем заздалегідь у вигляді фрагменту онтології предметної області;
- візуалізацію автоматично побудованої мережі понять у веб-браузерах, за рахунок чого досягається кросплатформеність представлення результатів;
- редагування мережі понять: додавання, вилучення, перейменування об'єктів та зв'язків між ними, зв'язування об'єктів з зовнішніми ресурсами, визначення ступеня зв'язку між об'єктами, перегляд фрагментів мережі заданої глибини та шляхів між окремими об'єктами;
- підключення через інтерфейси користувачів до інших інформаційних систем з метою розширення їх можливостей на основі використання технології обміну даними з іншими системами.

## ОРГАНІЗАЦІЯ ІНФОРМАЦІЙНИХ РЕСУРСІВ ПОРТАЛУ.

Організація і управління інформаційними ресурсами у середовищі ПОРТАЛУ базується на понятті категорія як ключового способу організації інформації, основаному на принципах і положеннях сучасної філософії, кібернетики, інформатики і синергетики.

Наступні категорії пропонується визначити базовими для створення онтології життєдіяльності Тараса Григоровича Шевченко:

ЛЮДИНА,  
МИТЕЦЬ,  
ПИСЬМЕННИК  
ГРОМАДСЬКО-ПОЛІТИЧНИЙ ДІЯЧ  
МИСЛИТЕЛЬ  
ЧАС

Кожна категорія будується на основі включення до її структури класів понять, які є іменами процесів, подій та фактів з життєдіяльності Тараса Шевченка[27].

Семантична зв'язність категорій і класів, формується на основі використання міждисциплінарних відношень, які забезпечують структурування та актуалізацію інформації, якою цікавиться чи яку досліджує користувач системи, визначають умови їх використання таблиця 9.1.

Таблиця 9.1

### Формальні основи структурування онтологій

<b>I. Онтологічні категорії верхнього рівня</b>	
Види абстракцій	Класифікація, узагальнення, агрегація і асоціація
Принципи класифікації	Фундаментальні принципи дихотомії Арістотеля, трихотомії Пірса і решіток Лейбніца, природна класифікація
Типи класифікаційного поділу	Таксономічні, мереологічні, просторові і часові
Спадкування ознак	Множинне
<b>II. Мовно-онтологічна картина світу</b>	
А. Рівень концептів	
Концептно-ролеві відношення	Виробляють два репрезентативи примітиву, відповідно названі приєднанням і кореферентним зв'язком. На природну мову переводяться різними граматичними формами дієслів <i>мати</i> і <i>бути</i>

Семантичні відношення	Класифікаційні – рід-вид, ціле-частина, клас-елемент, вищерозташований-нижчерозташований, клас-підклас; Ознакові – об'єкт-атрибут, об'єкт-дія; Кількісні – мати міру
Лінгвістичні відношення	Гіпонімія, метонімія, синонімія і антонімія
<b>Б. Рівень примітивів</b>	
Предикативні відношення	Узагальнено-одиничного характеру, окремо-одиничного характеру і одинично-одиничного характеру
Семантично-рольові відношення	Класифікаційні – мати ім'я, бути еталоном; ознакові – мати значення атрибута, мати значення способу дії; кількісні – мати значні міри; порівняння – рівне, порівняльне, більше, більше або дорівнює, менше, менше або дорівнює, незрівняне; приналежності – простої приналежності; прості часові – бути одночасно, бути раніше, бути пізніше, збігтися в часі, перетинатися в часі, бути всередині в часі, починатися одночасно, закінчуватися у часі; прості просторові – збігатися в просторі, бути ліворуч, бути справа, бути попереду, бути позаду, бути зверху, бути знизу, бути навскоси, перетинатися в просторі, торкати, розміщуватися на, розміщуватися в; каузальні – бути метою, бути мотивом, причина-наслідок; інструментальні – служити для, бути засобом для, сприяти, бути інструментом, бути допоміжним засобом; інформаційні – бути відправником, бути одержувачем, бути джерелом інформації; порядкові – бути наступним, бути поточним, бути найближчим; модальні – можливості, дійсного втілення, необхідності; модифікаційні; квантифікаційні – загальності, існування, унікальності, одиничності, можливості, необхідності; кореляційні



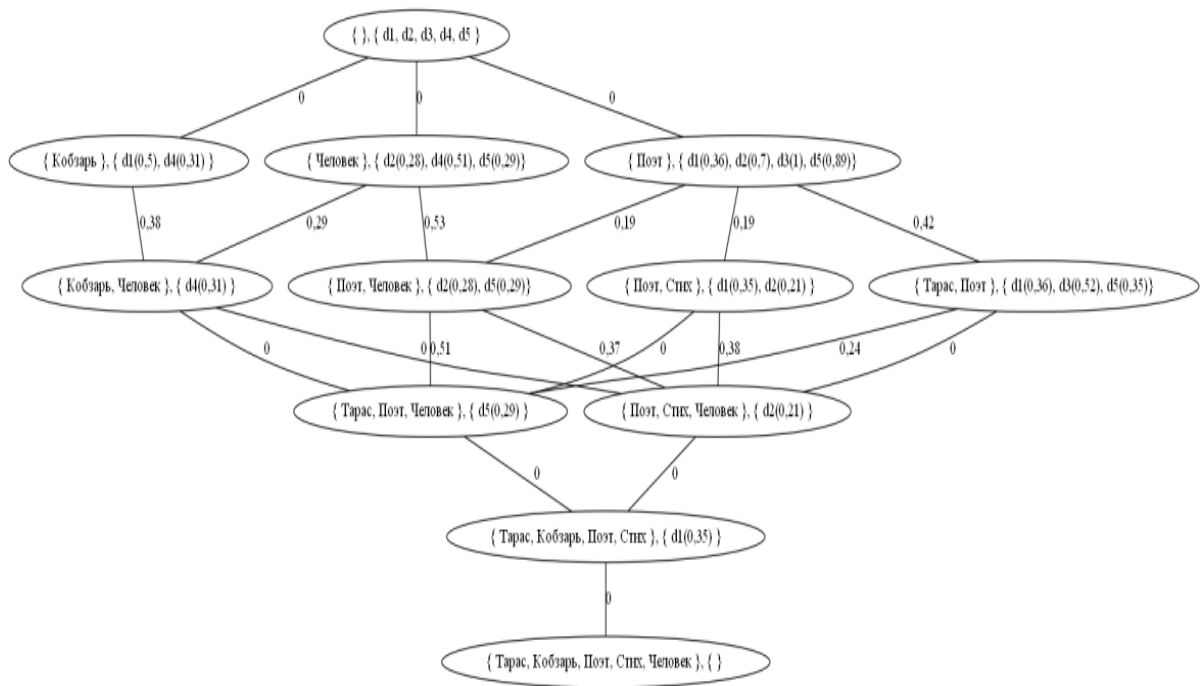


Рис. 9.7. Таксономія нечітких категорій онтологічного відображення життєдіяльності Т.Г.Шевченка

Кожна з вищезгаданих категорій включає до себе ряд класів, які сукупно відображають тематики життєдіяльності Тараса Шевченка. Приклад категорій та класів, які характеризують онтологію життєдіяльності Тараса Шевченка, представлено в таблиці 9.2.

Усі категорії та класи являють собою деяку класифікацію. Ця класифікація може бути змінена, усю ієрархію, назви та властивості кожної категорії та класу можливо відредагувати в будь-який час і за будь-яких умов. Також можливе поширення їх переліку. Тобто, онтологія життєдіяльності Тараса Шевченка має динамічний характер і забезпечує як можливе широке та досить глибоке відображення усіх подій та фактів які сталися за його життя, так і ті, що склалися та існують до сьогодні.

в таблиці 9.3 наведено приклад формування процесів, які відносяться до категорії музейної класифікації.

Інтеграція змісту та функціоналу порталу з іншими системами розширяє його можливості.

Одним з таких компонентів є геоінформаційні системи (ГІС). У середовищі ГІС існує багато можливостей більш виразно відобразити події які описуються певною онтологією. Розглянемо це на прикладах методики застосування.

Більш детально відображення комунікацій з онтологічними описами життєдіяльності Тараса Шевченка у інформаційному середовищі порталу представлено на рисунках 9.8-9.30.



Рис. 9.8. Узагальнена структура інформаційного середовища бази знань з житнеопису «Тарас Шевченко»

Таблиця 9.2

**Категорії та класи онтології життєдіяльності ШЕВЧЕНКА**

Тарас Григорович Шевченко	ЛЮДИНА	ПИСЬМЕННИК	МИТЕЦЬ	МИСЛИТЕЛЬ	ЧАС				
ЛЮДИНА	Дитинство і молодість	Викуп	В Академії мистецтв	Перша подорож Україною	Поет проги імперії	Друга подорож Україною	Арешт і заслання	Кохання Тараса Григоровича Шевченка	Смерть і поховання
ПОЕТ/ ПИСЬМЕННИК	Поезія	Драматичні твори	Повісті	Археологічні нотатки	Листи ЩОДЕННИК	Записи народної творчості			

Дослідження життя і творчості Шевченка (шевченкознавство)	Вплив творчості Т. Г. Шевченка	ЧАС	За часом	Незавершені роботи	За технікою	За жанром	Художня творчість	ГРОМАДСЬКИЙ/ ПОЛІТИЧНИЙ ДІЯЧ	МИТЕЦЬ
Документи про життя і творчість	Вплив творчості Т. Г. Шевченка на події минулого	Дослідження життя і творчості Шевченка	1830 — 1846	Етюди	Олія	Портрети	За жанром	Кирило-Мефодіївське братство	Малярська творчість
Шевченкознавці	на сучасність	Ушанування пам'яті	1847—1850	Нариси	Акварель	Композиції	За хронологією	Революціонери-декабристи	
Шевченко в образотворчому мистецтві	на майбутні покоління	Вплив творчості Т. Г. Шевченка	1851—1857	Ескізи	Сепія	Архітектурні пейзажі й краєвиди	За технікою	"Прогвіта"	
Посилання на інші сайти та матеріали			1857 — 1861		Туш		Незавершені роботи		
Шевченківська енциклопедія					Олівець				
					Офорт				

Зображення	Музеї	Ушанування пам'яті	Фільми про Шевченка	Шевченко в образотворчому мистецтві	Оцифровані першоджерела	Документи про життя і творчість
на фотографіях	3D віртуальні музеї	Пам'ятники	Художні	Музична Шевченкіана	Тексти творів	Наукові біографії
на листівках	Національний музей Тараса Шевченка	Музеї	Документальні	Фільми про Шевченка	Ілюстрація творів	Родовід
в медальористичі та нумізматиці		Премії України ім. Тараса Шевченка			Аудіокниги	Місця перебування
в філателістиці		Зображення				Опис рукописів
		Сувенірна продукція				Оцифровані першоджерела
		Форум				
		Інститут літератури ім. Т.Г. Шевченка Національної академії наук України				

Таблиця 9.3

## Приклад формування категорій

№ п/п	Назва музею	Країна	Місто	Адреса
1	Національний музей Тараса Шевченка	Україна	Київ	бул. Шевченка, 12
2	Літературно-меморіальний будинок-музей Т.Г. Шевченка	Україна	Київ	пров. Тараса Шевченка, 8а
3	Меморіальний будинок-музей Т.Г. Шевченка	Україна	Київ	вул. Вишгородська, 5
4	Музей «Флігель Тараса Шевченка»	Україна	м. Яготин, Київська обл.	вул. Незалежності, 69-а.
5	Хата на Пріорці	Україна	м. Київ	вул. Вишгородська, 5
6	Національний заповідник „Батьківщина Тараса Шевченка”	Україна	село Шевченкове Звенигородський район Черкаська область	вул. Бондарівська, 33
7	Музей-квартира	Росія	Санкт-Петербург	Університетська набережна, 17
8	Меморіальний музей-гауптвахта	Росія	Оренбург	вул. Советська, 24
9	Шевченківський національний заповідник	Україна	м. Канів, Черкаська область	
10	Тарасова світлиця		Канів, Черкаська обл.	Тарасова Гора
11	Меморіальний комплекс	Казахстан	Форт-Шевченко	25 км на схід від міста
12	Музей Т.Шевченка	Канада	Торонто	1614 Bloor St. West
13	Музей "Кобзаря"	Україна	Черкаси	вул. Хрещатик, 219
14	Орський музей Т. Шевченка	Росія	Орськ	вул. Шевченка
15	Баришівський музей Т. Г. Шевченка	Україна	Баришівка, Київська обл.	вул. Жовтнева, 5

16	Музей "Заповіту"	Україна	Переяслав-Хмельницький	вул. Шевченка, 8
17	Тарасова світлиця	Україна	Канів, Черкаська обл.	Тарасова Гора
18	Літературно-меморіальний музей	Україна	с. Шевченкове, Звенигородський р-н., Черкаська обл.	
19	Кімната-музей	Україна	Львів	вул. Коперника, 17
20	Шевченкова світлиця	Україна	м. Красно-перекопськ, АР Крим	вул. Калініна, 22, ЗОШ І-ІІІ ступенів №4,
21	Орський музей Т. Шевченка	Росія	Орськ	вул. Шевченка
22	Казахська державна художня галерея ім. Т. Шевченка	Казахстан	Алмати	мікрорайон Коктем-3, буд. 22/1

### 9.3. Методика застосування

Таксономічна структура відображення процесів життєдіяльності Тараса Шевченка, дозволяє досить ефективно застосувати у середовищі бази знань я засоби геоінформаційних систем (ГІС) [4], які забезпечують вирішення завдань, що пов'язані з дослідженням і аналізом фактів життя і діяльності Т. Г. Шевченка, з осмисленням і виділенням головних факторів і причин тих чи інших подій, пов'язаних з біографією Кобзаря, а також їх можливих наслідків, вплив творчості на сучасників і майбутні покоління. Розвиток геоінформаційних систем пов'язаний з необхідністю спільної обробки обсягів просторової і непросторової інформації, складних процесів обробки взаємозалежної різнопланової інформації, її інтеграції та взаємодії з іншими різними за призначенням системами. Додаткові вимоги до знаходження кращих рішень, зручності, змістовності, відображення міждисциплінарності інформаційних процесів які досліджуються, також вимагають розробки і розвитку адекватних методів та моделей.

Знайдені за допомогою інструментів пошукової машини електронні інформаційні джерела можуть бути оброблені засобами поверхневого лінгвістичного аналізу мережної системи КОНСПЕКТ[5]. Ці засоби забезпечують побудову різних класів таксономій, які в свою чергу, можуть бути доозначені або більш розширено визначені таким чином, щоб

конструкція з імен вершин у кожній таксономії починаючи з поняття нижнього рівня і закінчуючи поняттям верхнього рівня утворювала логічний ланцюжок. Для цього будьмо розглядати усі зв'язні вершини таксономічного графу як певні тавтології, які будуються за правилом: <концепт> – властивість-відношення – <концепт>/. Тобто таксономії, вершини яких утворюються певними, змістовно наповненими вершинами-концептами, можливо визначати як тавтології при урахуванні певних рамок обмежень інтуїтивного розуміння понять [3]. Тоді висловлювання, які формуються з усіх вершин, що мають між собою непусті відношення можуть розглядатися як множина всіх допустимих тавтологій і складатися з тверджень про приналежність концептів-понять до певної таксономії. Тавтології формуються на основі тематичного зв'язування концептів класів, що носять імена наведених вище таксономій [26].

Таксономії в середовищі ГІС, що забезпечують угруповання класів об'єктів онтології ПрО, відповідають шарам тематичної карти, а самі об'єкти, що входять до відповідного класу, – об'єктам шару. Таксономія об'єктів онтології, що відповідає легенді карти, формується на основі встановлення відносин між поняттями і класами, прикладом може бути відношення «частина – ціле».

Відображення класів таксономії у вигляді тематичних шарів карти дозволяє об'єднати поняття місця і часу з поняттями фактів і подій у невідомій до цього комбінації, під новим кутом зору.

Застосування функції приналежності дозволяє визначити класи таксономій, концепти-поняття яких можуть бути представлені тематичними шарами карти. Також концепти, безпосередньо складові таксономії, можуть знаходити своє відображення в легендах цих тематичних шарів.

Легенда карти включає в себе тематичні шари, аналогічні за назвою таксономіям, як класам понять, що мають однакові множини властивостей і об'єкти шарів. Відповідність між тематичними шарами карти і таксономіями, що відображають структуру інформаційних масивів, реалізується на основі визначення рівня схожості нечітких формальних понять досліджуваних предметних тематик.

Опис концептів предметної області у вигляді певних об'єктів на карті обмежений полями атрибутивної інформації, а сервіс вкладень дозволяє прикріпити тільки ту інформацію, яка фізично наявна у користувача. Завдяки об'єднанню різних типів баз даних, в таксономії атрибути об'єктів можуть бути представлені не тільки в табличному вигляді, але і в текстовому, а також у вигляді тематичних гіперпосилань на розподілені в мережі інформаційні ресурси.

Агрегація таксономії та тематичної карти дозволяє розширювати і доповнювати інформаційні описи понять-об'єктів (концептів) на основі мережевих інформаційних ресурсів, розподілених в Internet, а вбудована в онтологію пошукова машина – значно розширити уявлення про предметні області і міждисциплінарні зв'язки між предметними областями. Таке поєднання дозволяє створити єдине інформаційно-аналітичне середовище

експерта, яке перманентно поповнюється доробками територіально розподілених користувачів різних напрямків досліджень.

У середовищі ГІС в легенді карти символи об'єктів (шаблони) тематичних шарів залишаються незмінними, а, отже, дізнатися, що об'єкт належить до множини об'єктів N, ми можемо лише переглянувши атрибутивну інформацію. Таксономія може гнучко доповнюватися новими класами і поняттями, зберігаючи наочність сприйняття користувачем взаємозв'язків між елементами. Наприклад, всі музеї на карті мають однаковий символ, його назва міститься в атрибутивній таблиці. З таксономії онтології життєдіяльності Тараса Шевченка ми можемо бачити, що конкретний об'єкт «Літературно-меморіальний будинок-музей Т. Г. Шевченка» є музеєм і входить до класу об'єктів вшанування пам'яті, відповідний тематичному шару карти, присвяченої Т. Г. Шевченку.

При обробці інформаційних джерел «Тарас Шевченко в Киргизстані», «Тарас Шевченко. БІОГРАФІЧНИЙ НАРИС», «Серце моє трудне», «Поетичне слово кобзаря», «Шевченко. Біографія» засобами системи КОНСПЕКТ на основі введеного правила були виділені поняття, об'єднані в класи таксономій розглянутої предметної області. Причому з концептів сформованих таксономій можуть бути сформульовані тавтології у вигляді тверджень про події життєдіяльності. З множини всіх таксономій виділимо наступні:

**Тарас Григорович Шевченко** (місця перебування в іпостасі).

Людина

Поет

Письменник

Митець

Громадський/Політичний діяч

**Маршрути подорожей**

Перша подорож Україною

Друга подорож Україною

Заслання

Третя подорож Україною

**Вшанування пам'яті**

Музеї

Пам'ятники

Премії

Причому окремо взяте поняття, в свою чергу, може бути класом для наступних понять. Наприклад, поняття «Людина» є класом для понять «Дитинство і юність», «Кохання Тараса Шевченка», «Смерть і поховання», а поняття «Письменник» – клас для понять «Драматичні твори», «Повісті», «Археологічні нотатки», «Записи народної творчості».

Завдяки використанню пошукової машини, контекстне сприйняття кожного поняття логічно доповнюється семантично пов'язаною з ним інформацією, а відображення відповідного поняття як об'єкта тематичного



шару на карті в ГІС- середовищі дозволяє побачити загальну картину, охоплюючи категорії часу і географічної прив'язки. Наприклад, клас «Кохання Тараса Шевченка» включає поняття «Варвара Репіна». За допомогою використання інструментів пошукової машини дізнаємося, що в період з січня по березень 1845 року (час) Варвара Репіна вела переписку з Т. Г. Шевченко з Яготина (географічна прив'язка).

За результатами пошуку семантично пов'язаних понять з класом «Заслання» можемо виділити нові класи понять і об'єкти тематичних шарів (виділені курсивом): «Другой раз я проезжаю мимо *Симбирска*, и другой раз не удастся видеть мне монумент придворного историографа. Первый раз в 1847 меня провез фельдъегерь мимо *Симбирска*. Тогда было не до монумента *Карамзина*. Тогда я едва успел пообедать в какой-то харчевне, или, вернее сказать, в кабаке. Во мне была (как я после узнал) экстренная надобность в *Оренбурге*, и потому-то фельдъегерь неудобозабываемого *Тормоза (Николай I. — Ред.)* не дремал. Он меня из *Питера* на восьмые сутки доставил в *Оренбург*, убивши только одну почтовую лошадь на всем пространстве». Таким чином, клас «Заслання» доповнюється поняттями географічної прив'язки «Симбірськ», «Оренбург», «Пітер» і часу – «Епоха Миколи I», «1847».

Клас «Митець» доповнюється поняттям «Художня творчість», яке включає поняття «Олія», «Акварель», «Туш», «Олівець» тощо. Кожне з цих понять є класом для понять-назв художніх творів. Наприклад, в 1841 році (час) в Санкт-Петербурзі (географічна прив'язка) була опублікована збірка «Наши, списанные с натуры русскими людьми» з *ілюстрацією* («Туш») Т. Г. Шевченка до твору Г. Квітки-Основ'яненка «Знахар», про яку М. Некрасов написав: «Настал иллюстрированный В литературе век. С тех пор, как шутка с «Нашими» Пошла и удалась, Тьма книг с политипажами В столице развелась» («Громадський/Політичний діяч»). Таким чином, складається логічна таксономія «Ілюстрація до «Знахар» Г. Квітки-Основ'яненка» → «Туш» → «Художня творчість» → «Період перебування в Академії мистецтв у Петербурзі» → «Людина»/ «Митець»/ «Громадський/ Політичний діяч» → «Тарас Григорович Шевченко», яку можна прочитати як «Ілюстрація до твору Г. Квітки-Основ'яненка «Знахар», виконана тушшю і віднесена до художньої творчості періоду навчання в Петербурзькій Академії мистецтв, розцінювалася сучасниками як політична діяльність Т. Г. Шевченка.

Таким чином онтологічні описи різних сторін життєдіяльності і творчості Тараса Григоровича Шевченка ефективно забезпечують певні процеси вивчення його спадщини, дозволяють асоціативно переглядати досить велику кількість різномірних документів і поєднувати їх за семантичною змістовністю.

Це забезпечує дослідників його спадщини, шевченкознавців і усіх небайдужих до таланту митця новітніми інструментами дослідження, розглядаючи життєдіяльність Шевченка як визначну систему знань [28].

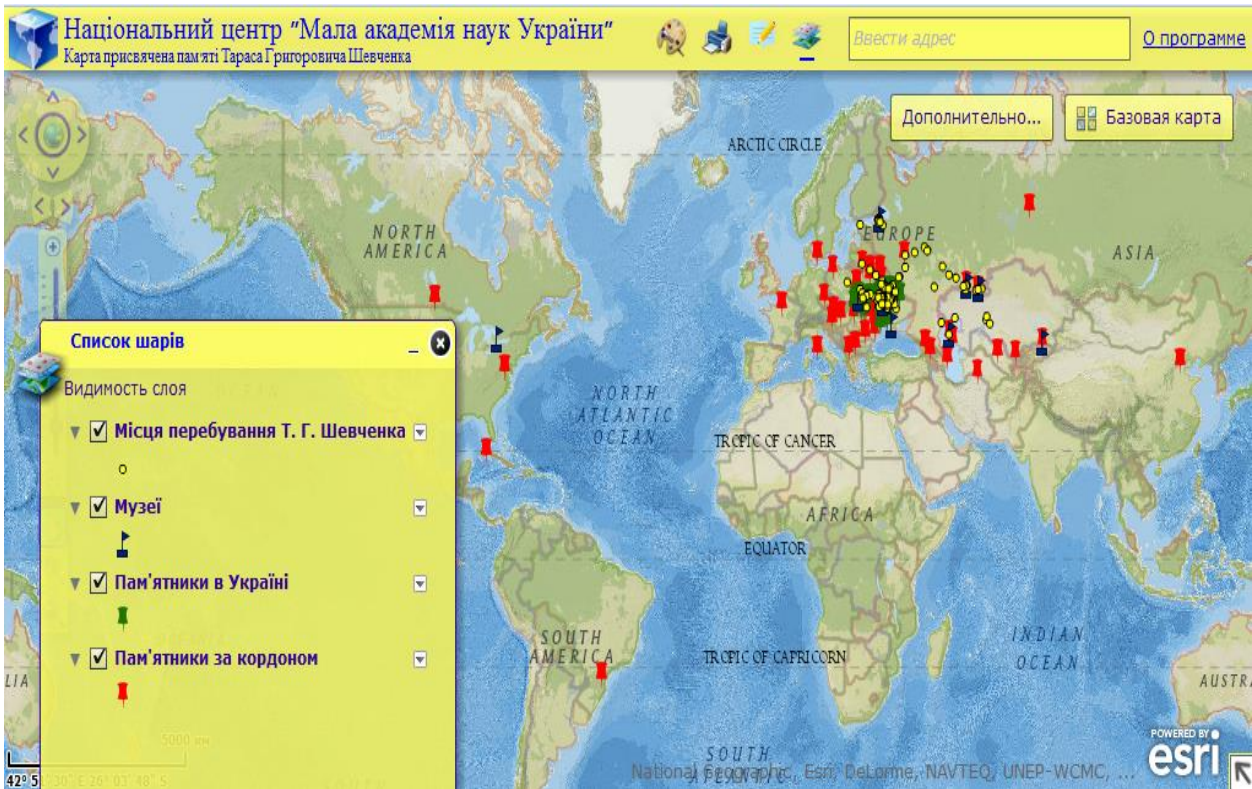


Рис. 9.9. Тематичні шари на карті, присвяченій життю та творчості Т.Г.Шевченка

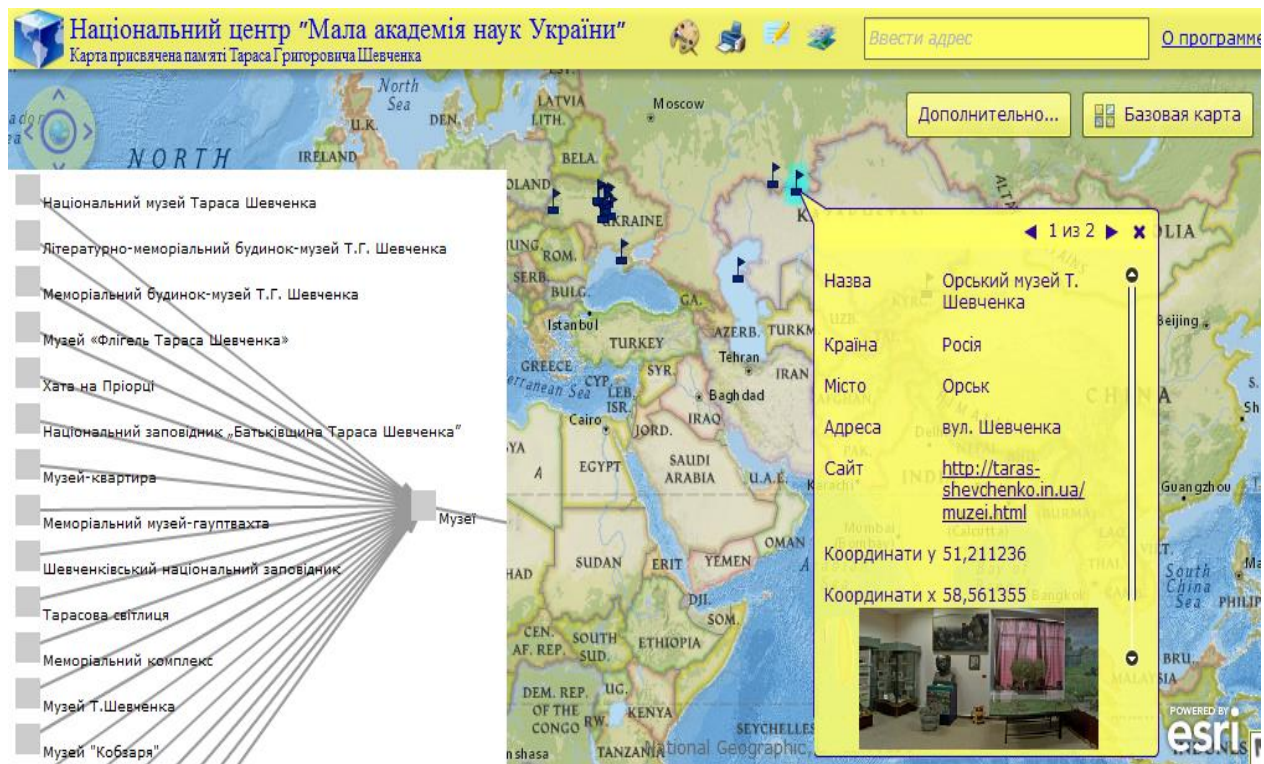


Рис. 9.10. Тематичний шар «Музеї», що відповідає однойменному класу

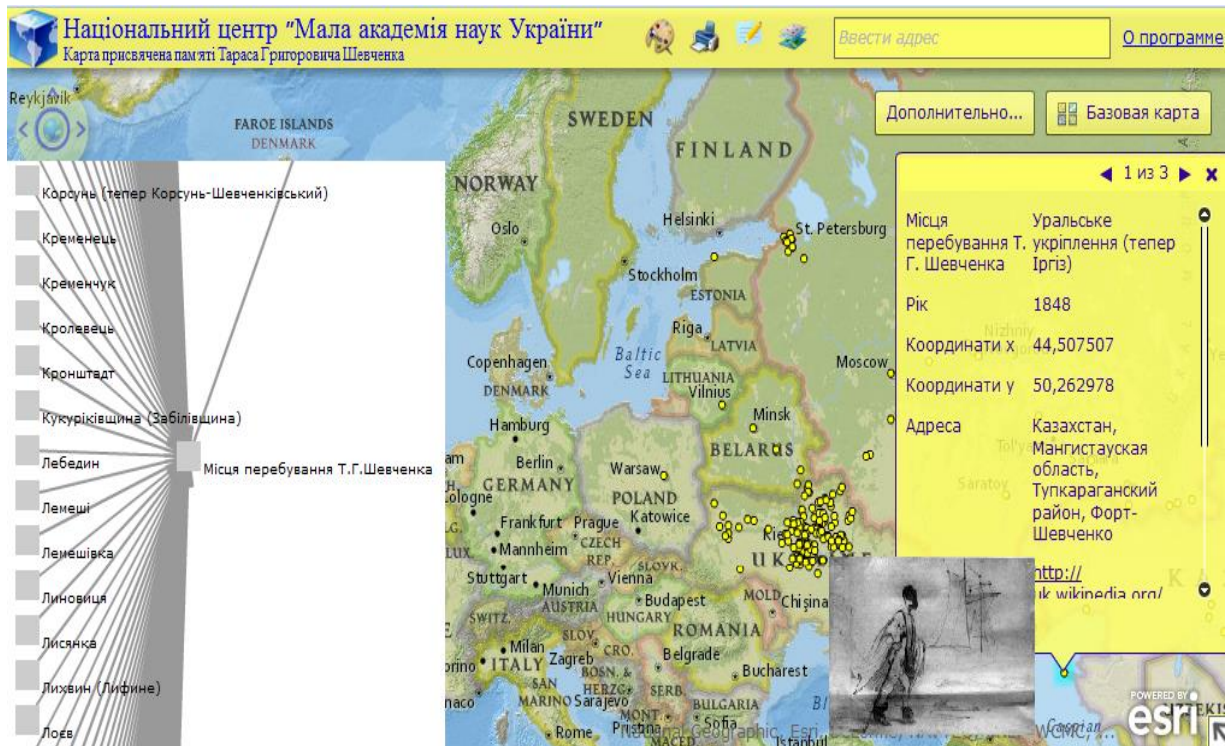


Рис. 9.11. Тематичний шар «Місця перебування Т.Г.Шевченка», що відповідає однойменному класу понять таксономії

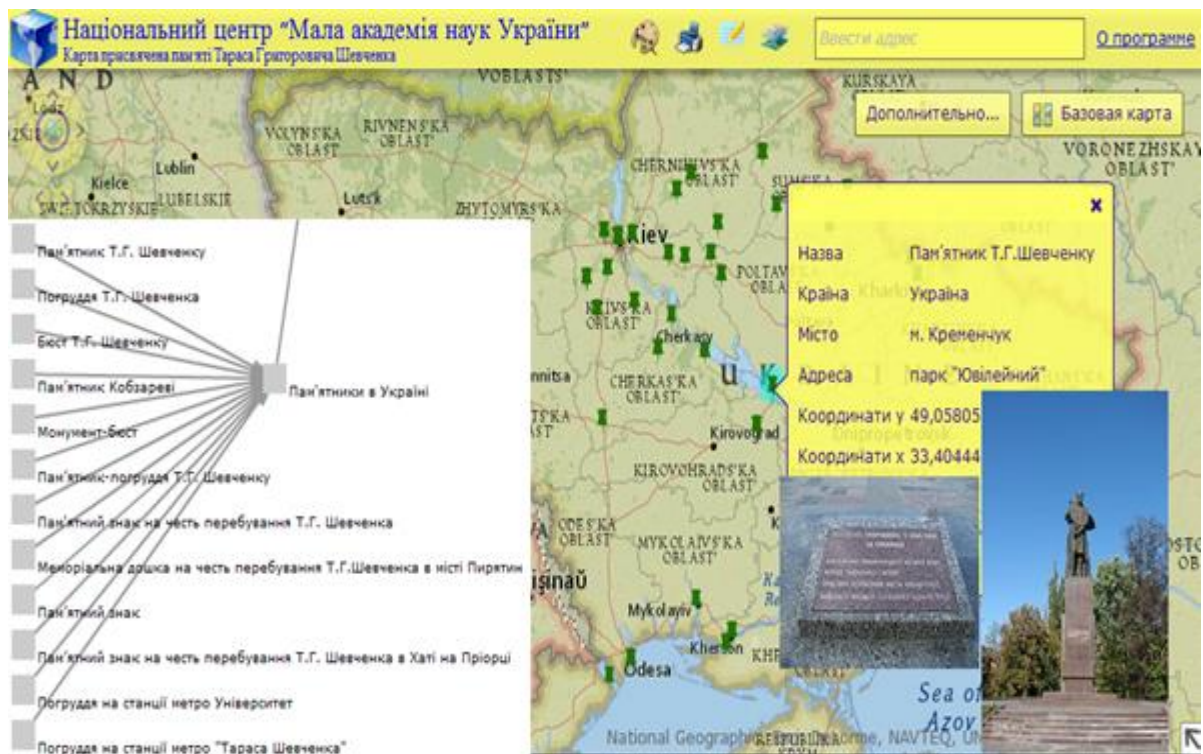


Рис. 9.12. Тематичний шар «Пам'ятники в Україні», що відповідає однойменному класу понять таксономії

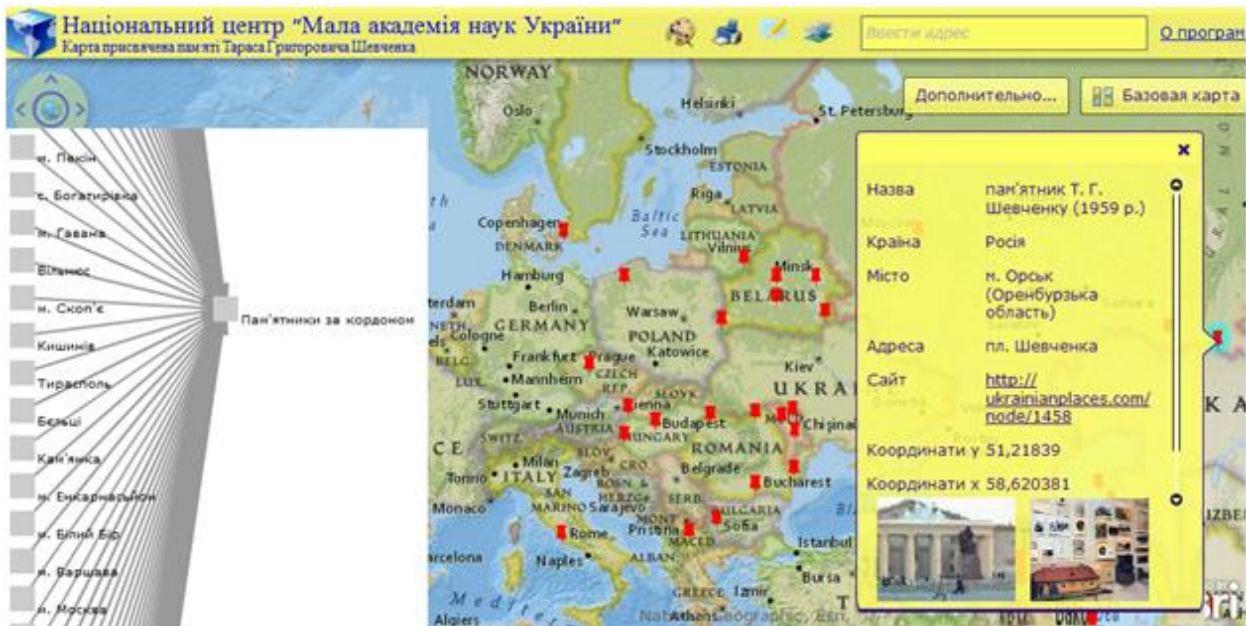


Рис. 9.13. Тематичний шар «Пам'ятники за кордоном», що відповідає однойменному класу понять таксономії

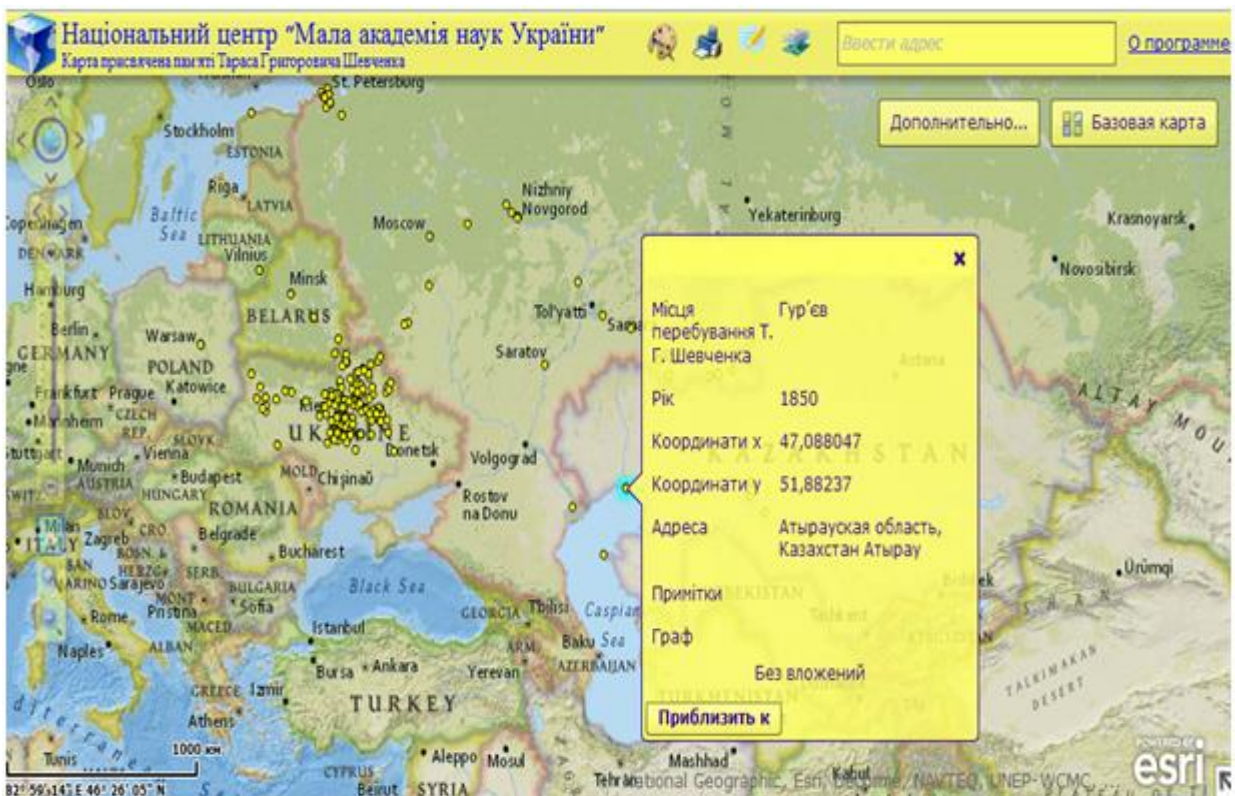


Рис. 9.14. Атрибутивна інформація об'єкту шару «Місця перебування Т.Г.Шевченка»

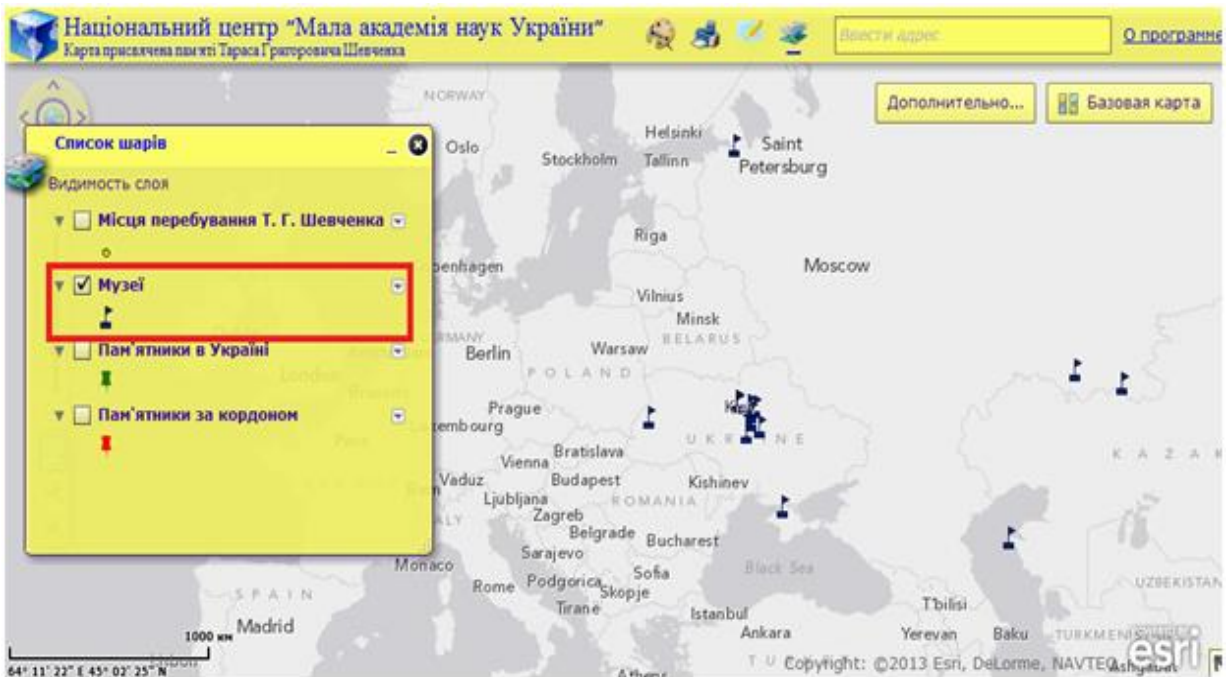


Рис. 9.15. Об'єкти шару «Музеї»

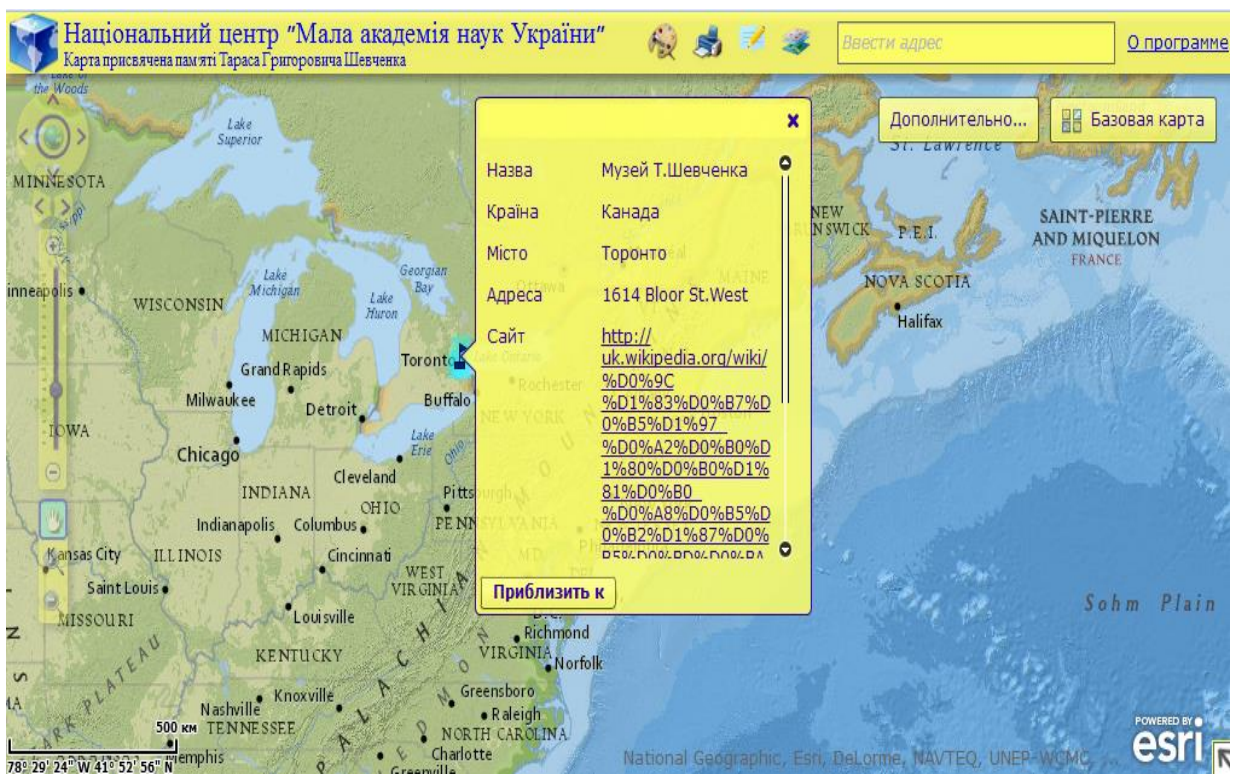


Рис. 9.16. Атрибутивна інформація об'єкту шару «Музеї»

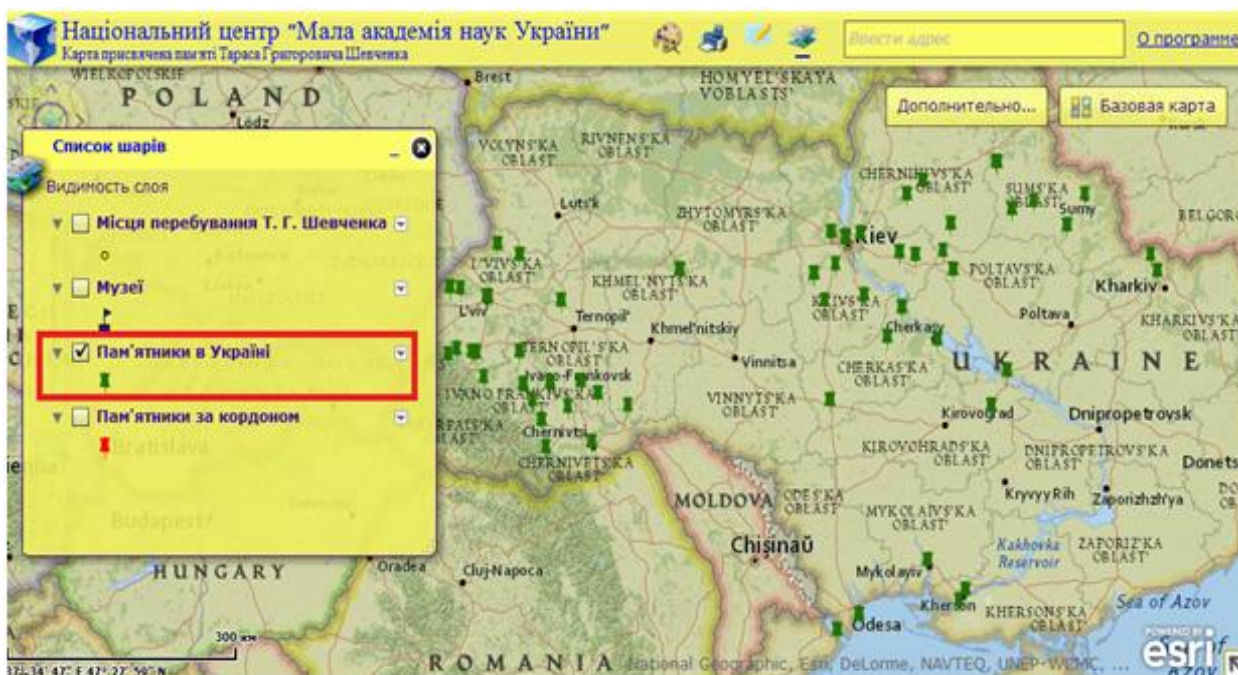


Рис. 9.17. Об'єкти шару «Пам'ятники в Україні»

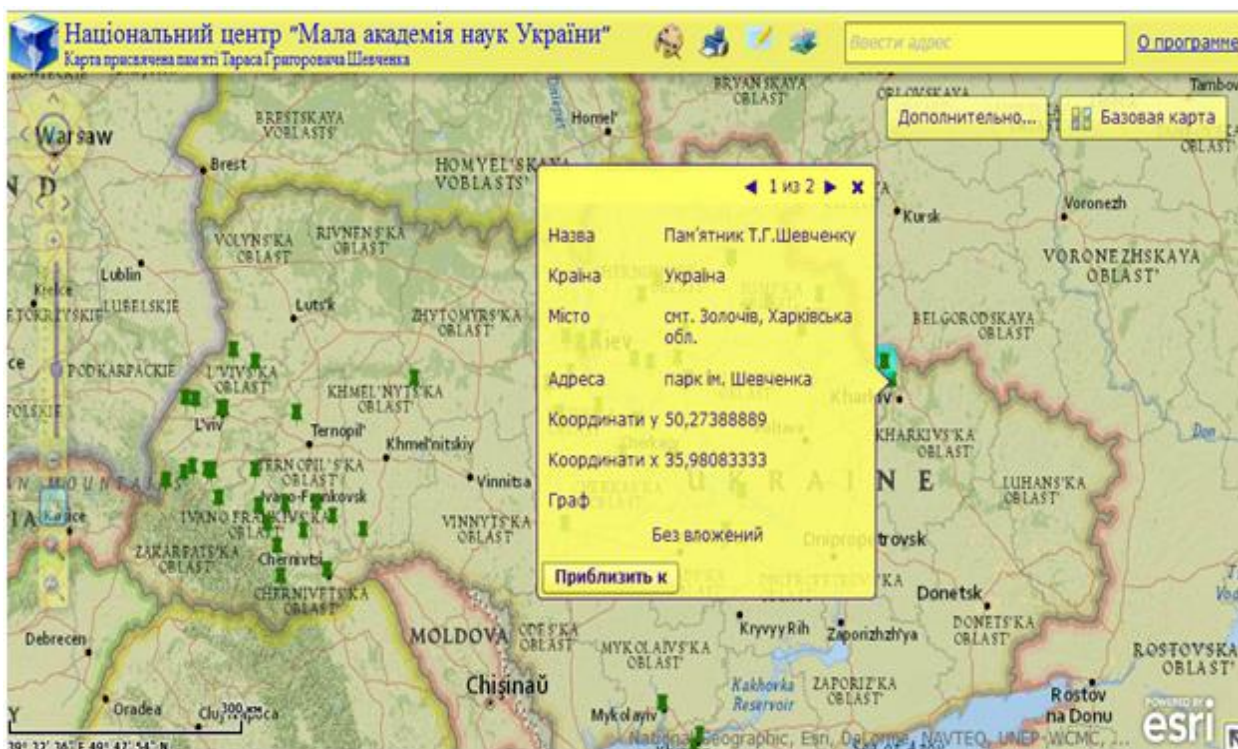


Рис. 9.18. Атрибутивна інформація об'єкту шару «Пам'ятники в Україні»

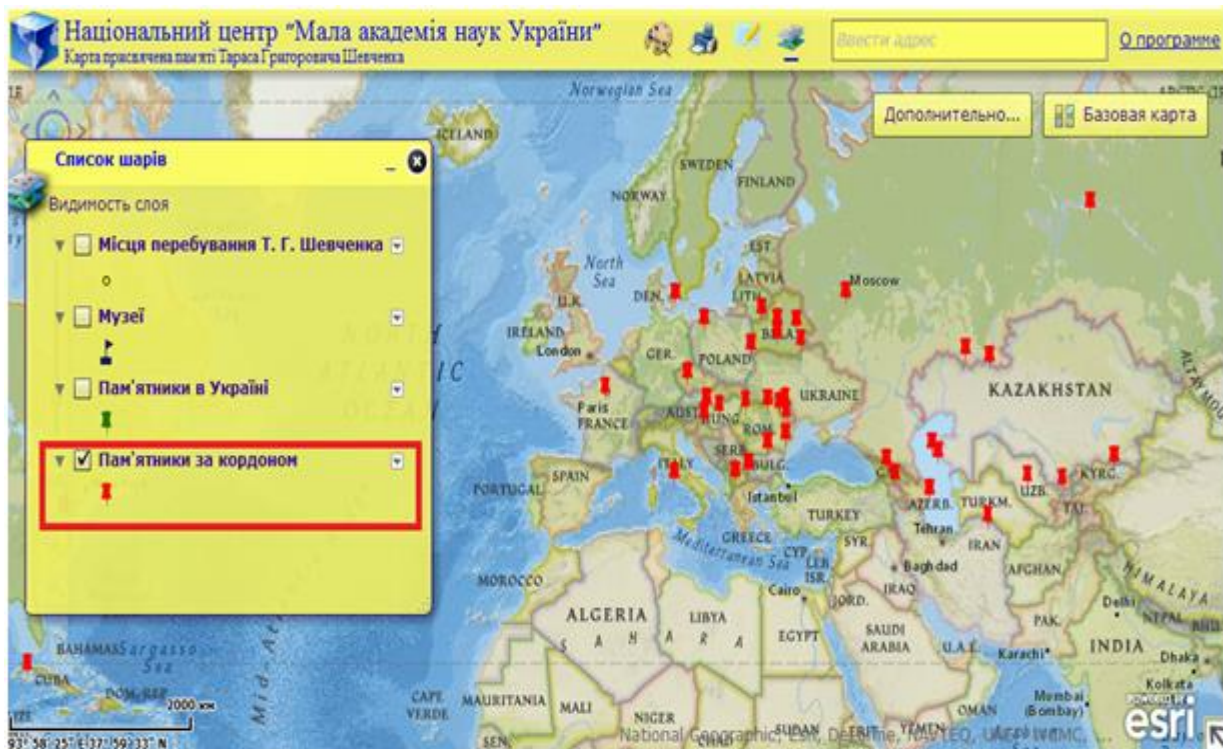


Рис. 9.19. Об'єкти шару «Пам'ятники за кордоном»

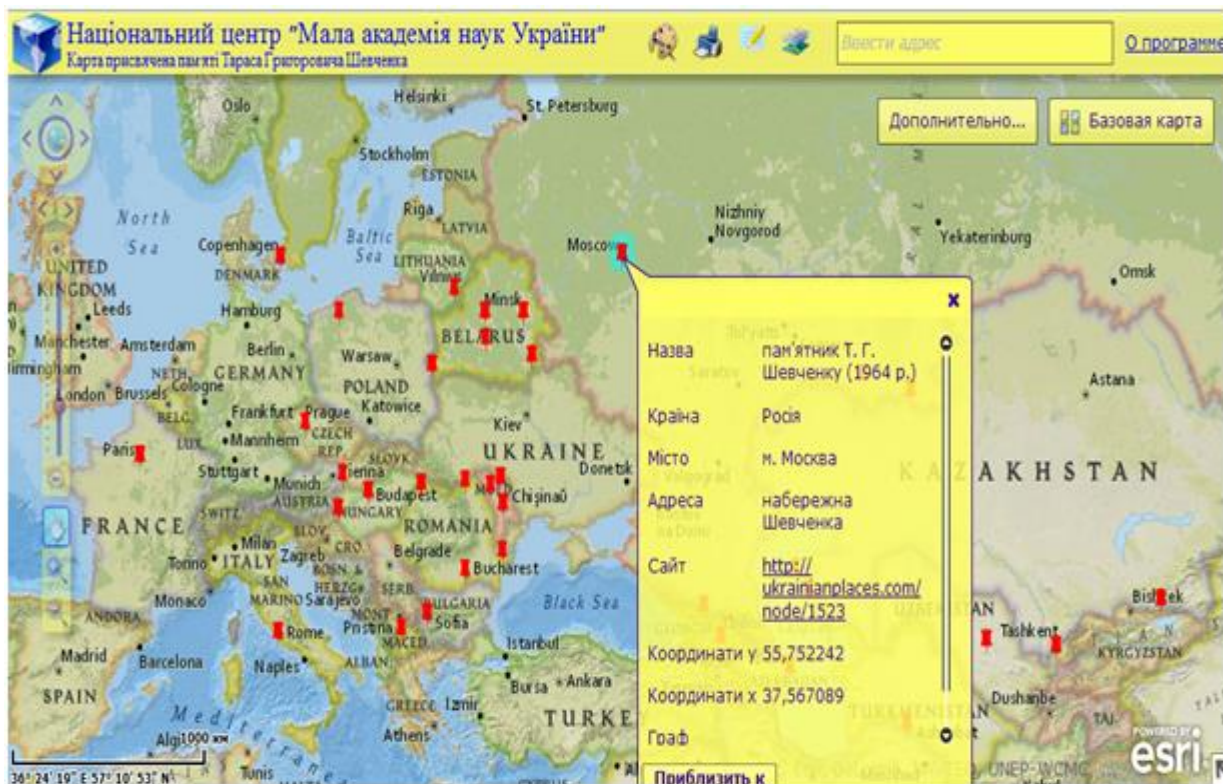


Рис. 9.20. Атрибутивна інформація об'єкту шару «Пам'ятники за кордоном»

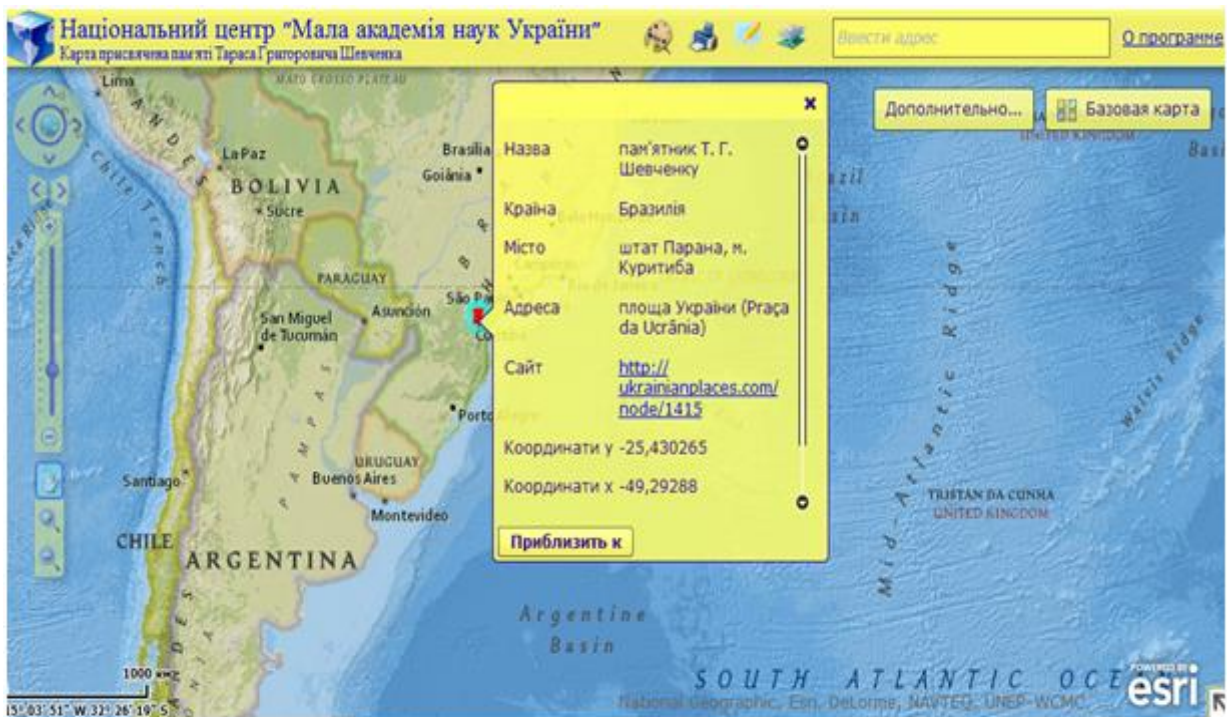


Рис. 9.21. Атрибутивна інформація об'єкту шару «Пам'ятники за кордоном»

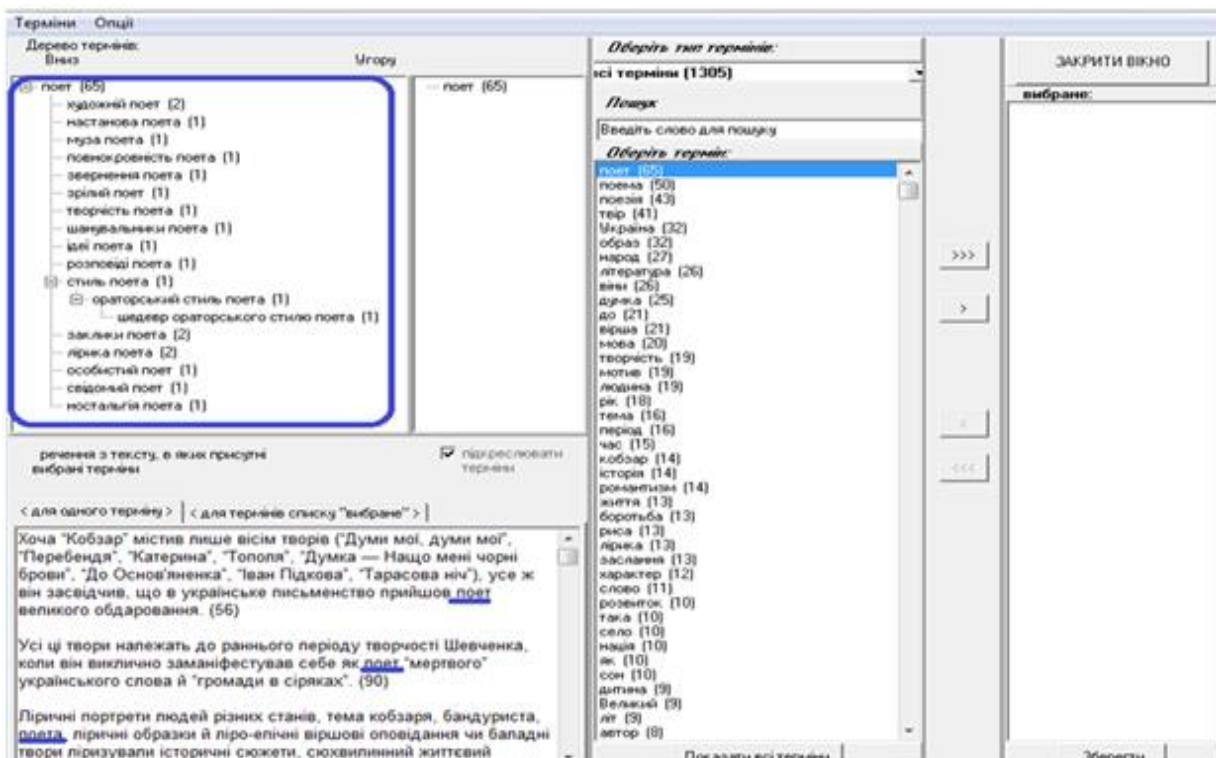


Рис. 9.22. Фрагмент множини термінологічних дерев опису творчості



The image shows a complex semantic network on the left side of a portal interface. The network is centered around two main nodes: 'ПОЕТ/ПИСЬМЕННИК' (Poet/Writer) and 'ГОМАДСЬКИЙ/ПОЛІТИЧНИЙ ДІЯЧ' (Civic/Political Figure). Various other nodes are connected to these, including 'Плакати', 'Археологічні знахідки', 'Листи', 'Залишки народної творчості', 'Кирило-Мефодієвське братство', 'Ремісничий/виробничий', 'Проста', 'на сучасність', 'на майбутнє покоління', 'Портрети', 'Композиції', 'Архітектурні пейзажі й композиції', and 'Факт'.

On the right side, there is a search results page for 'Наймичка' (The Maid) by Taras Shevchenko. The page shows the search results for 'Наймичка' with 236 results. The first result is a preview of the poem, including the text: 'А наймичка невислуха Шренєвир, небога, Свою долю прокигнає. Тяжко-важко плаче...'. The page also includes a 'Narrow your search' section with filters for Source, Document type, File extension, and Last modification.

Рис. 9.23. Семантично зв'язані між собою інформаційні компоненти порталу

This diagram illustrates a semantic network centered on 'НумізMATика' (Numismatics) and 'Зображення' (Images). The 'НумізMATика' node is connected to 'Монета, 5 копійок', 'Монета, 10 копійок', 'Монета пам'ятки (вартість 1 руб.) від 175 років від дня народження Т.Г. Шевченка', and 'Купон багаторазового використання номінальної вартістю 100000 карбованців'. The 'Зображення' node is connected to '3D віртуальні музеї', 'Национальний музей Тараса Шевченка', 'на Фотосфіді', 'на листівках', 'в наглядності та нумізматиці', and 'в філателістиці'. Other nodes include 'Ушанування пам'яті' (Honoring memory), 'Музеї' (Museums), 'Філателістика' (Philately), and 'НЕГАТИВИ' (Negatives). The 'НЕГАТИВИ' node is connected to various historical photographs and documents, such as 'Флігель-поміщиків садибі Репнік (м. Яготин) на Полтавщині, де жив Т.Г. Шевченко 1843-1844', 'Интер'єр вестибюлю палацу М.Терещенка', 'Відкриття Державного музею Т.Г.Шевченка в Києві', 'К. Бульдин, Я. Чайков. Провод на пам'ятника Т.Г. Шевченкові', 'Скульптор Воллунін. Проект пам'ятника Т.Г. Шевченкові в Москві 1918', 'М. Гейлан, М. Ашпназ. Проект пам'ятника Т.Г. Шевченкові 1928-193', 'Григорь. Смітський. Фотопортрет пам'ятника Т.Г. Шевченкові в Москві', 'М. Внштейн. Проект пам'ятника Т.Г. Шевченкові 1928-1933рр.', 'Крейслерде Г.П. Модел пам'ятника Т. Шевченкові в м. Полтаві. Пис. 1', 'Крейдельде Г.П. Модел пам'ятника Т.Г. Шевченкові 1928-1933 рр.', 'В. Касин, Х. Давидович, С. Крайкович. Проект пам'ятника Т.Г. Шевченкові', and 'Київський З.А. Купон на пам'ятника Т.Г. Шевченкові для М.'.

Рис. 9.24. НумізMATика і зв'язані категорії

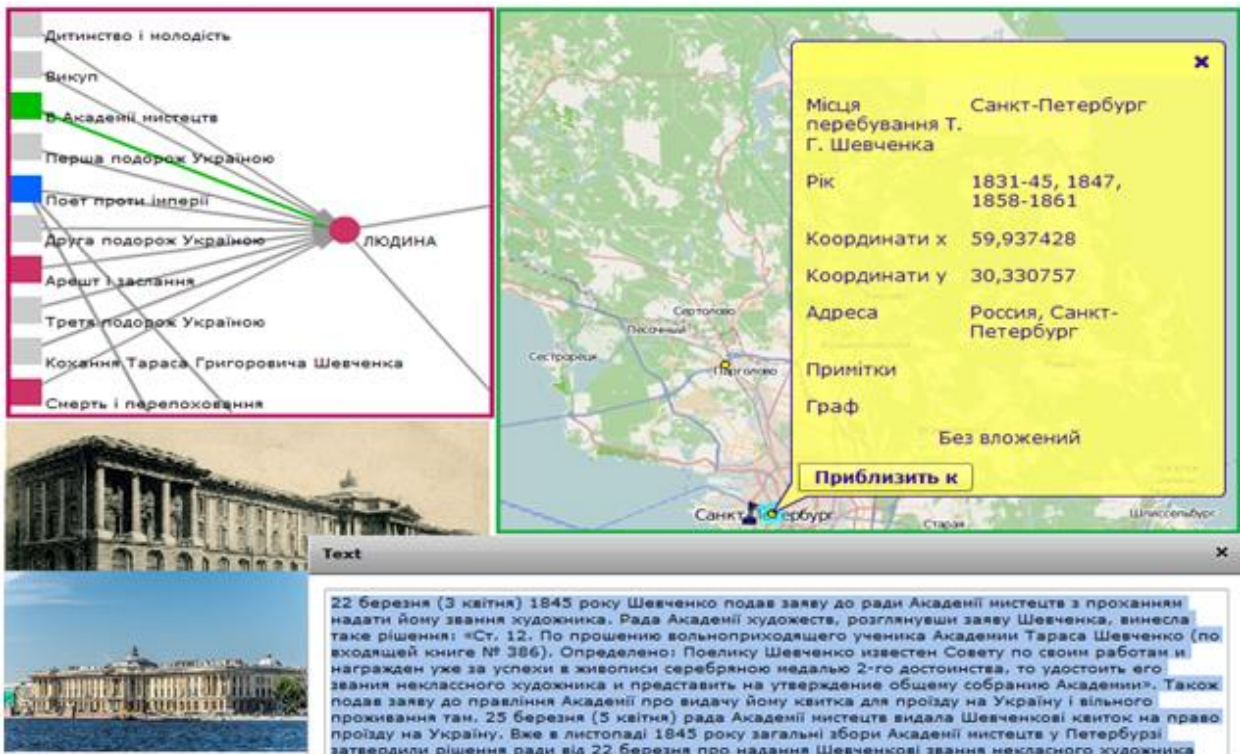


Рис. 9.25. Онтологія міст творчої діяльності



Рис. 9.26. Онтологія творчості

**Национальний центр "Мала академія наук України"**  
Карта проєктована на карті Тараса Григоровича Шевченка

**Vilnius**  
Місце перебування Т. Г. Шевченка  
Рік: 1829-31  
Координати x: 25,276664  
Координати y: 54,674914  
Адреса: Литва, Вільнюс  
Принітки: Без вложення  
Граф:

**Document preview** | occurrence: **Вільнюс**

У Вільні, городі преславнім,  
Оце случилося недавно,  
Ще був тойлі... От як на те  
Не вб'яго в віршу цього слова...  
Тоді здоровий-презоровий  
Зробили з його лазарет,  
А бакалярів розігнали  
За те, що шапки не ламати

Першим коханцем молодого Шевченка була молода дівчина, однокласниця — Оксана. Родина та сватів закоханої були вкрай ворожі до молодого українця, який дискусював з її батьком. Але навіть були нарешті — Тарас у квітні свого лавка Палла Бюгальдара група побігла до Вільнюса (тепер — Вільнюс). Залишилася була можливість і доїхати. Усе своє подальше життя Шевченко буде в містечку відкрити ту дівчину, яку кохав кохав. Наступним коханням, яку він називав своєю коханкою, була польська шевченка Дарія Гукієвська. У 1843 році Шевченко їде в Україну і там зустрічає Ганну Захарівну, з якою згодом присвятив вірш № 3-4. Наступним коханням, що займали його в серці Кобзаря, були Варвара Реліна, сільська дівчина Глофіра та Агата Русова, 18-річна актриса Катерина Пушова, яка, напевно, просто не мала жодної любові своєї життя в мовчанні, але сподівалася віддати своє життя, щоб тільки на тривалість роки був старшим від неї. С) Політиски, яка займалася в Петербурзі, подорожувала, але вона не заважала в світі. Я покинула поета, абишла замість її місцею молодого. Луврив Яковлева-П) віддала пророкості на голову Шевченка

**exalead** <> У Вільні, городі преславнім. Тарас Шевченко. Повне зібрання творів. Том 2. <http://tzboryk.org.ua/shevchenko/shev275.htm> - 6,8 KB - Sep 2, 2011

Back to results

Рис. 9.27. Трансдисциплінарне відображення онтології (ГІС, текст)

**Национальний центр "Мала академія наук України"**  
Карта проєктована на карті Тараса Григоровича Шевченка

**Fort Shevchenko**  
Місце перебування Т. Г. Шевченка (Форт Г. Шевченка)  
Рік: 1850-57  
Координати x: 44,507568  
Координати y: 50,26272  
Адреса: Казахстан, Мангістауська область, Тупкараганський район, Форт-Шевченко  
Принітки:

**Document preview** | occurrence: **Ужгород**

**ЛІТЕРАТУРНО-МЕМОРІАЛЬНИЙ МУЗЕЙ І. Г. ШЕВЧЕНКА** у Форті Шевченка — державний музей Шевченка в Каз. РСР. Відкритий 1.V 1932 за рішенням РНК Казахстану від 12.X 1925 г. м. Форт Шевченка на п-ві Мангішлак, де поет відбував заслання 1850 — 57. Розширений у кол. будинок Ужгород, де був Шевченко. У п'яти залах музею зібрано матеріали про життя, творчість і революц. діяльність Шевченка: численні фотодіагностичні документи, історичні картини Шевченка, картини рад. художників, документи про українців пам'яті поета в раз. час. тощо. Перед музеєм на високому постаменті — мармурове погруддя Шевченка (скульптор В. Бокланов), поруч же один пам'ятник поетові. В Шевченковому саду зберігаються стіги і земляники, де працював поет. Над ними зроблено знаменку споруду. Музей проводить велику культ.освітню роботу.

Лит. Пана В. І. Музей великого Кобзаря у Форті Шевченка. К., 1975.  
Є. Ужгород.

**«САД БІЛЯ НОВОПЕТРОВСЬКОГО УКРІПЛЕННЯ»** — малюнок, ескіз й начерки Шевченка, виконані 1853 — 57 в Новопетровському укріпленні. В 1853 за наказом команданта І. Усовою біля укріплення почали садити сад. Дерева для цього привезли з Астрахані, Гур'єва та Ханга-Баби. Садити сад допомагав і Шевченко. В саду було споруджено лінійний будинок для роботи команданта азіатців де Усовий «обидав» поставлено юрті; в якій малював Шевченко, збудовано землянку для вихованця Усова, в якій поет кохав свої начерки. Цей сад названо ім'ям Шевченка. Сад Шевченко відтворив у аварелі, рисунках олівцем, ескізі і начерках. Рисунок олівцем «Місце майбутнього саду біля Новопетровського укріплення» (12,1 × 32,5) створено олівцем 1853, коли на місці майбутнього саду були лише воєнопільні споруди в лінійному будинку команданта. З трьох основних малюнків «Сад біля Новопетровського укріплення» одну аварелю (12,4 × 29,8) виконано воском 1853, коли в саду вже було насаджено дерева й збудовано апарату. На малюнку зображено й каплі, юрті, розташовані поблизу саду. Другу аварелю (17,1 × 29) Шевченко виконав улітку 1854. Поверхню умітку — монохромна автора й дата: «Ш. 1854». Обидві аварелі художник подарував родніи Усовиці. Рисунок олівцем, на якому зображено сад з тиньового боку на тлі укріплення, виконаний між 1854 і 1855. Один з начерків олівцем — «Біля Новопетровського укріплення» (11,4 × 28,7) — між 1854 і липнем 1857, другий — «Службова споруда поблизу батареї № 2» (12,9 × 28,7) — між 1855 і 1857. На першому плані тут

В портрет використано К і новий релігійного значення каваки, що і в двох нерозриваних олівцях, про які Шевченко писав Бр. Захарівну в листі від 20.V 1857 р. до листа від 8.V 1857 р. (Лит. № 222 — 223). Далі йде на підставі олівця дитини з олівця 1856 р. «Елеголічний дитини». Сандвичівський «Кавказ» (Лит. № 48, 49, 50), а також з використанням олівця «Молтва по вмерлому» (Лит. пр.к. № 222 — 223), надані Бр. Захарівну 20.V 1857 р. Портрет був подарований Шевченком родніи Усовиці: «Усовиці на ланці о проєкти з ними картами і о картині зібрані і о картині портрети жими команданта Азіатів Бюгальдара на старій дівчині на дурак. Каві — імає чарівний в королівстві олівця в лів. пейзаж...» («Філософський К. Н. Усовий о Шевченку», «Кавказський старик», 1899, кн. II, стор. 304). Поверхню цієї зображення виконав Усовиці, Н. І. Сіпко, ІТШ, РКД. Дитини, літ. № 7 г. н. 246. Малюнок «Кавказська Катя» Фейделі Кобзар створив під час своєї «азиатського заслання». Теня землі та зображення кавакию наряду з пейзажем гешаліного теорія, імає виконати з своєї творчості.

**exalead** <> Лис-П'є. Шевченківський словник. У двох томах. Том 1. <http://tzboryk.org.ua/shevchenko/slov22.htm> - 45,3 KB - Sep 2, 2011

Back to results

Рис. 9.28. Трансдисциплінарне відображення (ГІС, текст, мульти-медіа)

Рис. 9.29. Трансдисциплінарність відображення (ГІС, текст, операції тематичного пошуку, мульті-медіа)

Таким чином, спадщина Шевченка відображається зі збереженням його поетичної сили і національного авторитету. Якщо говорити про фізичний вимір збереження Шевченкових цінностей, то тут онтологічний підхід забезпечує рішення проблем щодо відображення зв'язку творів Кобзаря з творами інших мистців, причому якщо говорити про відображення духовного та феноменологічного зв'язку - то це проблема не проста і загальна.

Онтологічне моделювання історичних особливостей формування української культури, долі Шевченкової малярської спадщини, специфіки поетичного і образотворчого видів мистецтва забезпечує більш поглиблене дослідити ці тенденції.

Між тим, онтологічні описи Шевченкової мистецької спадщини, дозволяють більш глибоко дослідити його новаторство у багатьох художніх галузях та вплив на подальший розвиток національного мистецтва, що дозволяє більш об'єктно говорити про нього як про одного з найвидатніших майстрів ХІХ ст.

Ми звично декларуємо, що в Шевченкові якнайповніше сконцентровано буття українського народу на всі часи. А отже, його творчість має пронизувати всі пласти духовної культури нашого суспільства. Духовну культуру можна означити як складне суперечливе явище. Вона включає взаємозв'язані рухомі пласти, які взаємодіють в процесі функціонування, визначають її статус.

Якщо останню умовно уявити у вигляді піраміди знань з горизонтальним означенням її складових, відповідно: офіційна культура, національна (народна), тіньова, молодіжна культура, то вертикаль Шевченкового мистецтва повинна бути присутня тою чи іншою мірою у кожній з них.

Передусім, це офіційна культура, яка в силу історичного буття України, утвердилась як базова, міцно сформована державними інституціями, що в різні форми соціального контролю скеровували провінціальний розвиток національної культури. Фактично геній, що приходить у заблоковану культуру, з перших же кроків починає задихатися. Він заблокований звідусіль державою, трагізмом сфальсифікованої історії, консерватизмом смаків, кон'юнктурою сучасного моменту.

## **ВИСНОВКИ**

Розглянутий у роботі підхід до побудови описів історичних особистостей дозволяє врахувати інформацію з різних джерел. Описи (ієрархії нечітких понять), отримані в результаті, можуть використовуватися при побудові і наповненні спеціалізованих порталів знань, а також формуванні глобальної онтології історичних особистостей.

Даний підхід не обмежений застосуванням для опису історичних особистостей, а може використовуватися, наприклад, для побудови опису історичних подій, визначення основного напрямку роботи вченого тощо. Основною умовою є семантично схожа тематична спрямованість досліджуваних інформаційних масивів. Як видно з прикладу, саме тематична спрямованість і впливає на значення функції приналежності в нечіткому формальному контексті досліджуваних розподілених інформаційних ресурсів.

Серед питань, які необхідно додатково досліджувати, можна виділити дослідження залежності побудованих ієрархій понять від різних варіантів визначення функції приналежності в нечіткому формальному контексті, а також розвиток розглянутого підходу для побудови онтологій історичних особистостей.

На основі аналізу процесів обробки інформації в системах управління знаннями показано, що обробка інформації в них здійснюється експертами в умовах неповноти, лінгвістичної невизначеності та необхідності врахування як кількісної так і якісної інформації. Програмні комплекси, що функціонують в таких системах розроблені під конкретну предметну область та надають обмежені можливості для обробки інформації. Тому практично важливою й актуальною науково-технічною задачею є розробка підходу до обробки інформації в системах управління знаннями, відповідних методу, моделей та алгоритмів, а також гнучкої технології та інструментальних засобів створення програмних модулів обробки інформації для їх практичного застосування.

## ЛІТЕРАТУРА

### Література до розділу 1

1. Бухарестская декларация (общеевропейская конференция в Бухаресте по подготовке всемирной встречи на высшем уровне по вопросам информационного общества). – Режим доступа: [www.itu.int/dms\\_pub/itus/md/03/wsispc2/doc/S03WSISPC2DOC0005!!\\_MSW-R.doc](http://www.itu.int/dms_pub/itus/md/03/wsispc2/doc/S03WSISPC2DOC0005!!_MSW-R.doc). – Дата доступа: 17.08.2010. – Название с экрана.
2. Ракитов А. И. Философия компьютерной революции / Ракитов А. И. – М. : Политиздат, 1991. – 287 с.
3. Башмаков А.И. Интеллектуальные информационные технологии: Учеб. пособие / А.И. Башмаков, И.А. Башмаков. – М.: Изд.-во МГТУ им. Н.Э. Баумана, 2005. – 304 с.
4. Конноли Т. Базы данных: проектирование, реализация и сопровождение. Теория и практика, 2-у изд.: Пер. с англ. / Конноли Т., Бегг К., Страчан А.– М.: Издательский дом «Вильямс», 2001. – 1120с.
5. Комов С.А. Управление знаниями – что это и как ими управлять?/Журнал «Корпоративные системы» -03.2005, с.17-25
6. Комов С.А. Аналитика и разведка в организации – чем работать?/Журнал «Корпоративные системы» -06.2005. с.22-29
7. Знание-ориентированные информационные системы с обработкой естественно-языковых объектов: онтологический подход / [Палагин А.В., Кривый С.Л., Петренко Н.Г., Величко В.Ю.]. – УСиМ, 2010. – № 4, 5. – С. 3–14.
8. Теоретичні основи проектування інформаційних середовищ, як педагогічних систем, спрямованих на підтримку творчої діяльності учнів : монографія / за ред. канд. техн. наук В.В. Камишина і канд. техн. наук О.Є. Стрижака. – К.: Інформаційні системи, 2010. – 194 с.
9. Формування авторських методик викладання предметних дисциплін в середовищі серверу підтримки навчальної взаємодії : монографія / за ред. В.В. Камишина, О.Є. Стрижака. – К.: Інформаційні системи, 2010. – 138 с.
10. Застосування серверів підтримки навчальної взаємодії в педагогічному процесі : методичні рекомендації / за ред. канд. техн. наук В.В.Камишина і канд. техн. наук О.Є.Стрижака. – К.: Інформаційні системи, 2011. – 232 с.
11. Використання розподільних інформаційних систем в навчальному процесі позашкільного навчального закладу Мала академія наук : методичні рекомендації / за ред. д-ра техн. наук, академіка НАНУ О.В. Палагіна і канд. техн. наук О.Є. Стрижака. – К.: Інформаційні системи, 2012. – 182 с.
12. Інформаційні системи підтримки навчальної та дослідницької групової діяльності обдарованих учнів за схемою - <учні-викладачі-науковці> : методичні рекомендації // за ред. д-ра техн. наук, академіка

НАНУ О.В. Палагіна і канд. техн. наук О.Є. Стрижака. – К.: Інформаційні системи, 2012 – 198 с

13. Андрусенко Т.Б, Стрижак О.Є. - Управление учебным процессом на основе тезаурусов.: e-LearningWorld (Москва). — 2007.– N 1. – с.56-62.

14. Семинар «Электронное образование (e-learning) в сфере высоких технологий» в МЭСИ / Электронное издание «Наука и технологии в России», 17 ноября 2009. – Режим доступа: <http://rusnanonet.ru/news/29170/>. – Дата доступа: 22.11.2010. – Название с экрана.

15. Капитонова Ю.В. Парадигмы и идеи академика В.М. Глушкова / Ю.В. Капитонова, А.А. Летичевский. – Киев, Наукова думка, 2003. – 456 с.

16. Искусственный интеллект: В 3 кн. Кн. 1. Системы общения и экспертные системы: [Справочник / Научн. ред. Э.В. Попов]. – М.: Радио и связь, 1990. – 464 с.

17. Искусственный интеллект: В 3 кн. Кн. 3. Программные и аппаратные средства: Справочник / [Научн. ред. В.Н. Захаров, В.Ф. Хорошевский]. – М.: Радио и связь, 1990.– 368 с.

18. Египко В.М., Организация и проектирование систем автоматизации научно-технических экспериментов. – Киев: Наукова думка, 1978.

19. Абрамов А.Е., Кошкина А.О. Пути автоматизации исследований процесса ЭМО деталей машин. – Современная техника и технологии. – Октябрь, 2011. [Электронный ресурс]. URL: <http://technology.snauka.ru/2011/10/107>.

20. Комплекс автоматизации экспериментальных и технологических установок АСTest. – ООО «Лаборатория автоматизированных систем (АС)» – Москва 2007.

21. Зінченко С.В. Проблеми представлення знань в інформаційних технологіях / С.В. Зінченко, І.М. Герасімчук. – К.: НТУУ «КПІ», 2010. – 240с.

22. Бельчусов А.А. Разработка онтологии школьного курса «Информатика и ИКТ» / А.А. Бельчусов // Матер. II Междунар. науч.-практ. конф. "Информационные технологии в образовании" ("ИТО-Сибирь-2008"). – Томск, 25–27 сентября 2008. – [http://ito.itdrom.com/files/ito\\_zip/51\\_1.doc](http://ito.itdrom.com/files/ito_zip/51_1.doc).

23. Египко В.М. Методика применения индивидуальных компьютерных средств для автоматизации элементов профессиональной деятельности / Египко В.М., Зинченко В.П. // Кибернетика и вычислительная техника. Дискретные системы управления. – 1993. – Вып. 97. – С. 86–91.

24. Функциональное наполнение прикладного программного обеспечения проблемно-ориентированных комплексов / Зинченко С.В., Зинченко Н.П., Криворученко В.С. и др. // Технології створення перспективних комп'ютерних засобів та систем з використанням новітньої елементної бази. – К.: НАН України, ІК ім. В.М. Глушкова, 2000. – С. 61–66.

25. Зинченко С.В. Концепция создания онтолого-управляемой информационной системы / С.В. Зинченко // Інформаційні системи, механіка та керування: наук. зб. – К.: НТУУ «КПІ», 2008. – Вип. № 1. – 11 с.

26. Онтологічні керовані інформаційні технології та їх застосування для отримання якісної освіти в загальноосвітніх школах м. Києва // Творчість та освіта в інтелектуальних пошуках та практиках сучасності: зб. тез допов. ІХ міжнар. наук.-практ. конф. – К.: ІВЦ “Політехнік“, 2007. – 2 с.
27. Энциклопедия кибернетики (в двух томах) /– Главная редакция Украинской советской энциклопедии (отв. научн. ред. В.М. Глушков), – Киев, 1974. – Т. 1,2. – 1233 с.
28. Федорук П.І. Адаптивна система дистанційного навчання та контролю знань на базі інтелектуальних Інтернет-технологій / П.І. Федорук. – Ів.-Фр.: – Прикарп. нац. ун.-т, 2008. – 326 с.
29. Moodle. – Режим доступа:<http://www.moodle.org>. – Дата доступа: 26.05.2011. – Название с экрана.
30. Ontologies and the Semantic Web for E-learning / Sampson, D. G., Lytras, M. D., Wagner, G., & Diaz, P. Educational Technology & Society, 7 (4). – 2004. – pp. 26–28. – Режим доступа: <http://www.miltiadislytras.net/8.pdf>. – Дата доступа: 21.03.2006. – Названиесэкрана.
31. К вопросу автоматизированного построения онтологии предметной дисциплины для электронных курсов обучения II / [Палагин А.В., Петренко Н.Г., Тихонов Ю.Л., Величко В.Ю.]. – Вісник східноукраїнського університету ім. В. Даля. – 2010. – №4 (150). – С. 171–178.
32. Палагин А.В. Об одном подходе к формализованному представлению онтологии текстового документа / А.В. Палагин, Н.Г. Петренко, Севрук А.О. – Комп'ютерні засоби, мережі та системи, 2007. – №6. – С. 14–20.
33. Чернецька Т. І. Сучасний урок: теорія і практика моделювання: [навч. посібник] / Т. І. Чернецька. – К. : ТОВ «Праймдрук». – 2011. – 352 с.
34. Чернецька Т. І. Освітнє середовище навчально-дослідницької діяльності дітей: особливості проектування, моделювання і функціонування / Т. І. Чернецька // Наукові записки Малої академії наук України : (збірник наукових праць). К. : ТОВ «Праймдрук». – 2012. – С. 63-78.
35. Чернецька Т. І. Формування мети і цілей в контексті реалізації компетентнісного підходу в початковій ланці освіти / Т. І. Чернецька // Імідж сучасного педагога. – 2012. – № 6(125). – С. 7–10.
36. Чернецька Т. І. Актуальні питання становлення змістового контенту поняття «процес навчання» / Т. І. Чернецька // Педагогічний вісник. – 2012. – № 1 (21) . – С. 80–83.
37. Чернецька Т. І. Основні тенденції розвитку інформаційного суспільства і глобального інформаційного простору та їх вплив на організацію навчання обдарованих дітей / Т.І. Чернецька // Актуальні питання, проблеми та перспективи залучення громадських організацій до підтримки обдарованих дітей та молоді в сучасному інформаційному просторі : матеріали круглого столу в рамках проведення чотирнадцятої міжнародної виставки навчальних закладів «Сучасна освіта в Україні –



2011», Київ, 16 лютого 2011р. – К : ТОВ «Інформаційні Системи», 2011. – С.5–10.

38. Чернецька Т. І. Формування освітнього середовища: сутнісний зміст ієрархічно-діяльнісного підходу / Т.І.Чернецька // Моделювання особистісно-розвивального середовища обдарованої дитини : матеріали Всеукраїнської науково-практичної конференції, Київ, 11-12 жовтня 2011р. – К : ІОД НАПН України, 2011. – С. 150–154.

39. Чернецька Т. І. Моделювання процесу навчання в контексті розвитку обдарувань учнів засобами освітньо-інформаційного середовища / Т.І.Чернецька // Інноваційні технології навчання обдарованої молоді : матеріали Міжнародної науково-практичної конференції, Київ, 8-9 грудня 2011р. – К : ІОД НАПН України, 2011. – С.68–73.

40. Чернецька Т. І. Сучасне моделювання освітніх процесів: актуальні питання та способи їх реалізації в умовах функціонування Малої академії наук України» / Т.І.Чернецька // Інноваційні технології в дошкільній освіті України: розвиток дитячої обдарованості та креативності» : матеріали Всеукраїнського науково-методичного семінару Інституту розвитку дитини НПУ імені М. П. Драгоманова / укл. І. І. Загарницька. – К. : НЦ «МАНУ», 2012. – С. 318–341.

41. ЧернецькаТ. І. Теоретичні засади проектування освітнього середовища навчально-дослідницької діяльності дітей дошкільного віку. / Т.І.Чернецька // Проблеми діагностики та проектування розвитку обдарованих дошкільників : матеріали Всеукраїнської науково-практичної конференції, 10-11 квітня 2012 р., м. Одеса. – К. : ТОВ «Інформаційні системи», 2012. – С. 109–115.

42. ЧернецькаТ. І. Структурно-функціональні основи формування освітнього середовища навчально-дослідницької діяльності дітей. / Т.І.Чернецька // Психолого-дидактичні засади формування освітнього середовища навчально-дослідницької діяльності дітей : матеріали Всеукр. наук-практ. конф. (Кіровоград, 24-25 квітня 2012 р.) М-во освіти і науки, молоді та спорту України, Кіровоградський обл. ін-т післядипл. педаг. Освіти ім. Василя Сухомлинського. – Кіровоград : ТОВ «ПОЛІМЕД-Сервіс», 2012. – С. 122–125.

43. Попова М. А. «Формування освітнього середовища навчально-дослідницької діяльності учнів на основі впровадження ГІС-технологій». - Наукові записки Малої академії наук України: (збірник наукових праць). – К. : ТОВ «Праймдрук». – 2012. – 304 с.

44. Палагин А.В., Величко В.Ю., Стрижак А.Е., Попова М.А. «Инструменты поддержки процессов аналитической деятельности эксперта при тематическом исследовании информационных ресурсов и источников». - International Journal “Information Technologies and Knowledge”. – 2010. -Vol. 4, Number 4. – С. 329-347.

45. Попова М. А. «Методика побудови онтолого-керованих інформаційних ресурсів як елементу комп'ютерних ділових ігор для

навчання фахівців в галузі екологічної безпеки»- Екологічна безпека та природокористування: Зб. наук. праць / М-во освіти і науки України, Київ. нац. ун-т буд-ва і архіт., НАН України, Ін-т телекомунікацій і глобал. інформ. простору; редкол.: О.С. Волошкіна, Е45 О.М. Трофимчук (голов. ред.) [та ін.]. — К., 2012. — Вип. 10. —258с.: іл.

46. Стрижак О.Є., Січкара Т.Г., Кальной С.П., Шут М.І. - Віртуальні фізичні кабінети як інструмент поглиблення пізнавального процесу. - Збірник наукових праць Кам'янець-Подільського національного університету. Серія педагогічна. Випуск 14. Інновації в навчанні фізиці та дисциплін технологічної освітньої галузі: Міжнародний та вітчизняний досвід. 2008р. с166-168.

47. Olexandr Strizhak – Use of intellectual information resources in education system – 21st International CODATA Conference. Scientific Information for Society – from Today to the future, October 5-8,2008,Kyiv, Ukraine, p.233-234.

48. Alexander Stryzhak – Instruments of management knowledges at teaching of the gifted children – Ucen zdolny wyzwaniem dla wspolczesnej edukacji. Praca zbiorowa pod redakcja Jana Laszczyka, Malgorzaty Jablonowskiej. Warszawa 2008. Wydawnictwo Akademii Pedagogiki Specjalnej. p. 268-272.

49. АндрусенкоТ.Б., КальнойС.П., СтрижакО.Є., ЯценкоГ. - Підтримкатворчогорозвиткушколярівнаосновізабезпеченнядоступудоінтелектуальнихінформаційнихресурсів, НоваПедагогічнадумка, №1, 2008. с.95-97

50. SergienkoV, Stryzhak A. - Instruments of ma-nagement knowled-ges at teaching of the gifted children /Medzinárodní vedecko - odborná konferenceZÁUJMOVÁ ČINNOSTŽIAKOV stav, problémy, trendy 27. 11. - 28. 11. 2008. p.182-189

51. П.І.Коваленко, О.Є.Стрижак, А.В.Кравець – Інформаційні технології: аграрне виробництво та розвиток сільських територій// Формування енергогенеруючих біоорганічних агроєкосистем. Науково-технологічне забезпечення аграрного виробництва/ за ред..член-корр. УААН Ю.А.Тараріко/ К.: Изд-во ПГІМ «Агроресурси», 2008, с.133-140.

52. Стрижак О.Є. – Віртуальна школа МАН, як забезпечення взаємодії зобдарованою молоддю в середовищі глобальної мережі Інтернет / Матеріали круглого столу «Інноваційні технології сприяння розвитку обдарованості» м.Тернопіль 15 січня 2009р., К.:Вид-во ІОД АПНУ, с.23-28.

53. Стрижак О.Є. – Комп'ютерні тезауруси як технологічна платформа створення авторських методик викладання предметних дисциплін// Актуальні проблеми психології: Психологічна теорія і технологія навчання/ за ред.. С.М. Максименко, М.Л.Смульсон. – К.: Вид-во НПУ імені М.П.Драгоманова, 2009,- Т.8, вип..6. с.259-266.

54. Стрижак О.Є., Кальной С.П. – Віртуальна школа МАН – платформа формування системи знань для підтримки пізнавальної діяльності учнів в мережі Інтернет – Матеріали науково-практичної конференції

«Виявлення та підтримка обдарованості учнів загальноосвітньої школи», м.Тернопіль 24-26 червня 2009 р., К.: Вид-во ІОД АПНУ, с.229-237.

55. Віртуальні класи. Використання в учбовому процесі позашкільних навчальних закладів. Методичні рекомендації / Стрижак О.Є., Кальной С.П., Довгий С.О., Лісовий О.В., Трофімчук О.М., - АПН, ІОД, 2009, 247 с.

56. Засоби доступу до джерел знань, та їх використання в учбовому процесі педагогічного навчального закладу. Методичні рекомендації / за редакцією члена-кореспондента НАН України С.О.Довгого і канд.техн.наук О.Є.Стрижака, - АПН, ІОД, 2009, 66 с.

57. „Методики використання сучасних інформаційних технологій при підтримці процесу навчання обдарованої молоді. Методичні рекомендації / за редакцією члена-кореспондента НАН України С.О.Довгого і канд.техн.наук О.Є.Стрижака, - АПН, ІОД, 2009, 199 с.

58. Олександр Стрижак – Управління знаннями в навчальному процесі, як системостворюючий фактор підтримки пізнавальної діяльності учнів// Світ виховання, №4 (35), 2009, с.5-8.

59. Стрижак О.Є., Кальной С.П., Ковбасюк Ю.Г. – Сервер підтримки навчальної взаємодії – технологічна платформа побудови інформаційно-розвивального середовища особистості//Матеріали міждисциплінарної науково-практичної конференції «Інноваційні технології навчання обдарованої молоді», 02-03 грудня 2009 р., с.181-190.

60. С.П.Кальной, О.Є.Стрижак – Сервер підтримки навчальної взаємодії – платформа надання дистанційних послуг в освіті./Матеріали міжнародної науково-технічної конференції «Сучасні методи, інформаційне, програмне та технічне забезпечення систем управління організаційно-технологічними комплексами», 26-27 листопада 2009 р., К.:НУХТ, с.91-92.

61. О.Є.Стрижак, О.О.Улашин – Технологічні аспекти формування системи знань навчального призначення / Матеріали міжнародної науково-технічної конференції «Сучасні методи, інформаційне, програмне та технічне забезпечення систем управління організаційно-технологічними комплексами», 26-27 листопада 2009 р., К.:НУХТ, с.92-93.

62. Свідоцтво про реєстрацію авторського права на твір №30844 – Комп'ютерна програма «Сервер підтримки навчальної взаємодії. Віртуальна школа Малої академії наук» («Сервер підтримки навчальної взаємодії (ВШ МАН)»).Стрижак О.Є., Кальной С.П., ДовгийС.О., Трофімчук С.М., Лісовий О.В. – 03.11.2009 р.

63. Стрижак О.Є. Управління знаннями в навчальному процесі, як системостворюючий фактор підтримки пізнавального розвитку обдарованої дитини / Матеріали всеукраїнської науково-практичної конференції «Обдаровані діти – інтелектуальний потенціал держави», 22-23 вересня 2009 р., К.: Вид-во ОІД АПНУ, с.167-174.

64. Кальной С.П., Стрижак О.Є. – Система надання дистанційних послуг для підтримки позашкільної пізнавальної діяльності обдарованої дитини / Матеріали всеукраїнської науково-практичної конференції

«Обдаровані діти – інтелектуальний потенціал держави», 22-23 вересня 2009 р., К.: Вид-во ОІД АПНУ, с.80-85.

65. Gruber, T.R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human and Computer Studies*, 1995. 43(5/6): 907–928.

66. Stafford, Philip You can believe the Hype Cycle's take on technology (англ.) // *Financial Times*. — 2008. — № May 30.

67. AN EXALEAD S.A. CLIENT SUPPORT DOCUMENT Doc. No. EN.120.0002.0-V4.6.1 - March 31, 2008 Copyright © 2003 - 2008 by Exalead S.A. All rights reserved.

68. HowE-learningWorks / L. A. Obringer. – Режим доступа: <http://communication.howstuff-works.com/elearning.htm>. – Дата доступа: 30.05.2010. – Название с экрана.

69. E-learning 2.0 by Stephen Downes, National Research Council of Canada October 17, 2005. -

Режимдоступа:<http://www.elearnmag.org/subpage.cfm?article=29-1&section=articles>. – Датадоступа: 17.05.2008. – Названиесэкрана.

70. Harms U. Virtual and remote labs in physics education / U. Harms// *Second European Conf. on Physics Teaching in Engineering Education*. – Budapest, Jun. 2000.

71. Leitner L.J. A virtual laboratory environment for online it education / Leitner L.J, Cane J.W. // *Proc. of the 6th Conf. on Information technology education (SIGITE '05)*. – NY: ACM Press, 2005. – P. 283–289.

72. Repenning A. Collaborative use and design of interactive simulations // Repenning A., Ioannidou J. // *Proc. of the CSCL*. – Stanford, CA, 1999.

73. Remote versus hands-on labs: a comparative study / Corter J.E., Nickerson J.V., Esche S.K., Chassapis C. // *Proc. of the 34th Annual Conf. on Frontiers in Education (FIE'04)*. – Oct 2004. – 2. – P. 17–21.

74. Менеджмент в некоммерческой организации: принципы и практика. — М.: «Вильямс», 2007. — С. 304.

75. Застосування серверів підтримки навчальної взаємодії в педагогічному процесі: Методичні рекомендації / за ред. канд. техн. наук В.В.Камишина і канд. техн. наук О.Є.Стрижака. – К.: ТОВ Інформаційні системи. -2011. -308с.

76. Засоби доступу до джерел знань, та їх використання в навчальному процесі педагогічного навчального закладу : Методичні рекомендації / За редакцією члена-кореспондента НАН України С. О. Довгого і канд. техн. наук О.Є. Стрижака. - 2-ге вид., доповн. - К.: Інститут обдарованої дитини, 2012. - 192с.

## **Література до розділу 2**

1. Сагатовский В.Н. Основы систематизации всеобщих категорий / Сагатовский В.Н. – Издательство Томского университета, Томск, 1973. – 432 с.

2. Палагин А.В. Системно-онтологический анализ предметной области / А.В. Палагин, Н.Г. Петренко. – УСиМ, 2009. – № 4. – С. 3–14.
3. Палагин А.В. Системная интеграция средств компьютерной техники / А.В. Палагин, Ю.С. Яковлев. – Винница: УНІВЕРСУМ, 2005. – 680 с.
4. Башмаков А.И. Интеллектуальные информационные технологии: Учеб. пособие / А.И. Башмаков, И.А. Башмаков. – М.: Изд.-во МГТУ им. Н.Э. Баумана, 2005. – 304 с.
5. Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – СПб.: Питер, 2001. – 384 с.
6. Guarino N. Ontologies and Knowledge Bases. Towards a Terminological Clarification / N. Guarino, P. Guaretta. – Amsterdam: IOS Press, 1995. – 7 p.
7. Gruber T. R. A translation approach to portable ontology specifications / Gruber T. R. – Knowledge Acquisition, 5 (2), 1993. – PP. 199–220.
8. Gruber, T.R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing / Gruber T. R.. – International Journal of Human and Computer Studies, 43(5/6), 1995. – PP. 907–928.
9. Niles I. Towards a Standard Upper Ontology / I. Niles, A. Pease // 2nd International Conference on Formal Ontology and Information Systems (FOIS-2001), Welty C., and Smith B., eds., Ogunquit, Maine. – 17–19, October, 2001. – 8pp. – Режим доступа <http://home.earthlink.net/~adampease/professional/FOIS.pdf>. – Дата доступа: 16.12.2008. – Название экрана.
10. Палагин А.В. Архитектура онтолого-управляемых компьютерных систем / Палагин А.В. – Кибернетика и системный анализ, 2006 – №2. – С. 111–124.
11. Палагин А.В. Организация и функции "языковой" картины мира в смысловой интерпретации ЕЯ-сообщений / А.В. Палагин. – Information Theories and Application, 2000. – Vol. 7, №4. – С. 155–163.
12. Чернавский Д.С. Синергетика и информация: Динамическая теория информации / Д.С. Чернавский – Изд. 2-е, испр., доп. М.: УРСС, 2004. – 288 с. – Режим доступа: <http://ru.wikipedia.org/wiki/BookSources/5354002419>. – Дата доступа: 16.12.2008. – Название с экрана.
13. Новейший философский словарь. Синергетика. – Режим доступа: [http://slovari.yandex.ru/dict/phil\\_dict/article/filo/filo-707.htm](http://slovari.yandex.ru/dict/phil_dict/article/filo/filo-707.htm). – Дата доступа: 17.05.2011. – Название с экрана.
14. Князева Е.Н. Основания синергетики. Синергетическое мировидение / Е.Н. Князева, С.П. Курдюмов. – М.: КомКнига, 2005. – 240 с.
15. Капитонова Ю.В. Парадигмы и идеи академика В.М. Глушкова / Ю.В. Капитонова, А.А. Летичевский. – Киев, Наукова думка, 2003. – 456 с.
16. Искусственный интеллект: В 3 кн. Кн. 1. Системы общения и экспертные системы: [Справочник / Научн. ред. Э.В. Попов]. – М.: Радио и связь, 1990. – 464 с.

17. Искусственный интеллект: В 3 кн. Кн. 3. Программные и аппаратные средства: Справочник / [Научн. ред. В.Н. Захаров, В.Ф. Хорошевский]. – М.: Радио и связь, 1990. – 368 с.
18. Палагин А.В. К решению основной задачи эмуляции / А.В. Палагин. – УсиМ, 1980. – №3. – С. 24–28.
19. Микропроцессорные системы обработки информации / [Палагин А.В., Денисенко Е.Л., Белицкий Р.И., Сигалов В.И.]. – Киев: Наук. Думка, 1993. – 352 с.
20. Ракитов А.И. Курс лекций по логике науки / А.И. Ракитов – М.: Изд-во «Высшая школа», 1971. – 176 с.
21. Кондаков Н.И. Логический словарь-справочник / Н.И. Кондаков – М.: Наука, 1975. – 720 с.
22. Гаазе-Рапопорт М.Г. Структура исследований в области искусственного интеллекта. Толковый словарь по искусственному интеллекту / М.Г. Гаазе-Рапопорт, Д.А. Поспелов. – М.: Радио и связь, 1992. – С. 5–20.
23. Налимов В.В. Спонтанность сознания: вероятностная теория смыслов и смысловая архитектура личности / В.В. Налимов – М.: Изд-во «Прометей» МГПИ им. Ленина, 1989. – 288 с.
24. Хилькевич А.П. Проблема расширения традиционной силлогистики / А.П. Хилькевич – Минск: Изд. БГУ, 1981. – 115 с.
25. Рыков В.В. Управление знаниями / В.В. Рыков – Режим доступа: <http://rykkкурс2.narod.ru/part2.doc>. – Дата доступа: 17.05.2011. – Название с экрана.
26. Sowa J. F. Knowledge Representation: Logical, Philosophical, and Computational Foundations / J. F. Sowa – Brooks Cole Publishing Co., Pacific Grove, CA, ©2000. – 594 pp.
27. Палагін О.В. Модель категоріального рівня мовно-онтологічної картини світу / О.В. Палагін, М.Г. Петренко. – Математичні машини і системи, 2006. – №3. – С. 91–104.
28. Палагін О.В. Розбудова абстрактної моделі мовно-онтологічної інформаційної системи / О.В. Палагін, М.Г. Петренко. – Математичні машини і системи, 2007. – №1. – С. 42–50.
29. Палагін О.В. Розвиток та порівняльні характеристики логіко-онтологічних формальних теорій / О.В. Палагін, М.Г. Петренко, А.В. Михайлюк. – Математичні машини і системи, 2007. – №2. – С. 3–18.
30. Палагин А.В. К вопросу системно-онтологической интеграции знаний предметной области / А.В. Палагин, Н.Г. Петренко. – Математические машины и системы, 2007. – №3,4. – С. 63–75.
31. Рубашкин В.Ш. Проблема интерпретации в физической теории / В.Ш. Рубашкин – Сб. «Логика и методология науки». – М.: 1967. – С. 35 – 49.
32. Соколовская Ж.П. «Картина мира» в значениях слов / Ж.П. Соколовская – Симферополь: Таврия, 1993. – 197 с.

33. Кузьмин Е.С. Система онтологических категорий / Е.С. Кузьмин – Иркутск, 1958. – 183 с.
34. Свидерский В.И. Некоторые вопросы диалектики изменения и развития / В.И. Свидерский – М.: Мысль, 1965. – 228 с.
35. Библер В.С. О системе категорий диалектической логики / В.С. Библер – Сталинабад, 1958. – 157 с.
36. Зиновьев А.А. Основы логической теории научных знаний / А.А. Зиновьев – АН СССР. Ин-т философии. – М.: Наука, 1967. – 202 с.
37. Апресян Ю.Д. Идеи и методы современной структурной лингвистики / Ю.Д. Апресян – (Краткий очерк). М.: Просвещение, 1966. – 302 с.
38. Широков В.А. Інформаційна теорія лексикографічних систем / В.А. Широков – К.: Довіра, 1998. – 331 с.
39. Кургаев А.Ф. Проблемная ориентация архитектуры компьютерных систем / А.Ф. Кургаев – Киев: Сталь, 2008. – 540 с.
40. Кургаев А.Ф. Анализ развития идеала структуры научной теории / А.Ф. Кургаев – Кибернетика и вычислительная техника, 2003. – Вып. 139. – С. 50–63.
41. Андон Ф.И. Логические модели интеллектуальных информационных систем / Ф.И. Андон, Л.Е. Яшунин, В.И. Резниченко – К.: Наук. Думка, 1999. – 397 с.
42. Люггер Дж. Ф. Искусственный интеллект. Стратегии и методы решения сложных проблем / Дж. Ф. Люггер – Пер. с англ. – 4-е изд. – М.: Издательский дом «Вильямс», 2003. – 864 с.
43. Палагин А.В. К вопросу проектирования онтолого-управляемой ИС обработки ЕЯО / А.В. Палагин, Н.Г. Петренко // International Book Series “INFORMATIONSCIENCE&COMPUTING”, Varna, Bulgaria. – 2008. – № 2 – PP. 160–164.
44. Логические методы и формы научного познания / [Павлов В.Т., Руденко К.Ф., Семенов И.С. и др.]. – К.: Вища шк., 1984. – 208 с.
45. Искусственный интеллект. – В 3-х кн. Кн.2. Модели и методы: [Справочник / Научн. ред. Д.А. Поспелов]. – М.: Наука, 1990. – 304 с.
46. Гладун В.П. Процессы формирования новых знаний / В.П. Гладун – София: СД "Педагог 6", 1994. – 192 с.
47. Алгебро-логічний підхід до аналізу та обробки текстової інформації / [Палагін О.В., Кривий С.Л., Петренко М.Г., Бібіков Д.С.]. – Проблеми програмування. Спеціальний випуск. – 7-а міжнарод. наук.-практ. конф. з програмування “УкрПРОГ’2010”. – [Україна, Київ], 25-27 травня, 2010 р. – № 2,3. – С. 318–329.
48. Корпусна лінгвістика / [Широков В.А., Бугайов О.В., Грязнухіна Т.О. та ін.] – К.: Довіра, 2005. – 471 с.
49. Палагин А.В. Знание-ориентированные информационные системы с обработкой естественно-языковых объектов: основы методологии и архитектурно-структурная организация / А.В. Палагин, С.Л. Кривый, Н.Г. Петренко. – УсиМ, 2009. – № 3. – С. 42–55.

50. Гаврилова Т.А. Об одном подходе к онтологическому инжинирингу / Т.А. Гаврилова – Новости искусств. Интеллекта, 2005. – 3. – С. 25–30.
51. Uschold M. Ontologies: Principles, Methods and Applications / M.Uschold, M.Gruninger. KnowledgeEngineering Review 11(2),1996. – PP. 93–136.
52. Гаврилова Т.А. Онтологический инжиниринг / Т.А. Гаврилова. – Режим доступа: [http://www.kmtec.ru/publications/library/authors/ontolog\\_engeneering.shtml](http://www.kmtec.ru/publications/library/authors/ontolog_engeneering.shtml). – Дата доступа: 17.05.2011. – Название с экрана.
53. Клини С. Математическая логика / С. Клини – М.: Мир, 1973. – 480 с.
54. Куратовский К. Теория множеств / К. Куратовский, А. Мостовский. – М.: Мир, 1970. – 416 с.
55. Френкель А. Основания теории множеств / А. Френкель, И. Бар-Хиллел. – М.: Мир, 1966. – 638 с.
56. Курош А.Г. Лекции по общей алгебре / А.Г. Курош – М.: Наука, 1972. – 399 с.
57. Мальцев А.И. Алгебраические системы / А.И. Мальцев – М.: Наука, 1970. – 370 с.
58. Стол Р. Множества, логика, аксиоматическая теория / Р. Стол – М.: Просвещение, 1968. – 230 с.
59. Кривий С.Л. Дискретна математика: Вибр. питання: Навч. посіб. для студ. вищ. навч. закл. / С.Л. Кривий – К.: Вид. дім. “Кієво-Могилянська академія”, 2007. – 572 с.
60. Логический подход к искусственному интеллекту: От классической логики к логическому программированию / [Тейз А., Грибомон П., Луи Ж. и др.]; пер. с франц. – М.: Мир, 1990. – 429 с.
61. Ивлев Ю.В. Логика: учебник для вузов / Ю.В. Ивлев – М.: «Логос», 1997. – 272 с.
62. Рубашкин В.Ш. Представление и анализ смысла в интеллектуальных информационных системах / В.Ш. Рубашкин – М.: Наука, 1989. – 191 с.
63. Логический подход к искусственному интеллекту: От модальной логики к логике баз данных / [Тейз А., Грибомон П., Юлен Г. и др.]. – М.: Мир, 1998. – 492 с.
64. Широков В.А. Феноменологія лексикографічних систем / В.А. Широков. – К.: Наукова думка, 2004. – 327 с.
65. Лингвистический процессор для сложных информационных систем / [Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. и др.]. – М.: Наука, 1992. – 256 с.
66. К анализу естественно-языковых объектов / [Палагин А.В., Крытый С.Л., Величко В.Ю., Петренко Н.Г.]. – International Book Series



“INFORMATIONSCIENCE&COMPUTING”, BOOK Intelligent Processing, Varna, Bulgaria. – 2009. – № 9. – PP. 36–43.

67. Новые материалы: сб. научн. трудов /– [Научн. ред. Ю.С. Карабасов]. – М.: МИСИС, 2002 – 736 с.

68. Стяжин И.И. Становление идей математической логики / И.И. Стяжин. – М.: Наука, 1964. – 304 с.

69. Гуц А.К. Математическая логика и теория алгоритмов: учебное пособие / А.К. Гуц. – Омск: Издательство «Наследие». Диалог-Сибирь, 2003. – 108 с.

70. Новиков Ф.А. Дискретная математика для программистов / Ф.А. Новиков. – СПб.: Питер, 2000. – 304 с.

71. Робинсон Дж. Логическое программирование – прошлое, настоящее и будущее / Дж. Робинсон. – В кн.: Логическое программирование. М.: Мир, 1988.

72. Чёрч А. Введение в математическую логику / А. Чёрч. – М.: ИЛ, 1960. – 484 с.

73. Гаврилов А.В. Системы искусственного интеллекта: методические указания для студентов заочной формы обучения / А.В. Гаврилов. – АВТФ. – Новосибирск: Новосибирский государственный технический университет, 2004. – 59 с.

74. Искусственный интеллект: история развития. – Режим доступа: <http://library.by/shpargalka/belarus/computers/001/com-019.htm>. – Дата доступа: 17.05.2011. – Название с экрана.

75. Минский М. Фреймы и представление знаний / М. Минский. – М.: Энергия, 1979. – 150 с.

76. NCITS T2 (1998) Conceptual Graphs, A Presentation Language for Knowledge in Conceptual Schemas, Working draft of proposed American national standard, Document No. X3T2/96-008.

77. NCITS T2 (1998) Knowledge Interchange Format, Working draft of proposed American national standard, document – Режим доступа: <http://logic.stanford.edu/kif/dpans.html>). – Дата доступа: 11.03.2007. – Название с экрана.

78. Режим доступа: <http://www-ksl.stanford.edu/htw/dme/thermal-kb-tour/kif-ontology.lisp.html>. – Дата доступа: 12.11.2010.

79. Логика. – Режим доступа: <http://ru.wikipedia.org/wiki/логика.htm>. – Дата доступа: 09.04.2009. – Название с экрана.

80. Теория выбора и принятия решений / [Макаров И.М., Виноградская Т.М., Рубчинский А.А., Соколов В.Б.]. – М.: Наука, 1982. – 328 с.

81. Братко И. Программирование на языке Пролог для искусственного интеллекта: пер. с англ. / И. Братко. – М.: Мир, 1990. – 560 с.

82. Палагин А.В. Концептуальные графы и семантические сети в системах обработки естественно-языковой информации / Палагин А.В., Кривый С.Л., Петренко Н.Г. – Математические машины и системы, 2009. – №3. – С. 67–79.

83. Sowa J. F. Conceptual Graphs as a universal knowledge representation / Sowa J. F. – In: Semantic Networks in Artificial Intelligence, Spec. Issue of An International Journal Computers & Mathematics with Applications. (Ed. F. Lehmann), Vol.23, Number 2–5, 1992. –Part 1, – PP. 75–95.

84. Леонтьева Н.Н. Семантический словарь РУСЛАН как инструментальный компьютерного понимания / Н. Леонтьева, С. Семенова – М.: МГГИИ, 2003. – С. 41–46.

85. Леонтьева Н.Н. К теории автоматического понимания естественных текстов. Часть 1 // Моделирование системы "мягкого понимания" текста: информационно-лингвистическая модель / Н.Н.Леонтьева – М., МГУ, 2000. – 43 с.

86. Леонтьева Н.Н. К теории автоматического понимания естественных текстов. Часть 2 // Семантические словари: состав, структура, методика создания / Н.Н.Леонтьева – М.: МГУ, 2001. – 41 с.

87. Толковый словарь по искусственному интеллекту / [авт.-сост. Аверкин А.Н., Гаазе-Рапопорт М.Г., Поспелов Д.А.] – М.: Радио и связь, 1992. – 256 с.

88. Object-oriented modeling and design / J. Rumbaugh, M. Blaha, W. Premerlani et al. // Englewood Cliffs, 1991. – New Jersey: Prentice Hall. – 180 p.

89. Natalya F. Noy. 'Ontology Development 101: A Guide to Creating Your First Ontology' / F. Noy Natalya, L. McGuinness Deborah // Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March, 2001. – Режимдоступа:

[http://protege.stanford.edu/publications/ontology\\_development/ontology101.html](http://protege.stanford.edu/publications/ontology_development/ontology101.html). – Датадоступа: 22.02.2006. – Названиеэкрана.

90. Guarino N. Formal Ontology and Information Systems / N. Guarino In N. Guarino (ed.) Formal Ontology and Information Systems / FOIS'98, 6–8 June 1998, Trento, Italy: – IOS Press, Amsterdam, 1998. – PP. 3–15.

91. Онтологии и тезаурусы: модели, инструменты, приложения / [Добров Б.В., Иванов В.В., Лукашевич Н.В., Соловьев В.Д.]. – Интернет-университет информационных технологий – ИНТУИТ.ру, БИНОМ. Лаборатория знаний, 2009 – 176 с.

92. Методы и средства автоматизированного проектирования прикладной онтологии / [Добров Б.В., Лукашевич Н.В., Невзорова О.А., Федун Б.Е.]. – Изв. РАН. Теория и системы управления. М., 2004. – № 2. – С. 58–68.

93. Невзорова О.А. Онтолингвистические системы: методологические основы построения / О.А. Невзорова. – Научная сессия МИФИ-2007: сб. научн. трудов. Интеллектуальные системы и технологии. – М., 2007. – Том 3. – С. 84–85.

94. OntologiesforNLP: Introduction. – Режим доступа: <http://crl.nmsn.edu/Research/Projects/mik-ro/htmls/asis.paper-htmls.node1.html>. – Дата доступа: 17.03.2007. – Название с экрана.
95. TheMikrokosmosMashineTranslationProject. – Режим доступа: <http://crl.nmsn.edu/Research/Projects/mikro/htmls/asis.paper-htmls.node4.html>. – Дата доступа: 17.03.2007. – Название с экрана.
96. Философско-методологические основания системных исследований: системный анализ и системное моделирование: сб. научн. трудов / ВНИИ системных исследований. – М.: Наука, 1983. – 324 с.
97. Згуровский М.З. Системный анализ: проблемы, методология, приложения / М.З. Згуровский, Н.Д. Панкратова. – К.: Наук. Думка, 2005. – 744 с.
98. Методический комплекс по дисциплине “Современные проблемы науки” / Методология анализа предметных знаний. – Режим доступа: [http://old.ulstu.ru/people/SOSNIN/umk/Modern\\_Scientific\\_Problems/beloborodov/item\\_znan.htm](http://old.ulstu.ru/people/SOSNIN/umk/Modern_Scientific_Problems/beloborodov/item_znan.htm). – Дата доступа: 11.10.2007. – Название с экрана.
99. Палагін О. Міждисциплінарні наукові дослідження: оптимізація системно-інформаційної підтримки / О. Палагін, О. Кургаєв. – Вісник НАН України, 2009. – № 3. – С. 14–25.
100. WordNet: AnElectronicLexicalDatabase / Fellbaum, Christiane, ed. MITPress, Cambridge, MA, 1998. – 422 p.
101. MillerG.A. "WordNet: a lexical database for English" / G. A. Miller. – Communications of the ACM 38: 11, 1995. – PP. 39–41. – Режимдоступа: <http://www.cogsci.princeton.edu/-wn/>. – Датадоступа: 25.05.2007. – Названиесэкрана.
102. Широков В.А. Элементы лексикографії / В.А. Широков. – К.: Довіра, 2005. – 304 с.
103. Режим доступа: <http://www.isi.edu/naturallanguage/projects/ontologies.html>. – Дата доступа: 15.03.2010.
104. Philpot A. The Omega Ontology / A. Philpot, E. Novy, P. Pantel. – Information Sciences Institute of University of Southern California, 2005. – 8pp. – Режимдоступа: <http://omega.isi.edu/doc/>, <http://www.isi.edu/div3/div3/pubs/papers/philpot/2005omega.pdf>. – Датадоступа: 17.11.2007. – Названиесэкрана.
105. OpenCycDocumentation. – Режим доступа: <http://www.opencyc.org/doc>. – Дата доступа: 17.11.2005. – Название с экрана.
106. WonderWeb Deliverable D18: Ontology Library (final) / [Masolo C., Borgo S., Gangemi A., Guarino N., Oltramari A.]; Laboratory For Applied Ontology – ISTC–CNR, 2003. – 349pp. – Режимдоступа: <http://www.loa-cnr.it/Papers/D18.pdf>. – Датадоступа: 27.06.2006. – Названиесэкрана.
107. GiugleaA.KnowledgeDiscoveringusingFrameNet, VerbNetandPropBank / A. Giuglea, A. Moschitti. – 2004. – 6pp. – Режимоступа:

<http://olp.dfki.de/pkdd04/giuglea-final.pdf>. – Датадоступа: 17.12.2006. – Названиесэкрана.

108. Kingsbury P. PropBank: the Next Level of the TreeBank / P. Kingsbury, M. Palmer. – University of Pennsylvania, Department of Computer and Information Science – 2003. – 12pp. – Режимдоступа: [http://w3.msi.vxu.se/~rics/TLT2003/doc/kingsbury\\_palmer.pdf](http://w3.msi.vxu.se/~rics/TLT2003/doc/kingsbury_palmer.pdf). – Датадоступа: 17.01.2008. – Названиесэкрана.

109. Fillmore C.J. Frame semantics and the nature of language / C.J. Fillmore. – Annals of the New York Academy of Sciences, Volume 280, 1976. – PP. 20–32.

110. Baker C.F. The Berkeley FrameNet Project / C.F. Baker, C.J. Fillmore, J.B. Lowe // COLING–ACL, Montreal, Canada, 1998. – бр. – Режимдоступа: <http://framenet.icsi.berkeley.edu/~framenet/papers/acl98.pdf>. – Датадоступа: 11.05.2006. – Названиесэкрана.

111. FrameNet II FrameGrapher. – Режим доступа: <http://framenet.icsi.berkeley.edu/FrameGrapher>. – Дата доступа: 17.05.2011. – Название с экрана.

112. Standart Upper Ontology Knowledge Interchange Format. – Режимдоступа: <http://suo.ieee.org./suo-kif.html>. – Датадоступа: 17.05.2011. – Названиесэкрана.

113. The Mikrokosmos Machine Translation Project. – Режимдоступа: <http://crl.nmsu.edu/Research/Projects/mikro/htmls/asis.paper-htmls/node1.html>. – Датадоступа: 17.05.2011. – Названиесэкрана.

114. The Mikrokosmos Ontology. – Режим доступа: <http://crl.nmsu.edu/Research/Projects/mikro/htmls/asis.paper-htmls/node4.html>. – Дата доступа: 17.03.2007. – Название с экрана.

115. Маторин С.И. Системологическое исследование структуры системы категорий / Маторин С.И. – НТИ. Сер.2, 1997. – №3. – С. 3–7.

116. Палагін О.В. Архітектурно-онтологічні принципи розбудови інтелектуальних інформаційних систем / О.В. Палагін, М.Г. Петренко. – Математичні машини і системи, 2006. – №4. – С.15–20.

117. Экспертные системы для персональных компьютеров: методы, средства, реализации: Справочное пособие / [Крисевич В.С., Кузьмич Л.А., Шиф А.М. и др.]; – Минск: Выш. шк., 1990. – 197 с.

118. Guarino N. Some Ontological Principles for Disigning Upper Level Lexical Resources / N. Guarino. – First International Conference on Language Resources and Evaluation. – Granada, Spain, 1998. – 28–30 May. – 8pp. – Режимдоступа: <http://www.ladseb.pd.cnr.it/infor/ontology/ontology.html>. – Датадоступа: 17.05.2006. – Названиесэкрана.

119. Niles I. Linking Lexicons and Ontologies: Mapping WordNet to the suggested Upper Merged Ontology / I. Niles, A. Pease. – In proceedings of the 2003 International Conference on Information and Knowledge Engineering (IKE2003). – Las-Vegas, Nevada. – 2003, june 23–26. – брр. – Режимдоступа:

<http://home.earthlink.net/~adampease/professional/Niles-IKE.pdf>. – Дата доступа: 17.05.2008. – Название экрана.

120. Scheffczyk I. Linking FrameNet to the Suggested Upper Merged Ontology / I. Scheffczyk, A. Pease, M. Ellsworth. – 2006 – 9pp. – Режим доступа: <http://adampease.org/Articulate/publications/FOIS2006.pdf>. – Дата доступа: 27.02.2009. – Название экрана.

121. Chow I.C. Integration of Linguas Resources for Verb Classification: FrameNet, WordNet and SUMO / I.C. Chow, J.J. Webster // Fifth Mexican International Conference on Artificial Intelligence (MICAI'06). – 2006. – PP. 262–268. – Режим доступа: <http://csdl2.computer.org/persagen/DLabsToc.jsp?resourcePth=/dl/proceedings/micai/&toc=comp/proceedings/micai/2006/2722/00/272>. – Дата доступа: 17.11.2008. – Название экрана.

122. Марка Д. А. Методология структурного анализа и проектирования / Д.А. Марка, К. МакГоуэн. – М.: "МетаТехнология", 1993. – 239 с.

123. Методология функционального моделирования IDEF0. РД IDEF0 – 2000 // Госстандарт России. – Москва. – 2000. – 75 с. – Режим доступа: [www.nsu.ru/smk/files/idef.pdf](http://www.nsu.ru/smk/files/idef.pdf). – Дата доступа: 17.05.2006. – Название с экрана.

124. IDEF5 Method Report. – Armstrong Laboratory AL/HRGA Wright-Patterson Air Force Base, Ohio 45433. – 187p. – Режим доступа: [www.idef.com/pdf/Idef5.pdf](http://www.idef.com/pdf/Idef5.pdf). – Дата доступа: 17.05.2008. – Название экрана.

125. Gomez-Perez A. Ontological engineering: With Examples from the Areas of Knowledge Management, E-commerce and the Semantic Web / A.Gomez-Perez, O. M. Fernandez-Lopez, OCorcho (2004). – Springer, 2004. – 403 с.

126. Клещев А.С. Отношения между онтологиями предметных областей. Ч.1. / А. Клещев, И. Артемьева. – Информационный анализ, Выпуск 1, 2002. – С. 4–9.

127. Андреев А.М. Особенности проектирования модели и онтологии предметной области для поиска противоречий в правовых электронных библиотеках / А.М. Андреев, Д.В. Березкин, К.В. Симаков. – Режим доступа: [www.inteltec.ru/publish/articles/textan/RCDL2004.shtml](http://www.inteltec.ru/publish/articles/textan/RCDL2004.shtml). – Дата доступа: 12.12.2006. – Название с экрана.

128. Войшвилло Е.К. Понятие / Е.К. Войшвилло. – М.: МГУ, 1967. – 286 с.

129. SUO, (2001), TheIEEEStandardUpperOntologywebsite – Режим доступа: <http://suo.ieee.org>. – Дата доступа: 17.05.2006. – Название с экрана.

130. Гуревич И.Б. Тезаурус и онтология предметной области “Анализ изображений” / И.Б. Гуревич, Ю.О. Трусова // Всероссийская конф. с междунар. участием “Знания – Онтологии – Теории” (ЗОНТ–09). – Новосибирск: Институт математики им. С.Л. Соболева СО РАН, 2009. – 10 с. – Режим доступа: <http://www.math.nsc.ru/conference/zont09/reports/95Gurevich-Trusova.pdf>. – Дата доступа: 17.12.2010. – Название с экрана.

131. Martins A.F. Models for Representing Task Ontologies / A.F. Martins, R.A. Falbo. – Режим доступа: [http://nemo.inf.ufes.br/files/models\\_for\\_representing\\_task\\_ontologies\\_2008.pdf](http://nemo.inf.ufes.br/files/models_for_representing_task_ontologies_2008.pdf). – Дата доступа: 17.02.2010. – Название с экрана.

### Література до розділу 3

1. Палагин А.В. Знание-ориентированные информационные системы с обработкой естественно-языковых объектов: основы методологии и архитектурно-структурная организация / А.В. Палагин, С.Л. Кривый, Н.Г. Петренко. – УсиМ, 2009. – № 3. – С. 42–55.
2. Палагин А.В. Системно-онтологический анализ предметной области / А.В. Палагин, Н.Г. Петренко. – УСиМ, 2009. – № 4. – С. 3–14.
3. Рыков В.В. Управление знаниями / В.В. Рыков – Режим доступа: <http://rykkurc2.narod.ru/part2.doc>. – Дата доступа: 17.05.2011. – Название с экрана.
4. Демьянков В.З. Основы теории интерпретации и её приложения в вычислительной лингвистике / В.З. Демьянков. – М.: Изд.-во Моск. ун-та, 1985. – 76 с.
5. Палагин А.В. Системная интеграция средств компьютерной техники / А.В. Палагин, Ю.С. Яковлев. – Винница: УНІВЕРСУМ, 2005. – 680 с.
6. Широков В.А. Інформаційна теорія лексикографічних систем / В.А. Широков – К.: Довіра, 1998. – 331 с.
7. Рубашкин В.Ш. Представление и анализ смысла в интеллектуальных информационных системах / В.Ш. Рубашкин – М.: Наука, 1989. – 191с.
8. Лингвистический процессор для сложных информационных систем / [Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. и др.]. – М.: Наука, 1992. – 256 с.
9. Тестелец Я.Г. Введение в общий синтаксис / Я.Г. Тестелец. – М., РГГУ, 2001. – 798 с.
10. Штерн І.Б. Вибрані топіки та лексикон сучасної лінгвістики. Енциклопедичний словник / І.Б. Штерн. – Київ: АртЕк, 1998. – 336 с.
11. Замаруева И.В. Об одном подходе к компьютерному моделированию процесса понимания естественно-языковых текстов / И.В. Замаруева // : VI Междунар. конф. "ЗНАНИЕ–ДИАЛОГ–РЕШЕНИЕ", KDS-97, 15–20 сент. 1997г., Крым, Украина: материалы конф. – Ялта – С. 241–248.
12. Петренко М. Г. Особливості розробки знання-орієнтованого лінгвістичного процесора / Петренко М. Г. – Комп'ютерні засоби, мережі та системи, 2006. – №5. – С.18–22.
13. Палагін О.В. Модель категоріального рівня мовно-онтологічної картини світу / О.В. Палагін, М.Г. Петренко. – Математичні машини і системи, 2006. – №3. – С. 91–104.

14. Палагін О.В. Розбудова абстрактної моделі мовно-онтологічної інформаційної системи /О.В. Палагін, М.Г. Петренко. – Математичні машини і системи, 2007. – №1. – С. 42–50.
15. Попов Э.В. Общение с ЭВМ на естественном языке / Э.В. Попов. – М.: Наука, 1982. – 360 с.
16. Найханова Л.В. Основные аспекты построения онтологий верхнего уровня и предметной области / Л.В. Найханова. – Интернет-порталы: содержание и технологии, ФГУ ГНИИ ИТТ “Информатика”. – М.: Просвещение, 2005. – С. 452–479.
17. Палагин А.В. К решению основной задачи эмуляции / А.В. Палагин. – УсиМ, 1980. – №3. – С. 24–28.
18. Микропроцессорные системы обработки информации / [Палагин А.В., Денисенко Е.Л., Белицкий Р.И., Сигалов В.И.]. – Киев: Наук. Думка, 1993. – 352 с.
19. Палагин А.В. Архитектура онтолого-управляемых компьютерных систем / Палагин А.В. – Кибернетика и системный анализ, 2006 – №2. – С. 111–124.
20. Палагін О.В. Розбудова абстрактної моделі мовно-онтологічної інформаційної системи /О.В. Палагін, М.Г. Петренко. – Математичні машини і системи, 2007. – №1. – С. 42–50.
21. Палагин А.В. Реконфигурируемые вычислительные системы / А.В. Палагин, В.Н. Опанасенко. – К.: Просвіта, 2006. – 280 с.
22. Палагін О.В. Модель категоріального рівня мовно-онтологічної картини світу / О.В. Палагін, М.Г. Петренко. – Математичні машини і системи, 2006. – №3. – С. 91–104.
23. Андон Ф.И. Логические модели интеллектуальных информационных систем / Ф.И. Андон, Л.Е. Яшунин, В.И. Резниченко – К.: Наук. Думка, 1999. – 397 с.
24. Палагин А.В. Информационная модель обработки естественно-языковых текстов / А.В. Палагин, В.П. Гладун, Н.Г. Петренко и др. // “Knowledge-Dialogue-Solution” KDS’2007: XI-th Intern. conf. – Varna, Bulgaria, 2007. – vol.2. – PP. 525–527.
25. Петренко М. Г. Особливості розробки знання-орієнтованого лінгвістичного процесора / Петренко М. Г. – Комп’ютерні засоби, мережі та системи, 2006. – №5. – С.18–22.
26. Палагін О.В. Розбудова абстрактної моделі мовно-онтологічної інформаційної системи /О.В. Палагін, М.Г. Петренко. – Математичні машини і системи, 2007. – №1. – С. 42–50.
27. Попов Э.В. Общение с ЭВМ на естественном языке / Э.В. Попов. – М.: Наука, 1982. – 360 с.
28. Найханова Л.В. Основные аспекты построения онтологий верхнего уровня и предметной области / Л.В. Найханова. – Интернет-порталы: содержание и технологии, ФГУ ГНИИ ИТТ “Информатика”. – М.: Просвещение, 2005. – С. 452–479.

29. Палагин А.В. К решению основной задачи эмуляции / А.В. Палагин. – УсиМ, 1980. – №3. – С. 24–28.
30. Рубашкин В.Ш. Представление и анализ смысла в интеллектуальных информационных системах / В.Ш. Рубашкин – М.: Наука, 1989. – 191с.
31. Палагин А.В. Организация и функции "языковой" картины мира в смысловой интерпретации ЕЯ – сообщений //InformationTheoriesandApplication. – 2000. – Vol. 7, №4. С.155-163.
32. Сагатовский В.Н. Основы систематизации всеобщих категорий / Сагатовский В.Н. – Издательство Томского университета, Томск, 1973. – 432 с.
33. Палагін О.В., Петренко М.Г. Модель категоріального рівня мовно-онтологічної картини світу //Математичні машини і системи. – 2006. - №3. - С.91-104.
34. Гладун В.П., Величко В.Ю. Конспектирование естественно-языковых текстов. Proceedings of the XI-th International Conference “Knowledge-Dialogue-Solution”(KDS’2005).- Varna, Bulgaria.-2005.- pp.344-347 vol.2.
35. Гладун В.П. Процессы формирования новых знаний. – София: СД "Педагог 6", 1994. – 192 с.
36. ПетренкоМ.Г.Особливості розробки знання-орієнтованого лінгвістичного процесора. – Комп’ютерні засоби, мережі та системи. – 2006, №5. – С.18-22.
37. Палагін О.В., Петренко М.Г. Архітектурно-онтологічні принципи розбудови інтелектуальних інформаційних систем // Математичні машини і системи. – 2006. - №4. - С.15-20.
38. Ножов И.М. Морфологическая и синтаксическая обработка текста(модели и программы). Автореферат диссертации на соискание ученой степени кандидата технических наук.-М.-2003. – 19с.
39. Тестелец Я.Г. Введение в общий синтаксис. – М., РГГУ, 2001.
40. Попов Э.В. Общение с ЭВМ на естественном языке. – М.: Наука, 1982.
41. Шереметьева С.О., Ниренбург С. Эмпирическое моделирование в вычислительной морфологии//НТИ, №7, 1996.
42. Про один підхід до аналізу та розуміння природномовних об’єктів / [Палагін О.В., Світла С.Ю., Петренко М.Г., Величко В.Ю.]. – Комп’ютерні засоби, мережі та системи, 2008, №7. – С. 128–137.
43. Найханова Л.В. Основные аспекты построения онтологий верхнего уровня и предметной области // В сборнике научных статей “Интернет порталы: содержание и технологии”. Выпуск 3. / Редкол.: А.Н. Тихонов (пред.) и др.; ФГУ ГНИИ ИТТ “Информатика”. – М.: Просвещение, 2005. – С. 452-479.
44. Белоногов Г.Г., Кузнецов Б.А. Языковые средства автоматизированных информационных систем. М.: Наука, 1983.



45. B. Dobrov, N. Loukachevitch, O. Nevzorova. The technology of new domains' ontologies development // Proceedings of the X-th International Conference "Knowledge-Dialogue-Solution" (KDS'2003).- Varna, Bulgaria.- 2003.- pp.283-290.

46. Виталий Величко, Павел Волошин, Светлана Свитла. Автоматизированное создание тезауруса терминов предметной области для локальных поисковых систем. "Knowledge – Dialogue – Solution" International Book Series "INFORMATION SCIENCE & COMPUTING", Number 15. – FOI ITHEA Sofia, Bulgaria. - 2009. – pp.24-31.

47. Засобидоступу до джерел знань, та їх використання в навчальному процесі педагогічного навчального закладу : Методичні рекомендації / За редакцією члена-кореспондента НАН України С. О. Довгого і канд. техн. наук О.Є. Стрижака. - 2-ге вид., доповн. - К.: Інститут обдарованої дитини, 2012. - 192с.

#### Література до розділу 4

1. Згуровский М.З., Панкратова Н.Д. Системный анализ. Проблемы, методология, приложение. – К: Наук. думка, 2005. – 744 с.

2. Згуровский М.З., Доброногов А.В., Померанцева Т.Н. Исследование социальных процессов на основе методологии системного анализа. –К.: Наукова думка, 1997. – 224 с.

3. Checland P.B. Soft systems methodology: an overview // J. Appl. Syst. Anal. – 1988. –15. – P.27–36.

4. Плотинский Ю.М. Математическое моделирование динамики социальных процессов. – М.: Изд-во Моск. ун-та, 1992. –133 с.

5. Перегудов Ф.И., Тарасенко В.П. Введение в системный анализ. – М.: Высш. шк., 1989. – 367 с.

6. Ротштейн А.П. Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети. – Винница: УНИВЕРСУМ, 1999. –320 с.

7. Беляев В. И. Теория сложных систем. – К: Наукова думка, 1976. – 156 с.

8. Удовиченко Л.В. Модели и алгоритмы планирования и управления в организационно-технических системах в условиях неопределенности: Дис. канд. техн. наук: 05.13.06/Херсонский гос. ун-т. – Херсон, 1997. –167 с.

9. Жлуктенко С.В. Розробка і дослідження формальних процедур системного аналізу для вирішення задач моделювання і прийняття рішень в складних системах з невизначеністю опису: Дис...канд. техн. наук:05.13.02/ Національний технічний університет України «КПІ».–К., 1996. –219 л.

10. Молчанов А. А. Моделирование и проектирование сложных систем. - К. , Выща шк., 1988. – 359 с.

11. Моисеев Н.Н. Математические задачи системного анализа. М: Наука, 1981.- 488с.

12. Глоба Л.С., Терновой М.Ю. Комплексная оценка сложного объекта в условиях неопределенности. // Электроника и связь. – 2003. – №18. – С 65-68.
13. Толковый словарь по искусственному интеллекту / А.Н. Аверкин, М.Г. Гаазе-Рапопорт, Д.А. Поспелов. – М.: Радио и связь, 1992. – 256 с.
14. Основы системного анализа и проектирования АСУ: Учеб. пособие / А.А. Павлов, С.Н. Гриша, В.Н. Томашевский и др.; Под ред. А.А. Павлова. – К.: Выща шк., 1991. – 367 с.
15. Негойце К. Применение теории систем к проблемам управления. М: Мир, 1981. – 179с.
16. Панкратова Н.Д., Пугач О.В. Рациональный выбор структуры многоуровневых иерархических технических систем // III Междунар. науч. – практ. конф. «Системный анализ и информ. технологии». К.: НТУУ «КПИ», 2001. – С. 117-119.
17. Глушков В.М. Основы безбумажной информатики. – М.: Наука, 1982. – 552 с.
18. Дейнека В. С., Сергиенко И.В. Оптимальное управление неоднородными системами. – К.: Наук. думка, 2003. – 505 с.
19. Кунцевич В.М., Лычак М.М. Синтез оптимальных и адаптивных систем управления: Игровой подход. – К.: Наук. думка, 1985. – 248 с.
20. Зайченко Ю.П. Основы проектирования интеллектуальных систем. Навчальний посібник. – К.: «Слово», 2004. – 352 с.
21. Зайченко Ю.П. Исследование операций: Учебник. – 6 из., перераб. и доп. – К.: «Слово», 2003. – 688 с.
22. Системный анализ в исследовании сложных физических процессов и полей. / М.З. Згуровский, А.М. Демченко, А.Н. Новиков, И.И. Коваленко – К.: 1993. – 37с. – (Препр./АН Украины. Ин-т кибернетики им. В.М. Глушкова; 93-30).
23. Павлов А.А. Алгоритмическое обеспечение сложных систем управления. – К.: Выща шк., 1986. – 166 с.
24. Павлов А.А., Теленик С.Ф. Информационные технологии и алгоритмизация в управлении. – К.: Техніка, 2002. – 344с.
25. Теленик С.Ф. Интеллектуальные системы и логики высших порядков // Радиоэлектроника, информатика, управление. – 1999. – № 1. – С. 96-105.
26. Гончаров В.Н., Колосов А.Н., Дибнис Г.И. Оперативное управление производством: опыт разработки и совершенствования систем. – М.: Экономика, 1987. – 120 с.
27. Тищенко Н.М. Введение в проектирование систем управления. – М.: Энергоатомиздат, 1986. – 248 с.
28. Емельянов С.В. Управление гибкими производственными системами. Модели и алгоритмы. – Л.: Машиностроение; Берлин: Техник, 1987. – 364 с.

29. Многоуровневая система оперативного управления ГПС в машиностроении /С.А. Соколицын, В.А. Дуболазов и др.; Под ред. С.А. Соколицына. – Новосибирск: Политехника, 1991. – 208 с.
30. Сергиенко И.В, Математические модели и методы решения задач дискретной оптимизации. – К.: Наук. думка, 1988. – 472 с.
31. Саати Т., Керне К. Аналитическое планирование. Организация систем: Пер с англ. - М.: Радио и связь, 1991. –224 с.
32. Системы автоматизированного планирования и диспетчирования групповых производственных процессов / А.А. Павлов, З. Банашак, С.Н. Гриша, Е.Б. Мисюра; Под ред. А.А. Павлова. – К.: Техніка, 1990. – 180 с.
33. Планирование и управление в автоматизированном производстве / В.В. Шкурба, С.А. Белецкий, Т.П. Почасова и др.; Под ред. В.В. Шкурбы. – К.: Наук. думка, 1985. – 224 с.
34. Люгер Дж. Ф. Искусственный интеллект: стратегии и методы решения сложных проблем, 4-е издание. – М.: «Вильямс», 2003. – 864 с.
35. Корнеев В.В., Гарев А.Ф., Васютин С.В., Райх В.В. Базы данных. Интеллектуальная обработка информации. М.: Нолидж, 2000. – 351с.
36. Ефимов Е. И. Решатели интеллектуальных задач. - М.: Наука, 1982. – 320с.
37. Уоссермен Ф. Нейрокомпьютерная техника. М.:Мир, 1992. – 240с.
38. Jain A.K., Mohiuddin J.M. Artificial Neural Networks: A Tutorial. Computer, March, 1996, p. 31-44.
39. Lawrence J. Introduction in Neural Networks: Design, Theory and Applications. California Scientific Software. 1994. 423p.
40. Lawrence S., Giles C., Tsoi A. Lessons in Neural Network Training: Overfitting May be harder than Expected, Proceeding of the Fourteenth National Conference on Artificial Intelligence, AAAI-97, AAAI Press, Menlo Park, California, 1997, pp. 540-545.
41. Байдык Т.Н. Нейронные сети и задачи искусственного интеллекта. – К.: Наук. думка, 2001. – 263 с.
42. Куссуль Э. М. Ассоциативные нейроподобные структуры. – К.: Наук. думка, 1992. –144с.
43. Заде Л. А. Основы нового подхода к анализу сложных систем и процессов принятия решений. - В кн.: Математика сегодня. – М.: Знание, 1974. – с. 5-49.
44. Заде Л. А. Понятие лингвистической переменной и его применение к принятию приближенных решений. – М: Мир, 1976. – 165 с.
45. Кофман А. Введение в теорию нечетких множеств.- М.: Радио и связь, 1982. – 432 с.
46. Беллман Р., Заде Л.А. Принятие решений в расплывчатых условиях - В кн.: Вопросы анализа и процедуры принятия решений. – М.: Мир, 1976. – С. 172-215.

47. Борисов А. Н., Алексеев А.В., Меркурьева Г.В. Обработка нечеткой информации в системах принятия решений. – М.: Радио и связь, 1989.– 304 с.
48. Зиман Э., Бьюменен О. Толерантные пространства и мозг. - В кн.:На пути к теоретической биологии. М.: Мир, 1970.– С. 134-144.
49. Ивахненко А.Г. Долгосрочное прогнозирование и управление сложными системами. – К.: Техника, 1975. –312с.
50. Поспелов Д. А. Ситуационное управление, теория и практика. - М.: Наука. - Гл. ред. физ.-мат лит., 1986 – 288с.
51. Поспелов Д. А. Логико-лингвистические модели в системах управления. – М.: Энергоатомиздат, 1981. – 190с.
52. Shortliffe E. MYCIN: Computer-Based Medical Consultations. - American Elsevier, New York, 1976.
53. Buchanan B. G., Shortliffe E. H. Rule-Based Expert System - The MYCIN Experiments of the Stanford Heuristic Programming Project. - Addison-Wesley. - 1984.
54. Экспертные системы. Принципы работы и примеры: Пер. с англ, / А. Брукинг, П. Джоне, Ф. Кокс и др.; Под ред. Р.Форсайта. - М. . Радио и связь, 1987. – 224с.
55. Демченко А. М.,Жлуктенко С. В.,Скуцкий А. В., Власенко Ю.Н.Применение байесовского вывода в задачах диагностики //Вестник КПИ. Техническая кибернетика. –1990. – Вып. 14. – С. 34-37.
56. Зельнер А. Байесовские методы в эконометрии. - М.: Финансы и статистика,1980.– 438 с.
57. Хей Дж. Введение в методы байесовского статистического вывода. - М.: Финансы и статистика, 1987. -335с.
58. Dempster A. P. Upper and lower probabilities induced by a multivalued mapping. //Ann. Math. Statis., 1967 v. 38, p. 325-333.
59. Shafer G. A Mathematical Theory of Evidence. -1976 - Princeton University Press, Princeton N. J.
60. Zadeh L.A. Fuzzy logic and approximate reasoning //Synthese. 1975, N30
61. Zadeh L. A. Theory of fuzzy sets. In:Encyclopedia of ComputerScienceand Technology(J.Belzer, A.Holzman, A Kent, eds.), Maroel Deckker, New-York,1977.
62. Zadeh L. A Fuzzy Sets as a Basic for a Theory Possibility. Fuzzy Sets and Systems, 1978, No. I, pp. 3-28.
63. ДюбуаД., ПрадаА. Теориявозможностей. Приложения к представлению знаний в информатике: Пер. с фр. - М.: Радио и связь, 1990 – 288с.
64. Прикладные нечеткие системы: Пер с япон. /К. Асаи, Л Батада, С. Иван и др. ; под ред. Т. Терано, К. Асаи, М. Сугэно. – М.:Мир, 1993. – 368с.

65. Справочник по прикладной математике. Том 2: Пер. с англ. /Подред.Э. Ллойда,У. Лидермана,С. А. Айвазяна, Ю. Н. Тюриня. – М.: Финансы и статистика, 1990. – 526 с.
66. Mamdani E.H., and Assiliani S. An experiment in linguistic synthesis with a fuzzy logic controller //International Journal of Man-Machine Studies, 1975, N 7, P. 1-13.
67. Мелихов А.Н., Бернштейн Л.С., Коровин С. Я. Ситуационные советующие системы с нечеткой логикой. – М.:Наука. Гл. ред. физ.-мат. лит., 1990.– 272с.
68. Малышев Н.Г., Берштейн Л.С., Боженюк А.Б. Нечеткие модели для экспертных систем в САПР. – М.: Энергоатомиздат, 1991. –136с.
69. Принятие решений при нечетких основаниях. I. Универсальная шкала // И.В. Ежкова, Д. А. Поспелов. Изв. АН СССР. Техническая кибернетика. Кб, 1977.– С. 3-11.
70. Принятие решений при нечетких основаниях. II. Схема вывода. // И. В. Ежкова, Д. А. Поспелов. Изв. АН СССР. Техническая кибернетика. N2, 1978. – С. 5-11.
71. Ринке Д. Б. Эвристический подход к обобщенному календарному планированию производства с использованием лингвистических переменных: методология и применение. – В кн. Нечеткие множества и теория возможностей. Последние достижения. /Под ред. Р. Р. Ягера. – М.: Радио и связь, 1986. – 408с
72. Гудмен И. Нечеткие множества как классы эквивалентности случайных множеств. В сб.: Нечеткие множества и теория возможностей. М.: Радио и связь, 1986.– с. 241-264.
73. Альбеверьо С., Фенстад И., Хеэг-Крон Р., Линдстем Т. Нестандартные методы в стохастическом анализе и математической физике: Пер. с англ.- М.: Мир, 1990. – 616с.
74. Нечеткие множества и теория возможностей. Последние достижения. – М.: Радио и связь, 1986.– 408с.
75. Герасимов Б.М., Грабовский Г.Г., Рюмшин Н.А. Нечеткие множества в задачах проектирования, управления и обработки информации. –К.: Техніка, 2002. –140 с.
76. Мітюшкін Ю. І. Синтез і настройка баз нечітких знань для моделювання багатовимірних залежностей: Дис... канд. техн. наук: 01.05.02 / Вінницький держ. техн. ун-т. – Вінниця, 2001. – 160 арк.
77. Штовба С.Д. Введение в теорию нечетких множеств и нечеткую логику. [Электронный ресурс] [matlab.exponenta.ru/fuzzylogic/book1/](http://matlab.exponenta.ru/fuzzylogic/book1/)
78. Дудка Т.М. Моделі побудови експертних систем з асоціативними моделями подання знань: Автореф. дис... канд. фіз.-мат. наук : 01.05.03 / НАН України, Ін-т кібернетики ім. В.М. Глушкова. – К., 2001. – 16 с.
79. Лавров О.А. Моделювання гнучких дискретно-подійних систем на основі методів з комбінованою семантикою : Дис...д-ра техн.. наук : 05.13.20 / Національний технічний університет України «КПІ». – К., 2000. – 367 арк.

80. Ершова О.Л. Моделі, методи та засоби інформаційної технології прийняття управлінських рішень в соціально-економічних системах : Дис... канд.. екон. наук : 08.03.02 / НАН України, Міжнародний науково-навчальний центр інформаційних технологій і систем. – К., 1999. – 114 л.
81. Support learning mechanism for fuzzy rule-based modeling: a new approach // J.-H. Ciang, P.-Y. Hao. IEEE Transactions on Fuzzy Systems, vol. 12, No. 1, February 2004. pp. 1-12.
82. Fuzzy repertory table: a method for acquiring knowledge about input variables to machine learning algorithm // J.J. Castro-Schez, J.L. Castro, J.M. Zurita // IEEE Transactions on Fuzzy Systems, vol. 12, No. 1, February 2004. pp. 123-139.
83. Фу К.С., Исидзука. М., Яс. Дз. Т. П. Применение нечетких множеств для оценки устойчивости строительных конструкции при землетрясениях. В кн. Нечеткие множества и теория возможностей. Последние достижения. /Под ред. Р. Р. Ягера. – М.: Радио и связь, 1986. – С. 312-332.
84. Гуверне Ж. , Эме С. , Санчес Э. Приложение нечеткой логики к медицинской генетике. - В кн. Нечеткие множества и теория возможностей. Последние достижения. /Под ред. Р.Р. Ягера. – М.: Радио и связь, 1986. – С. 312-332.
85. Леунг И. Разделение на торговые зоны в нечетких условиях. - В кн. Нечеткие множества и теория возможностей. Последние достижения /Под ред. Р.Р. Ягера. – М.: Радио и связь, 1986. – С. 312-332.
86. Гвоздик А. А. Решение нечетких уравнений. //Известия АН СССР. Техническая кибернетика N5, 1984. – С. 176-183.
87. Основы логико-информационного моделирования сложных геосистем //Беляев В. И. , Худошина М. Ю. , Отв. ред. А.Г. Ивахненко; АН УССР. Мор. гидрофиз. ин-т. - К.: Наук. думка, 1933. – 160 с.
88. Алтунин А.Е., Семухин М.В. Модели и алгоритмы принятия решений в нечетких условиях: Монография. Тюмень: Издательство Тюменского государственного университета, 2000. –352 с.
89. Алтунин А.Е., Востров Н.Н. Оптимизация многоуровневых иерархических систем на основе теории размытых множеств и методов самоорганизации. В сб.: «Проблемы нефти и газа Тюмени», Тюмень, вып. 42, 1979.– С. 68-72.
90. Алиев И.М. Об одной нечеткой модели надежности информационных систем. // Автоматика и вычислительная техника, Рига: Институт электроники и вычислительной техники, 2004. –№ 3. С. 64-71.
91. Глоба Л.С., Терновой М.Ю. Основні підходи до побудови інформаційно-аналітичної системи УДАІ УМВС України. //Електроніка і зв'язь. – 2002. – №17. – С. 97-104.
92. Глоба Л.С., Терновой М.Ю. Интеллектуальная обработка информации в производственных системах //Труды четвертой

международной конференции «Современные информационные и электронные технологии» (СИЭТ 2003). – Одеса: ОНПУ. – 2003. – С. 98.

93. Jen-Christophe Buisson, Alexandre Garel. Balancing meals using fuzzy arithmetic and heuristic search algorithms. //IEEE Transactions on Fuzzy Systems, Vol. 11, No. 1, February 2003 pp.68-78.

94. Pedrycz W., Reformat M. Evolutionary fuzzy modeling //IEEE Transactions on Fuzzy Systems, Vol. 11, No. 5, October 2003 pp.652-665.

95. Вознюк О.В. Нейронная сеть для параллельной и последовательной обработки нечетких правил // «Электроника и связь», Київ: Поліграфічний центр НДІ прикладної електроніки, ФЕЛ НТУУ «КПІ», 2001.- №10. С. 75-77.

96. Вознюк О.В. Объединение нейронной сети и нечеткого контроллера // «Электроника и связь», Київ: Поліграфічний центр НДІ прикладної електроніки, ФЕЛ НТУУ «КПІ», 2000.-№9. С. 24-27.

97. Nauck D., Kruse R. How the Learning of Rule Weights Affects the Interpretability of Fuzzy Systems. In Proc. IEEE International Conference on Fuzzy Systems 1998 (FUZZ-IEEE'98), Anchorage, AK, May 4-9, 1998, pp. 1235-1240.

98. Nauck D., Kruse R. NEFCLASS – A Neuro-Fuzzy Approach for the Classification of Data. In K. M. George, Janice H. Carrol, Ed Deaton, Dave Oppenheim and Jim Hightower, editors, Applied Computing 1995. Proc. 1995 ACM Symposium on Applied Computing, Nashville, Feb. 26-28, pages 461–465, ACM Press, New York, February 1995.

99. Nauck D., Kruse R. New learning strategies for NEFCLASS. In Proc. Seventh International Fuzzy Systems Association World Congress IFSA'97, volume IV, pages 50–55, Prague, 1997

100. [idisy.iae.nsk.su/fuzzy\\_book/](http://idisy.iae.nsk.su/fuzzy_book/)

101. Саати Т. Математические модели конфликтных ситуаций. – М.: Сов. радио.–1977.–304 с.

102. Ternovoy M. Modified Fuzzy Intelligence System // Proceedings of the International Conference TCSET'2006. – Lviv-Slavsko, Ukraine. – pp. 420-422.

103. Дьяконов В., Круглов В. Математические пакеты расширения MATLAB. Специальный справочник. – СПб: Питер, 2001 – 480 с.

104. Терновой М.Ю. Використання апарату нечіткої логіки для аналізу інформації в розподілених корпоративних системах // Труды 6-й Междунар.науч.-практ. конф «Современные информационные и электронные технологии» (СИЭТ 2005). – Одеса: ОНПУ.– 2005 . – С.163.

105. Маркушевич Н.С. Автоматизированная система диспетчерского управления. –М.: Энергоатомиздат, 1986. – 136с.

106. Мартин Дж. Вычислительные сети и распределенная обработка данных. Вып. 1. – М.: Финансы и статистика, 1985. – 256с.

107. Мелихов А.Н. и др. Лингвистический терминальный комплекс. Модели выбора альтернатив в нечеткой среде. Рига: РПИ, 1984. – С.140-142.

108. Норвич А.М., Турксен И.Б. Построение функций принадлежности. В сб.: Нечеткие множества и теория возможностей. – М.: Радио и связь, 1986. – С.64-71.

109. Норвич А.М., Турксен И.Б. Фундаментальное измерение нечеткости. В сб.: Нечеткие множества и теория возможностей. – М.: Радио и связь, 1986. – С.54-64.

110. Зуенков М.А. Приближение характеристических функций нечетких множеств. Автоматика и телемеханика, N 10, 1984.– С.138-149.

111. Борисов А.Н. и др. Модели принятия решений на основе лингвистической переменной. – Рига: Зинатне, 1982. – 256 с.

112. Заде Л.А. Размытые множества и их применение в распознавании образов и кластер-анализе. - В сб.: Классификация и кластер. – М.: Мир, 1980.– С. 208-247.

113. Орловский С.А. Проблемы принятия решений при нечеткой исходной информации. – М.: Наука, 1981.– 203с.

114. Фишберн П. Теория полезности для принятия решений. – М.: Наука, 1978. – 352 с.

115. Prade H. A computational approach to approximate and plausible reasoning with applications to expert systems. "IEEE Trans. Pattern Anal. and Mach. Intel.", 1985, N3, p.260-283.

116. Ягер Р.Р. Множества уровня для оценки принадлежности нечетких подмножеств. В сб.: Нечеткие множества и теория возможностей. М: Радио и связь, 1986. – С.71-78.

117. Bellman R., Kalaba K., Zadeh L.A. Abstraction and pattern classification. J.Math. Anal. and Appl., v.13, No1, Jan, 1966.

118. Xiao-Jun Zeng; Singh, M.G Approximation theory of fuzzy systems-MIMO case. Fuzzy Systems, IEEE Transactions on. Volume 3, Issue 2, May 1995 Page(s):219 - 235

#### **Література до розділу 5**

1. Ротштейн А.П. Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети. – Винница: УНИВЕРСУМ, 1999. –320 с.

2. Глоба Л.С., Терновой М.Ю. Оптимизация использования базы знаний экспертной системы введением иерархии правил. // Труды седьмой международной научно-практической конференции «Современные информационные и электронные технологии» (СИЭТ 2006). – Одеса: ОНПУ. – 2006 . – С.163.

#### **Література до розділу 6**

1. Ротштейн А.П. Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети. – Винница: УНИВЕРСУМ, 1999. –320 с.

2. Зайченко Ю.П. Основы проектирования интеллектуальных систем. Навчальний посібник. – К.: «Слово», 2004. – 352 с.



3. Герасимов Б.М., Грабовский Г.Г., Рюмшин Н.А. Нечеткие множества в задачах проектирования, управления и обработки информации. –К.: Техніка, 2002. –140 с.
4. Мітюшкін Ю. І. Синтез і настройка баз нечітких знань для моделювання багатовимірних залежностей: Дис... канд. техн. наук: 01.05.02 / Вінницький держ. техн. ун-т. – Вінниця, 2001. – 160 арк.
5. Pedrycz W., Reformat M. Evolutionary fuzzy modeling //IEEE Transactions on Fuzzy Systems, Vol. 11, No. 5, October 2003 pp.652-665.
6. Nauck D., Kruse R. NEFCLASS – A Neuro-Fuzzy Approach for the Classification of Data. In K. M. George, Janice H. Carrol, Ed Deaton, Dave Oppenheim and Jim Hightower, editors, Applied Computing 1995. Proc. 1995 ACM Symposium on Applied Computing, Nashville, Feb. 26-28, pages 461–465, ACM Press, New York, February 1995.
7. Nauck D., Kruse R. New learning strategies for NEFCLASS. In Proc. Seventh International Fuzzy Systems Association World Congress IFSA'97, volume IV, pages 50–55, Prague, 1997
8. Xiao-Jun Zeng; Singh, M.G Approximation theory of fuzzy systems-MIMO case. Fuzzy Systems, IEEE Transactions on. Volume 3, Issue 2, May 1995 Page(s):219 - 235
9. Setnes M., Babuska R., KaymakU., van Nauta Lemke H.R. (1998) Similarity measures in fuzzy rule base simplification. IEEE Trans. SMC-B 28, P. 376-386.
10. Roubos J.A., Setnes M., Abonyi J. Learning Fuzzy Classification Rules From Data. [Електроний ресурс]. <http://www.fmt.vein.hu/softcomp/>

### **Література до розділу 7**

1. Глоба Л.С., Терновой М.Ю. Разработка универсальной компоненты анализа информации в распределенной корпоративной системе. // Труды пятой международной научно-практической конференции «Современные информационные и электронные технологии» (СИЭТ-2004).–Одесса: ОНПУ. – 2004.– С. 150.
2. Терновой М.Ю. Гибкая технология создания программных модулей аналитической обработки информации // Труды 16-й международной конференции Крымико'2006.– Севастополь:Вебер.– 2006. – С. 400-401.
3. Ільченко М.Ю., Терещенко П.І., Глоба Л.С., Терновой М.Ю., Лисенко А.М. Технології Intranet в задачах аналітичної обробки інформації та підтримки прийняття управлінських рішень щодо підвищення безпеки дорожнього руху. // Безпека дорожнього руху України. – 2004. – № 1-2. – С.148-160.
4. Коломієць С.Г., Глоба Л.С., Прокопенко Ю.М., Христич М.Й., Терновой М.Ю. Програмні технології побудови розподіленої корпоративної системи департаменту Державтоінспекції МВС України. // Безпека дорожнього руху України. – 2002. – № 2 (13). – С. 21-24.
5. Глоба Л.С., Терновой М.Ю. Проблемы и технологии построения информационных ресурсов корпорации. // Материалы III международной

научно-технической конференции «Электронные ресурсы: проблемы формирования, обработки, распространения и использования- 2002». – К.: УкрИНТЭИ. – 2002. – С. 34.

6. Глоба Л.С., Терновой М.Ю. Створення інформаційних ресурсів корпорації: архітектурні рішення та технології. // Науково-технічна інформація. – 2003. – №1. – С.19-22.

7. Коннолли Т., Бегг К., Страчан А. Базы данных: проектирование, реализация и сопровождение. Теория и практика. – М.:«Вильямс», 2000. – 1120 с.

8. Зайченко Ю.П. Основы проектирования интеллектуальных систем. Навчальний посібник. – К.: «Слово», 2004. – 352 с.

9. Себеста Р.У. Основные концепции языков программирования, 5-е изд.: Пер. с англ. – М.: «Вильямс», 2001. – 672 с.

10. Дж. Д. Ульман, Дж. Уидом. Введение в системы баз данных. – М.: Издательство «Лори», 2000. – 374 с.

11. Касьянов В. Н., Евстигнеев В. А. Графы в программировании: обработка, визуализация и применение. – СПб.: БХВ-Петербург, 2003.– 1104 с.

12. Андерсон Дж. А. Дискретная математика и комбинаторика: Пер. с англ. – М. : «Вильямс», 2003. – 960 с.

13. Оре О. Теория Графов. –2-е изд. – М.: Наука, 1980. – 336 с.

14. Баканов А.А., Глоба Л.С., Терновой М.Ю Алгоритм динамического формирования SQL-запроса в системах аналитической обработки информации. // Электроника и связь. – 2005. – №27. – С. 65-68.

### **Література до розділу 8**

1. Гладун В.П. Процессы формирования новых знаний / В.П. Гладун – София: СД "Педагог 6", 1994. – 192 с.

2. Про один підхід до аналізу та розуміння природномовних об'єктів / [Палагін О.В., Світла С.Ю., Петренко М.Г., Величко В.Ю.]. – Комп'ютерні засоби, мережі та системи, 2008, №7. – С. 128–137.

3. Найханова Л.В. Основные аспекты построения онтологий верхнего уровня и предметной области / Л.В. Найханова. – Интернет–порталы: содержание и технологии, ФГУ ГНИИ ИТТ “Информатика”. – М.: Просвещение, 2005. – С. 452–479.

4. Онтологии и тезаурусы: модели, инструменты, приложения / [Добров Б.В., Иванов В.В., Лукашевич Н.В., Соловьев В.Д.]. – Интернет-университет информационных технологий – ИНТУИТ.ру, БИНОМ. Лаборатория знаний, 2009 – 176 с.

5. Величко В.Ю. Автоматизированное создание тезауруса терминов предметной области для локальных поисковых систем / Виталий Величко, Павел Волошин, Светлана Свитла // “Knowledge – Dialogue – Solution” InternationalBookSeries “INFORMATIONSCIENCE&COMPUTING”,

SupplementstotheInternationalJournal“INFORMATIONTECHNOLOGIES&KNOWLEDGE” № 14. – Vol. 3,2009 – FOITHEASofia, Bulgaria. – 2009. –pp. 24–31

6. Стрижак О.Є. – Комп’ютерні тезауруси як технологічна платформа створення авторських методик викладання предметних дисциплін// Актуальні проблеми психології: Психологічна теорія і технологія навчання/ за ред.. С.М. Максименко, М.Л.Смульсон. – К.: Вид-во НПУ імені М.П.Драгоманова, 2009,- Т.8, вип..6. с.259-266.

7. Л. І. Мацько "Тарас Шевченко – основоположник нової української літературної мови" // Українська мова в освітньому просторі: навчальний посібник для студентів-філологів освітньо-кваліфікаційного рівня "магістр". — К.: Вид-во НПУ імені М. П. Драгоманова, 2009. — С. 185-191.

8. Методики написання наукових робіт на основі онтологічного аналізу текстів: методичний посібник / упоряд. Андрущенко Т. І., Величко В. Ю., Гальченко С. А., Глоба Л. С., Гуляев К. Д., Клімова Е. Я., Комова О. Б.,

Лісовий О. В., Попова М. А., Приходнюк В.В., Сидоренко В. І., Стрижак О.Є., Стус Д. М.,- К. : ТОВ «СІТПРІНТ», 2013. - 124 с.

#### Література до розділу 9

1. Гладун В.П. Процессы формирования новых знаний / В.П. Гладун – София: СД "Педагог 6", 1994. – 192 с.

2. Гладун В.П. Планирование решений / Гладун В.П. – Киев: Наукова думка, 1987. – 168с.

3. Мендельсон Э. Введение в математическую логику. — М. Наука, 1971. – 320 с.

4. Ішук О. О. Просторовий аналіз і моделювання в ГІС: Навчальний посібник / О. О.Ішук, М. М. Коржнев, О. Е. Кошляков; за ред. акад. Д.М.Гродзинського. – К.: Видавничо-поліграфічний центр "Київський університет", 2003.– 200 с.

5. Про один підхід до аналізу та розуміння природномовних об’єктів / [Палагін О.В., Світла С.Ю., Петренко М.Г., Величко В.Ю.]. – Комп’ютерні засоби, мережі та системи, 2008, №7. – С. 128–137.

6. Величко, В. Автоматизированное создание тезауруса терминов предметной области для локальных поисковых систем / В. Величко, П. Волошин, С. Свитла // “Knowledge – Dialogue – Solution” International Book Series “INFORMATION SCIENCE & COMPUTING”, Number 15. – FOI ITHEA Sofia, Bulgaria. - 2009. – pp.24-31.

7. Заде, Л. Понятие лингвистической переменной и его применение к принятию приближенных решений / Л. Заде // М.: Мир, 1976. 166с.

8. Ganter, B. Formal Concept Analysis: Mathematical Foundations / B. Ganter, R.Wille // Berlin-Heidelberg, Germany: Springer-Verlag, 1999.

9. Quan Thanh Tho. Automatic Fuzzy Ontology Generation for Semantic Web. / Quan Thanh Tho, Siu Cheung Hui, Tru Hoang Cao // IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 18, NO. 6, JUNE 2006.

10. Мендельсон Э. Введение в математическую логику. — М. Наука, 1971. — 320 с.
11. Князева Е. Н. ТРАНСДИСЦИПЛИНАРНЫЕ СТРАТЕГИИ ИССЛЕДОВАНИЙ // Вестник ТГПУ. 2011. №10. .
12. Андрущенко Т.І. Співець народного єднання // В літопис шани і любові: Збірник наукових праць, присвячений 175-річчю від дня народження Т.Г.Шевченка – К.: УМК ВО України, 1989. – С.205-217.
13. Андрущенко Т.І. Образотворча шевченкіана Музею Тараса Шевченка// Образотворче мистецтво. – 1999. – № 3/4. – С. 12-14.
14. Андрущенко Т.І. Національне в мистецькій спадщині Тараса Шевченка і сучасна образотворчість // Образотворче мистецтво – 2000. - № 1-2. С. 22-25.
15. Андрущенко Т.І. Тарас Шевченко // Великие художники: их жизнь, вдохновение творчество. – Ч.24: Тарас Шевченко - К.: Иглмосс Юкрейн, 2003. – 31 с
16. Андрущенко Т.І. Малярство Тараса Шевченка. Примітки // Тарас Шевченко. Вибрана поезія. Живопис. Графіка. / Укл. І.М.Дзюба, С.А.Гальченко. – К.: Мистецтво, 2007. – С.482-603
17. Андрущенко Т.І. Малярство Тараса Шевченка. Коментарі до малярських творів // Т.Г.Шевченко Кобзар: Повна ілюстрована збірка / Укл. І.М.Дзюба, С.А.Гальченко Харків:Клуб семейного досуга, 2009 – С. 695-714.
18. Андрущенко Т.І. Проблема естетичного в культурі (матеріали до спецкурсу) К.: Знання України, 2010. – 550с.
19. Скуратовский Вадим. Тарас Шевченко: Прочтения и имитация прочтения // Столичніє новости. - 2003. - 18 июня - С.17.
20. Матвієнко Світлана. Любити Шевченка: демонтаж риторизму // Дзеркало тижня. - 2003. - 7 червня. - С.19.
21. Гніздовський Яків. Масова механічна репродукція. - 1967.
22. Овсійчук Володимир. Стан мистецтвознавства в Україні. Народознавчі зошити. - 1999. - Зошит 5. - С.597.
23. Костенко Ліна. Гуманітарна аура нації, або Дефект головного дзеркала. - К., 1999.
24. Ортега-і-Гассет Хосе. Дегуманізація культури. - К., 1994. - С.243.
25. Вдовин Геннадий. Кризис музея как мировоззрение // Мир музея. - 2002. - № 6. - С.38-39.
26. Відображення культурної спадщини України з використанням і електронних карт в середовищі геоінформаційної системи «МАН України»: навчально-методичний посібник / упоряд. Андрущенко Т. І., Бревус С. М., Гальченко С. А., Кудряк В. М., Ляшук К. В., Попова М. А., Рафальський О.О., Сидоренко В. І., Слюсаренко О. О., Стрижак О. Є., Стус Д. В. - К.: ТОВ «СІТІПРІНТ», 2013- 120 с.
27. Інформаційні системи підтримки навчальної та дослідницької групової діяльності обдарованих учнів за схемою <учні - викладачі - науковці>:

28. Методичні рекомендації / Т. І. Андрущенко, С. А. Гальченко, Л. С. Глоба, О. Є. Стрижак, В. Ю. Величко, В. М. Кудляк, К. В. Ляшук, М. А. Попова, В. І. Сидоренко, О. О. Слюсаренко, Д. В. Стус, І. С. Чернецький; За ред. О. Є. Стрижака. - К.: ТОВ «Інформаційні системи». - 2012. - 200 с.

29. Використання розподільних інформаційних систем в навчальному процесі позашкільного навчального закладу Мала Академія Наук: Метод. рекомендації / Т. І. Андрущенко, С. А. Гальченко, Л. С. Глоба, О. Є. Стрижак, Н. І. Алішов, В. Ю. Величко, А. М. Галелюка, О. В. Гупал, В. М. Кудляк, В. А. Лісовий, К. В. Ляшук, В. А. Марченко, В. М. Опанасенко, О. В. Палагін, М. А. Попова, В. В. Прокопчук, О. О. Рафальський, В. М. Романов, Є. В. Сарахан, О. М. Трофимчук; За ред. д. т. н., академіка НАНУ О. В. Палагіна і к. т. н. О. Є. Стрижака. - К.: ТОВ «Інформаційні системи». - 2012. - 465 с.

НАУКОВЕ ВИДАННЯ

**Комп'ютерні онтології та їх використання у  
навчальному процесі. Теорія і практика**

**МОНОГРАФІЯ**

Підписано до друку 27.03.2013. Формат 60x84 1/16  
Друк цифровий. Папір офс. 80 г/м<sup>2</sup>. Ум. друк. арк. 18,02  
Наклад 300 прим. Зам. № 110

Видавництво Інститут обдарованої дитини  
вул. Артема, 52-Д, м. Київ, 04053  
тел./факс (044) 481-38-38  
Свідоцтво про внесення до Державного реєстру  
серія ДК №3366 від 13.01.2009 р.