

НАЦІОНАЛЬНА АКАДЕМІЯ ПЕДАГОГІЧНИХ НАУК УКРАЇНИ
ІНСТИТУТ ОБДАРОВАНОЇ ДИТИНИ

**Програмно-інформаційні
засоби формування систем знань
навчального призначення**

Посібник

Київ
2014

УДК 004.89:658.5:37.091.3

ББК 74.202.5

П78

*Рекомендовано до друку Вченою радою
Інституту обдарованої дитини НАПН України
(протокол № 6 від 25.06.2014 р.)*

Рецензенти:

О. В. Палагін – академік Національної академії наук України, професор, доктор технічних наук

Л. В. Струганець – доктор філологічних наук, професор, завідувач кафедри методики викладання української мови і культури мовлення Тернопільського національного педагогічного університету імені В. Гнатюка

Авторський колектив:

О. Є. Стрижак, В. Ю. Величко, Л. С. Глоба, К. Я. Климова, О. Б. Комова, М. Г. Петренко, М. А. Попова, В. В. Приходнюк

П78 **Програмно-інформаційні засоби формування систем знань навчального призначення : посібник / за ред. О. Є. Стрижака. – К. : Інститут обдарованої дитини, 2014. – 144 с.**

ISBN 978-966-2633-43-6

У посібнику викладено технологічні підходи щодо термінологічного розбору речень, виділення контексту для певних термінів, формування таксономії текстового документу та послідовний семантичний аналіз змісту. Виклад матеріалу оснований на використанні програми TextTegmIn.

Рекомендовано для викладачів, учителів, студентів педагогічних університетів та учнів старших класів.

УДК 004.89:658.5:37.091.3

ББК 74.202.5

ISBN 978-966-2633-43-6

© Інститут обдарованої дитини
НАПН України, 2014

ЗМІСТ

ВСТУП	5
РОЗДІЛ 1. <i>Концептуальні аспекти побудови систем знань</i>	7
1.1. Дослідження предметно-орієнтованих знань. Концептуальний аналіз предметних знань	7
1.2. Концептуальна парадигма роботи зі знаннями	14
1.3. Проблеми, методи та засоби обробки природномовних об'єктів	27
1.4. Формалізовані підходи добування та представлення знань	31
1.5. Формальна постановка задачі добування знань з ПМТ	33
1.6. Комп'ютерні онтології ПдО	34
1.7. Лінгвістичні онтології	41
1.8. Системна інтеграція лінгвістичних і предметних онтологій	58
1.9. Методологія розробки онтології ПдО	64
РОЗДІЛ 2. <i>Інструкція для користувача програмою TextTermin:</i> <i>автоматичне виділення термінів на основі лінгвістичного аналізу</i>	79
2.1. Вимоги до робочого місця (комп'ютера)	80
2.2. Вимоги до тексту	80
2.3. Налаштування	81
2.4. Установлення програми	83
2.5. Опис основного вікна «Terms' Extraction»	86
2.6. Опис вікна «Дерево термінів»	86
2.7. Опис вікна «Речення з тексту, у яких присутні вибрані терміни»	87
2.8. Список термінів	88
2.9. Визначення типу термінів	89

2.10. Пошук	91
2.11. Збереження даних	93
2.12. Перенесення термінів у таблицю	94
2.13. Збереження таблиці	97
2.14. Конвертація даних за допомогою програми Confor	98

РОЗДІЛ 3. <i>Зразки виконання дослідницької роботи з використанням програми TextTermin</i>	100
--------------------------------------------------------------------------------------------------	-----

3.1. Завдання I. За допомогою програми TextTermin побудувати дерево термінів та провести статистичний (кількісний) контент-аналіз: які слова та словосполучення найчастіше використовує поет у своєму листуванні	100
3.2. Завдання II. Skorиставшись даними програми TextTermin, побудувати онтологічний граф з вершиною «Лист» та прикріпити відповідні дані та гіперпосилання	101

РОЗДІЛ 4. <i>Технологія створення онтологічних моделей та профільних баз знань</i>	112
------------------------------------------------------------------------------------------	-----

4.1. Лінгвістична термінологія	113
4.2. Математична термінологія	114
4.3. Хімічна термінологія	115
4.4. Психологічна термінологія	116

РОЗДІЛ 5. <i>Технологія контент-аналізу лексики художнього твору</i>	117
----------------------------------------------------------------------------	-----

5.1. Зміст технології: дослідження концепту «Україна» у текстах збірки Т. Г. Шевченка «Кобзар» за допомогою програми TextTermin	119
5.2. Основний етап: вирішення завдань дослідження	120

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	125
----------------------------------	-----

ДОДАТОК А. Зразки завдань для дослідження життя та творчості Т. Г. Шевченка з використанням онтологій, побудованих за допомогою програми TextTermin	129
-----------------------------------------------------------------------------------------------------------------------------------------------------------	-----

ВСТУП

У посібнику викладено основи створення методичних та інформаційно-програмних засобів побудови онтологічних моделей опису функціональних об'єктів та процесів під час формування загальної мережі знань для вирішення актуальних проблем підвищення ефективності підтримки процесу дослідницької діяльності талановитої молоді. Вони охоплюють науково-методологічні засади та сучасні інформаційні технології, що забезпечують розробку формалізованої системи знань в конкретних предметних галузях, зокрема з використанням віртуальних дослідно-проектувальних лабораторій (ВДПЛ). Ці лабораторії надаватимуть учням можливість дистанційно та локально виконувати дослідження за допомогою аналітичного устаткування, а також аналіз та обробку отриманих результатів.

У посібнику описано та наведено:

- методологія системно-онтологічного аналізу предметної області, інформаційних та лінгвістичних моделей обробки текстових документів із заданої предметної області, формалізації та представлення набутих знань;
- архітектура і структура інструментального комплексу підтримки процесів онтологічного опису предметних галузей.

Наведено приклади та фрагменти програм опрацювання онтологічних даних (ПООД), що є складовими інструментального комплексу і підтримують процес автоматизованої побудови онтологічних структур. Серед різних варіантів практичного використання сховищ текстової інформації превалює потреба виділення інформації, що характеризується тематичною єдністю. Тематичний пошук вимагає копійної роботи з текстами, що зберігаються в бібліотеках, архівах,

Інтернеті, текстових баз даних (БД) тощо. Складність цієї роботи, зокрема полягає в тому, що частіше доводиться відбирати не цілі тексти, а лише фрагменти, які відповідають обраній темі. Зміст багатьох текстів є переплетінням низки тем. Виникає проблема пошуку у текстових документах фрагментів, які релевантні заданій темі.

Існуючі методи інформаційного пошуку, зокрема в Інтернет-мережі, серед результатів видають багато непотрібної «сміттевої» інформації, фільтрація якої займає багато часу. Вирішення проблеми полягає у використанні семантичних критеріїв, що забезпечують відбір найбільш істотних характеристик понять, щодо яких збирається інформація.

Програма обробки лінгвістичних даних «Конспект» виконує такі функції:

- лінгвістичний аналіз тексту до рівня поверхневого синтактико-семантичного аналізу;
- виділення термінів предметної області з релевантних текстів;
- виділення і стисле конспектування фрагментів природномовних текстів, що належать до обраної 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 тих, що найбільшою мірою релевантні заданій темі.

РОЗДІЛ 1

Концептуальні аспекти побудови систем знань

1.1. 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 поєднанні дослідницьких методів онтології, гносеології та логіки. Перш ніж перейти до детального розгляду науково-практичних задач аналізу знань, варто звернутись до його філософських аспектів, а точніше до тієї сукупності спільних наукових дисциплін, підходи, методи та принципи яких безпосередньо впливають на формування і становлення вищезазначеного аналізу. Такими дисциплінами є Онтологія¹, гносеологія (або теорія пізнання), загальна логіка і теорія відображення [1]².

Зміст зазначених дисциплін, тією чи іншою мірою, характеризує тріаду {суб'єкт (*S*) – об'єкт (*O*) – відношення (*R*)}. Під суб'єктом розуміють те, що діє і пізнає (в певному розумінні синонім терміну спостерігач). У більш широкому сенсі під суб'єктом можна розуміти систему будь-якого рівня, яка впливає та відображає. Під об'єктом

¹ Цей термін написаний з великої літери, тому що в цьому контексті він застосований до онтології як філософської дисципліни, щоб відрізнити від комп'ютерної онтології.

² Сказане можна рівною мірою віднести і до двох інших аспектів дослідження – методологічного і логічного.

розуміють систему, яка відображається та піддається впливу, щось пізнаване і змінюване суб'єктом. Із точки зору системи $\Lambda = \langle O, S, R \rangle$ відомі наступні (нестрогі) визначення вищезгаданих наукових дисциплін [1].

Онтологія – це вчення про загальні властивості та відношення об'єктивного світу, загальна теорія, що розглядає об'єкт поза його відношення 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 результатом специфікації загальної теорії відображення.

Роль вищезазначених наукових дисциплін розглянемо на прикладі, де в ролі об'єкта виступає наукова теорія.

Логіка розглядає структуру цієї теорії незалежно від того, що вона відображає, які об'єктивні основи цієї структури. Із 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 операцій, але абстрагується від соціальних і психологічних умов, їх генезису та функціонування.

Загальна взаємодія перелічених наукових дисциплін і системи Λ представлена на рисунку 1.1.

Таким чином, у системі Λ різні фахівці, які займають різні позиції, вивчають різні предмети.

Звісно, описана модель взаємодії наукових дисциплін і системи Λ є наближеною. У ній не розглянуто декілька уточнюючих моментів, зокрема:

- 1) межі загальної логіки;
- 2) процес впливу суб'єкта на об'єкт;
- 3) процес, коли суб'єкт і об'єкт можуть мінятися місцями.

Важливо наголосити, що застосування математичних методів у методології породжує нову дисципліну – **технологію** пізнання, де

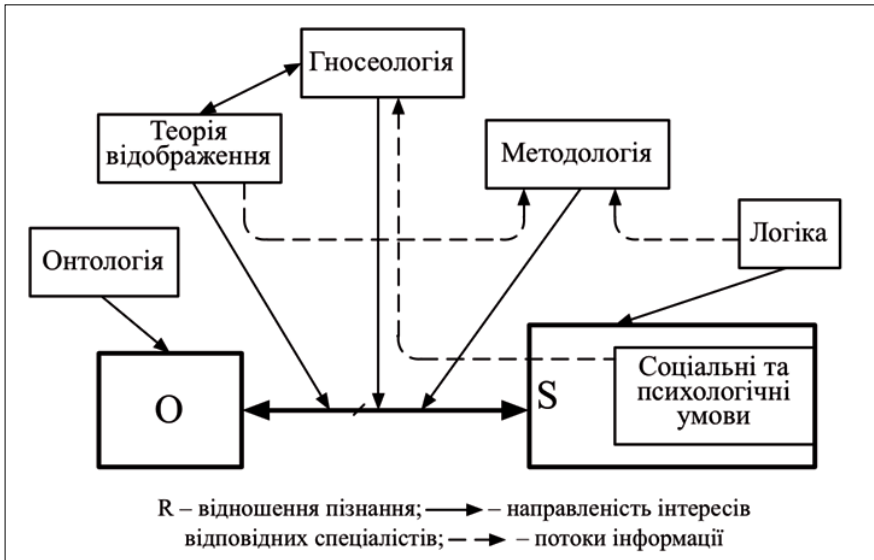


Рис. 1.1. Загальна структура і взаємозв'язки наукових теорій, що вивчають систему «об'єкт ↔ суб'єкт»

методологія переходить від загальних принципів аналізу до процесу синтезу (проекування).

Якщо система Λ розглядається на абстрактному рівні, то її аналізом повинна займатися онтологія, як філософська дисципліна. Якщо об'єктом обрано певну ПдО (що представляє найбільший практичний інтерес), то його повинна розглядати «комп'ютерна онтологія предметної області» (О ПдО). Аналогічно 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 сфер штучного інтелекту, які колись розвивалися паралельно і незалежно: knowledge-engineering і комп'ютерної лінгвістики (когнітивної семантики). Варто

наголосити, що це прагнення відображає природну схему взаємодії людини з навколишнім світом: *вхідна інформація* → «свідомість – розуміння – знання» (щоб виділити «конструктивну» тріаду в останній не вказано проміжні процедури когнітивного циклу: відчуття, рефлексія та ін.). Свідомість відіграє роль персоніфікованого інструменту, що виробляє сукупність предметно, ситуативно або причинно-зв'язаних сутностей, які утворюють «свідоому» 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». Свідома картина світу представлена «мовною» картиною світу, яка є не лише системою знань, а й певним буфером, що 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 (простір, час, рух, якість, кількість, відношення) або більш місткі категорії пізнання і сфери буття (матерія, свідомість, практика), містить три взаємопов'язаних рівня:

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

Перший рівень орієнтовано на формалізацію побудови і правил виведення на семантичній (концептуальній) мові понять. Другий рівень може бути аналогізовано з сучасними тлумачними словниками (в електронному вигляді), проте володіє 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 (третій) забезпечує інформацію про морфологічні (рід, число, відмінок, вид, час тощо), синтаксичні (атрибутивні, кількісні, обставинні, службові тощо) і семантичні (об'єктні, ситуаційні,

ознакові, актантно-рольові, лексико-функціональні) характеристики лексем та їх семантико-синтаксичних конструкцій. Технологічною основою МКС можуть бути моделі реляційних БД, що забезпечують функції взаємодії у формі «ПІМ-SQL» запитів як природномовного інтерфейсу [11].

Таким чином, МКС можна представити через віртуальні індексовані інформаційні поля таблиць лексико-семантичних відношень («ціле – частина», «рід – вид», «множина – елемент», «причина – наслідок», «об’єкт – функція» тощо) та їх компонент (стовпців і рядків) з наборами семантичних характеристик лексем і складних 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 (ПдОі) є i -м розширенням загальної картини світу таким чином, що повну картину світу (або піраміду знань) можна представити у вигляді, поданому на рисунку 1.2. Тут і проходить через позиції філософської, біологічної, хімічної, фізичної картин світу (КС) або (що те саме) через позиції наукових знань у кожній із названих предметних областей. Кожна з них (КМі = ПдОі) формується за тими ж законами системного підходу (цілісності, ієрархії, класифікації та структурування), що і МКС. Застосовуючи методи системних переходів [11], процедуру, що реалізує перетворення за формулою Брукса ($K(S) + \Delta I = K(S + \Delta S)$, де $K(S)$ – початкові знання, ΔI – нова порція інформації, $K(S + \Delta S)$ – вихідні знання), можна сформулювати в термінах узагальнення сенсу. Йдеться про ($i \rightarrow j$)–переходи, з чого випливає, що нові, невідомі раніше знання 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 визначено. По-друге, отримання нових знань про навколишній реальний світ часто пов’язано зі зміною уявлень (запереченням) старих знань. Цей процес можна схематично проілюструвати (див. рис. 1.3). Тут $j = 1, 2, \dots$ – рівні

узагальнення знань: від елементарних (аксіоматичних) знань до повних. Стрілки вказують напрям розвитку знань: \rightarrow – розширення елементарних знань, \uparrow – узагальнення існуючих знань. Показники dI_j , dI_i вказують на збільшення знань за відповідними осями координат. Зона, обмежена перетином двох кривих, – це область знань, інваріантних щодо нових і старих теорій. Зони Sc і Sn – це області взаємовиключних розділів теорій. Цю інтерпретацію процесу розвитку знань можна використати як в межах однієї предметної області, так і для опису повної картини світу. Другий варіант передбачає представлення нетривіальними міждисциплінарних функцій МКС в плані креативної взаємодії ПдОі і процесів конвергенції їх теорій і технологій (рис. 1.2).

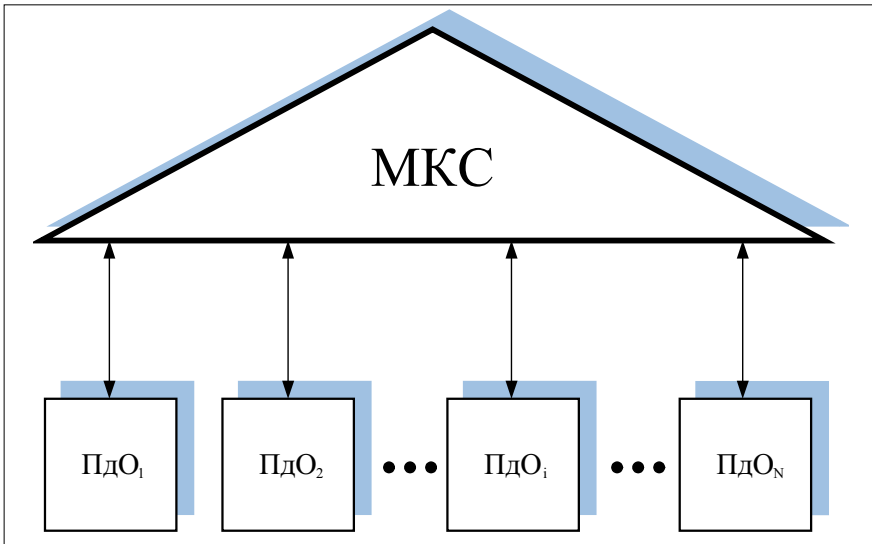


Рис. 1.2. «Піраміда» знань – повна картина світу

Йдеться про побудову формальної мови категоріального рівня, що забезпечує ефективне узагальнення, перенесення і впровадження прийомів, методів і теорій з однієї ПдО в іншу, що має підвищити ефективність процесу міждисциплінарної взаємодії. Як впливає з вищерозглянутого аналізу взаємодії спільних наукових дисциплін,

сама теорія пізнання визначає основні принципи та напрями дослідницького процесу в системі Λ . Разом із тим важливо відзначити радикальну зміну і самої системи наукового пізнання.

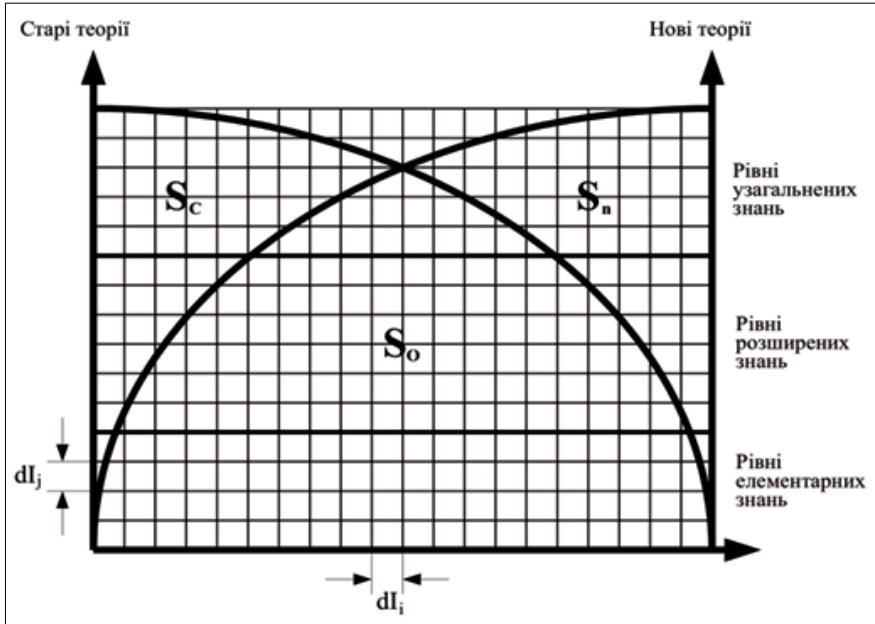


Рис. 1.3. Розвиток знань

Між практичною та пізнавальною діяльністю розвиваються чіткі межі. У системі наукового знання відбуваються інтенсивні процеси їх диференціації та інтеграції, розвиваються комплексні міждисциплінарні та трансдисциплінарні дослідження, нові способи і методи пізнання, методологічні установки. Також з'являються нові елементи картини світу, виділяються нові, більш складні типи об'єктів пізнання, що характеризуються історизмом, універсальністю, складністю організації, які раніше не піддавалися теоретичному (математичному) моделюванню. Варто зауважити, що проблеми і закономірності зазначеної взаємодії досліджуються в межах нового міждисциплінарного напрямку – синергетики, яке можна розглядати як сучасний етап розвитку кібернетики і системних досліджень.

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

1.2. Концептуальна парадигма роботи зі знаннями

Свого часу В. М. Глушков виділив *пошук нових принципів побудови ЕОМ* як один з основних напрямів розвитку кібернетики [15]. У межах цього напрямку, в результаті «еволюції людського знання, що нагадує сходження по спіралі або сходами з великими сходінками» було сформовано і розвинуто *концептуальну парадигму роботи зі знаннями*. Це складне поняття, що спирається на два інших, які визначаються термінами «концептуальна парадигма» і «робота із знаннями».

Парадигма часто виглядає як проста, елегантна і правдоподібна концептуалізація даних, на підставі якої можливо пояснення більшої частини спостережуваних фактів. Вона висловлює певне 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 розпочинається етап поступального розвитку науки, більшість відомих фактів знаходять пояснення. У процесі досліджень і спостережень неминуче виявляються такі феномени (їх стає все більше і більше), що не знаходять свого пояснення. Це призводить до необхідності зміни існуючої парадигми. Таким чином, надриває нова революція і підйом на новий щабель розвитку знань.

Варто сказати, що парадигми несуть в собі не лише пізнавальний, але і стверджуючий характер, а 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 зі знаннями» за належністю відноситься до галузі інженерії знань, а за належністю до онтологічної категорії – до поняття-процесу, що загальноприйнято визначати екстенціонально (через перерахування назв підпроцесів). До таких підпроцесів роботи зі знаннями зазвичай відносять *вилучення* та *добування, представлення знань*, а також маніпулювання знаннями. Сфера аспектів роботи зі знаннями, окрім зазначених підпроцесів, охоплює також методи і засоби. В онтологічному інжинірингу до цієї області долучають онтологію процесів, задач і об'єктів.

1. Вилучення знань з різних джерел, у т. ч. з природномовних об'єктів (ПМО), містить два основні розділи: аналіз вихідної інформації, формалізацію якісних знань й інтеграцію знань. Перший розділ пов'язано зі створенням методів, що дозволяють переходити від знань, виражених в природномовній формі, до їх аналогів, придатних для введення в пам'ять знання-орієнтованої інформаційної системи (ЗОІС). Другий розділ пов'язано з інтеграцією знань, набутих з різних джерел, у певну взаємопов'язану і несуперечливу систему знань про предметну область. Без вирішення цієї проблеми навряд чи можливо сформуванню 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 отримати такі знання, потрібні спеціальні прийоми і методи. Їх використовують в інструментальних системах добування знань, створення яких є одним із завдань інженерії знань. Отримані від експертів знання потрібно оцінити з точки зору їх відповідності до раніше накопичених знань і формалізувати для введення в пам'ять системи. Окрім того, знання, отримані від різних експертів, необхідно узгодити між собою, оскільки часто буває, що ці знання виявляються несумісними і навіть суперечливими. Розглянутий підпроцес охоплює 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 охоплює побудову *моделі знань, системи подання знань і бази знань* (БЗ). В інженерії знань системи подання знань містять сукупність процедур, необхідних для запису знань, вилучення їх з пам'яті та підтримки збереження знань в робочому стані. Системи подання знань оформлюються як БЗ. Вони є природним розвитком БД. Представлення знань – це угода про те, як і в якій формальній теорії описувати реальний світ. У природничих і технічних науках використовують традиційний спосіб представлення знань. Природною мовою вводяться основні поняття і відношення між ними, але при цьому використовуються раніше визначені поняття та відношення, зміст яких уже відомий. Потім відбувається етап встановлення відповідності між характеристиками (найчастіше кількісними) понять і підходящою математичною моделлю. Основна мета подання знань – побудова математичної 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);
- виокремлювати 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 належать такі розділи як *поповнення знань, класифікація знань, узагальнення знань, виводи на основі знань* (резольційні методи, квазіаксіоматичні системи і системи правдоподібного виводу), *міркування за допомогою знань, пояснення на знаннях, вирішення прикладних задач ПДО*.

Нові знання, що надходять до БЗ, повинні сформувати розширення знань за допомогою відомостей, які вже були раніше записані до неї. Особливе місце серед цих процедур займають псевдофізичні логіки

(часу, простору, дій тощо), 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, типу «частина – ціле», ситуативні (коли в одну множину об'єднуються знання, релевантні певній типовій ситуації). Зазначені процедури формуються 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 призводять до абстрактних знань, для яких немає прямого прообразу в зовнішньому світі.

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

Можливість появи в пам'яті ЗОІС нових фактів і відомостей призводить до того, що починає порушуватися принцип монотонності. Цей принцип лежить в основі функціонування всіх систем, що вивчаються традиційною математичною логікою. Немонотонність виведення у відкритих системах викликає чимало труднощів. В останні роки прихильники логічних методів у штучному інтелекті роблять спроби побудувати нові логічні системи, у межах 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 системи, що виникають за таких умов, варто називати *квазіаксіоматичними*. У таких системах цілком можлива зміна вихідних аксіом в процесі виведення. Ще однією особливістю виводів на основі знань є неповнота відомостей про предметні області та процеси, що в ній протікають, а також неточність вхідної інформації. Це означає, що виводи в ЗОІС носять не абсолютно достовірний характер, як у традиційних логічних системах, а наближений, правдоподібний характер. Такі виводи вимагають

розвинуеного апарату обчислення оцінок правдивості та методів оперування ними.

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

На основі вищепроведеного аналізу на рисунку 1.4 синтезовано узагальнену схему введення і обробки вхідних даних, семантичної переробки інформації та роботи зі знаннями. У ній також відображено зв'язки між відповідними процесами, послідовності яких становлять інформаційні технології та обробки даних, текст-процесингу, роботи зі знаннями і логіко-інформаційного підходу. Вони становлять основу методології проектування знання-орієнтованих інформаційних систем з обробкою ПМО і формалізованих знань.

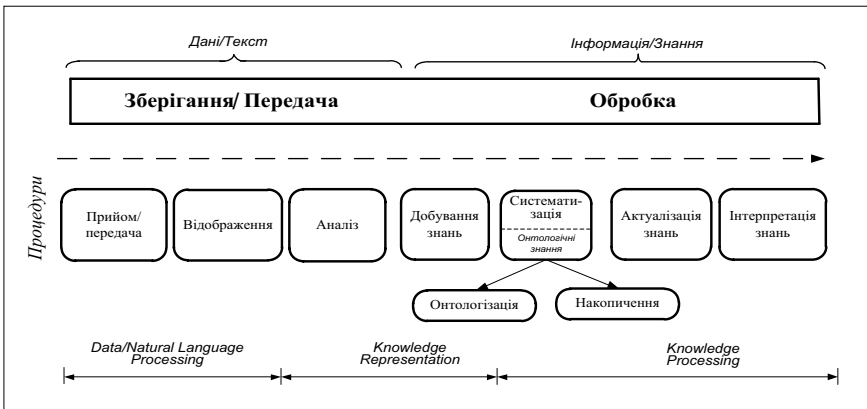


Рис. 1.4. Схема роботи зі знаннями

Концептуальна парадигма роботи зі знаннями утвердилась завдяки розвитку концепції інтелектуальних інформаційних систем (ІС). Згідно з нею, головною інформаційною одиницею комп'ютерної об-

робки стало «знання». У межах різних областей предметних знань і незалежно один від одного розвиваються інформаційні технології:

- автоматичної обробки природної мови – у комп'ютерній лінгвістиці,
- представлення знань – у 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»), варто зауважити, що сукупність наукових дисциплін, які мають безпосереднє відношення до процесів мислення, розуміння, усвідомлення та набуття нових знань, не функціонує завдяки виробленню своїх власних понять, певним способом фіксує властивості і закономірності її об'єктів. Із сукупності основоположних понять (*смысл, знання, (знакова)* система, текст, об'єкт, ставлення, предмет, мова, структура, зв'язок) розглянемо лише перші два, які виражають найбільшу проблемність досліджень в області побудови систем обробки знань, у т. ч. і ЗОІС, що оперують ПМО. Спершу розглянемо визначення категорії «поняття», що є фундаментом для складніших форм мислення – судження і висновку.

Зазвичай поняття визначають як одну з основних форм мислення, що підкреслює його важливу роль в пізнанні. Саме мислення може розглядатися як процес оперування поняттями, а завдяки поняттям воно набуває характеру узагальненого відображення дійсності. Виділення предметів – одна з основних функцій поняття у процесі пізнання. Виділені в поняттях класи предметів – це опорні пункти в пізнанні, навколо яких концентруються всі знання [20].

Поняття – це цілісна сукупність суджень, в яких стверджуються визначальні ознаки досліджуваної сутності, ядром якої є судження (або твердження) про найбільш загальні і zarazом істотні ознаки цієї сутності.

Кожне поняття характеризується об'ємом і змістом. Об'єм і зміст поняття – дві взаємопов'язані сторони поняття. **Об'єм** – це

клас узагальнених в понятті предметів, а **зміст** – це сукупність істотних ознак, за якими вироблено узагальнення і виділення предметів у даному понятті. Об'єм поняття є визначальним під час формування ієрархічної структури відповідного онтографа (ОГ), а зміст – під час аксіоматизації його (ОГ) вершин.

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

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

Смысл – це певний безперервний, невербальний конструкт, а осмислення – інтерпретація в індивідуальній концептуальній системі [11].

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

Варто зазначити, що «... природа смыслу може бути розкрита лише через одночасний аналіз *семантичної тріади*: «смысл–текст–мова» [23]. Текстове розкриття змісту відбувається через знакові системи, які ми сприймаємо як мову. Таким чином, кожен елемент вищезазначеної тріади розкривається через два інших. Включаючи в тріаду мову, ми вносимо представлення про те, що сама тріада стає можливою, лише коли є СПОСТЕРІГАЧ – носій свідомості, що сприймає тексти і оцінює смисли. Тріада стає синонімом свідомості, або якщо сказати точніше – мовної свідомості.

Визначення 1.1

1. **Розуміння** – це фіксація смыслу природномовного об'єкта.
2. **Фіксація** смыслу природномовного об'єкта – це проектування ПМО на мовну картину світу.
3. **Смысл** – 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 щодо їх комп'ютерної обробки.

З першої групи можна виділити наступні, найбільш характерні визначення.

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

Знання є сукупністю понять і уявлень про об'єктивну дійсність, їх внутрішньо взаємозалежних систем (суджень, положень, концепцій, теорій і т. д.), що виробляються суспільством у процесі пізнання і перетворення світу. Знання зародилось в епоху первісного духовно-фізичного синкретизму трудової діяльності людини. Воно і надалі продукується у спеціалізованій (наукової) діяльності суспільства, насамперед для досягнення практичних цілей. Знання все ширше охоплює об'єкт, все глибше пронизує його, тобто розвивається як в екстенсивному, так і в інтенсивному плані [24].

Ці приклади визначень досить коректні. Вони 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.

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

Знання – це складно організована інформаційна структура, що збережена в пам'яті інтелектуальних інформаційних систем і містить відомості про об'єкти та відношення предметної області, процеси взаємодії об'єктів у часі та просторі, а також правила провадження логічного виводу [3].

Визначення другої групи надають певне узагальнене уявлення про систему, засновану на знаннях, але все ж цього недостатньо, щоб побудувати зазначену систему.

Для цього необхідно 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.

Визначення 1.2

Під терміном «знання» варто 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 низкою властивостей, що відрізняють їх від традиційних моделей даних. До таких властивостей можна віднести наступні.

Внутрішня інтерпретація. У процесі зберігання знань у пам'яті ПС, разом з традиційними елементами даних зберігаються інформаційні структури, що дозволяють інтерпретувати вміст відповідних чарунок пам'яті.

Структурованість. Знання складаються з окремих інформаційних одиниць, між якими можна встановити класифікуючі типи відношень: «рід – вид», «клас – елемент», «тип – підтип», «частина – ціле» тощо.

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

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

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

Зазначимо, що ця властивість є фундаментальною для поняття «знання». На його основі введено поняття онтолого-керованої архі-

текстури КС, оскільки поняття «*онтологічна база знань*» є видовим відносно поняття «*база знань*».

Окрім того, властивість актуальності знань може породжувати процес актуалізації, який має визначальне значення у ланцюгу переходів «*сховище_даних* → *затребувана_інформація* → *знання*». Однак, з практичної точки зору, більш актуальною є проблема не первинності компонент ланцюга, а їх приналежності.

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 формі символів, якими оперує людина чи комп'ютер.

На будь-якому етапі перетворення вимога актуалізації відповідної компоненти є обов'язковою (але не достатньою) умовою переходу до вищого ступеню абстрактності. І якщо для першого переходу («*сховище_даних* → *затребувана_інформація*») механізми актуалізації досить добре вивчено та опрацьовано, то для другого переходу («*затребувана_інформація* → *знання*»), зазначені механізми слабо вивчено і реалізовано в розрізних експериментальних розробках.

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

Виконаний аналіз дозволяє синтезувати певну узагальнену схему «*еволюціонування*» знань, представлену на рисунку 1.5. Зауважимо, що в мові даного звіту словоформи термінів «*знання*», «*свідомість*» (у російськомовному варіанті «*сознание*») і «*пізнання*» мають спільний корінь, а самі терміни також утворюють тріаду, компоненти якої фіксують об'єкт, суб'єкт і процес пізнання.

Входом є деяка знакова система, яку можна інтерпретувати як сукупність даних. Потім відбувається «*осмислення*» вхідних даних, для

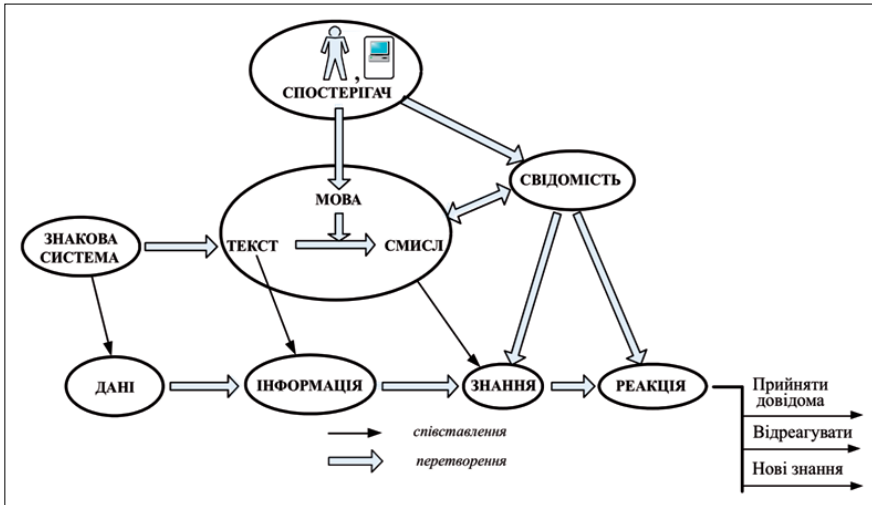


Рис. 1.5. Узагальнена схема «еволюціонування» знань

чого обов'язкова присутність спостерігача (людина, комп'ютер), який є носієм вищих форм мислення – мовної свідомості. І якщо у нього є «внутрішній мовний інтерпретатор, налаштований на дану знакову систему», то дані перетворюються в інформацію. Далі до роботи долучається «смісловий інтерпретатор» (якщо він є у спостерігача), який перетворює інформацію на знання або відбувається усвідомлення нової інформації. На заключному етапі обробки вхідного повідомлення відбувається природна реакція людини: взяти до відома, відреагувати або поповнити БЗ новою порцією знань [10].

Джерела знань та схема комп'ютерної обробки знань

Виділяють два основних джерела знань – це експерти, фахівці у своїй предметній області та лінгвістичний корпус текстів (або множина ПМО) [3–5, 11, 16, 22, 27–30]. Якщо для першого джерела досить добре вивчено і опрацьовано методи здобуття знань, а також відомими є відповідні промислові експертні системи, то для другого – розроблено лише окремі методи, не пов'язані єдиною інтегрованою технологією, а відповідні інформаційні системи носять експериментальний характер

і не досконалі. Актуальним є удосконалення уже відомих і розроблення нових, більш ефективних формальних підходів і моделей обробки. Цей процес розпочинається з пошуку релевантної інформації, її аналізу, усунення різного роду неоднозначностей, формально-логічного, онтолого-семантичного, інформаційно-кодового представлень, після чого відбувається перехід до формалізованого опису знань, розроблення процедур роботи зі знаннями і відповідних їм алгоритмів, закінчуючи усе отриманням конкретних результатів користувачем. У більш загальному вигляді вищенаведена послідовність етапів обробки знань, що містяться в природномовних об'єктах, представляється ланцюгом технологій Natural Language Processing (NLP) → Knowledge Representation (KR) → Knowledge Processing (KP).

Загальна схема комп'ютерної обробки знань, що містяться в ПМО, зображена на рисунку 1.6. Тут у загальному вигляді відображено інформаційну 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) на прості, ситуаційні та нові (формальні) знання.

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

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

Під *новими знаннями* розуміють розпізнавання збільшення ΔS і поповнення ними БЗ, заданої ПДО у певному формалізованому вигляді.

Схема на рисунку 1.6 являє 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), що визнано перспективним міждисциплінарним напрямом досліджень.

Аксіоматизація наукових знань визначається такими структурними елементами, як:

- 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 аксіоматизуються.

Усі вищезазвані аспекти становлять базу формального наукового дослідження. Разом з ними науковий підхід передбачає ще одну важливу компоненту – інтерпретацію. Зрозуміло, що науковий смисл мають лише ті абстракції, які можуть бути зараховані до чого-небудь. Їх можна зіставити з істиною, особливо в тому разі, коли це практично важливо й можливо. Механізмом такого зіставлення, а також пошуку логічних можливостей встановлення істини виступає інтерпретація як логічний прийом щодо встановлення значення термінологічних виразів теорій, знакової (формальної) системи.

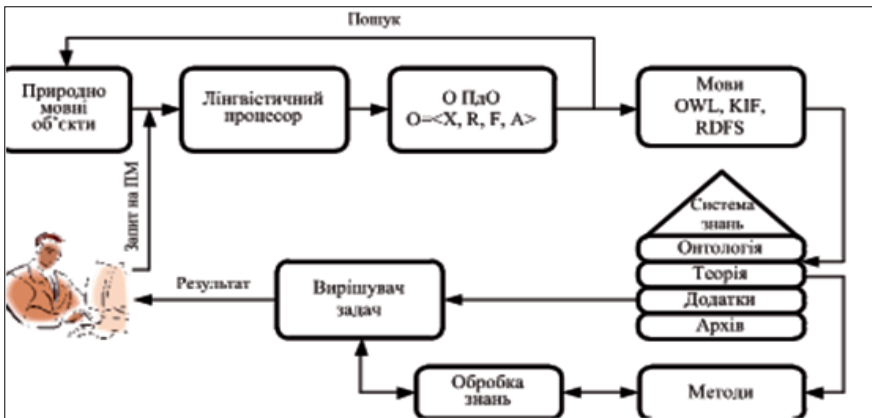


Рис. 1.6. Комп'ютерна обробка знань, що містяться в ПМО

Кожному реченню, логічно правильно побудованому, в інтерпретації надається певне значення (наприклад, речення може бути істинним або хибним). Питання про те, яка з цих можливостей реалізована, не вирішується інтерпретацією. Тому одна і та ж формальна система

передбачає безліч інтерпретацій і лише одну істину. Інтерпретація гарантує несуперечність теорії, але не забезпечує змістовної істинності цієї теорії.

Інтерпретація має свої гносеологічні передумови. Їх суть полягає в тому, що в уявному процесі об'єктивна дійсність не просто пасивно відбивається, але певним чином реконструюється, перебудовується. Під час пізнання відбувається безперервний процес утворення абстракцій та їх виключення, перехід одного рівня абстракцій на інший. Інтерпретація виступає логічним засобом переходу від одного рівня знань до іншого, засобом встановлення логічних зв'язків між цими рівнями.

Відомо досить багато моделей подання знань та відповідних їм методів обробки в КС [3, 5, 17, 25]. 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.

Варто зазначити, що існуючі механізми вилучення нових знань оперують лише знаннями силогістичної типу. Якщо взяти за основу викладену в [23] ймовірнісну модель людського мислення, то перехід до формування нових знань більш високого рівня абстракції методами штучного інтелекту відповідає рівню передсвідомості спостерігача. На цьому рівні знання організовані в складну онтологічну структуру з повним набором зв'язків між концептами. З вищесказаного випливає, що наразі автоматична побудова БЗ для широких предметних областей є *складною науковою проблемою*. При цьому ззовні (відносно КС) фіксуються рівні розташування базових понять ПдО і базові відношення між ними. Таку структуру, створену вручну інженером за знаннями на основі тезауруса ПдО, можна презентувати як *початкову онтологію ПдО*.

1.3. Проблеми, методи та засоби обробки природномовних об'єктів

Аналіз проблеми. Теорія і практика створення та використання систем, заснованих на знаннях (або знання-орієнтованих інформаційних систем) швидко розвивається. Напрямок Computer Science наразі є

найбільш актуальним та таким, що інтенсивно розвивається, оскільки використання його результатів підвищить ефективність створення інструментальних засобів, прикладних систем і застосування комп'ютерів. На жаль, наразі навіть на світовому рівні не отримано відчутних та загально визначених результатів у широких прикладних областях знань, хоча можна зазначити наявні вдалі рішення в цьому напрямі для вузькоспеціалізованих програм. Труднощі вирішення зазначеної проблеми визначаються складністю побудови, організації та використання великих баз формалізованих знань, а також залученням низки наукових теорій (логіки, комп'ютерної та психологічної лінгвістики, нейрокібернетики, теорії семантичних мереж тощо). Очевидно вони повинні сприяти вирішенню проблеми вилучення, формального представлення, обробки і системної інтеграції знань і скласти концептуально-методологічну основу *теорії міждисциплінарних наукових досліджень*.

Як впливає зі вищесказаного, вихідною інформацією для більшості наукових досліджень є знакові системи, представлені у вигляді ПМО. 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*, що досліджує лексичний і структурно-граматичний аспекти ПМ, а також *логіку*, що розглядає мову в одному обмеженому аспекті: лише такою мірою, за якою вона є засобом *фіксації, переробки і передачі знань*. Розглядаючи мовний опис, логіку цікавлять відношення його до елементів об'єктів. За допомогою певних зв'язків з цих елементів утворюються складні знакові системи, що виражають справжні знання про об'єктивний світ. Починаючи з аналізу ПМ, логіка розглядає в ролі пріоритетних – штучні мови науки, що виникають на базі природничих, але відрізняються низкою важливих рис. Для дослідження мов (природних або штучних) логіка використовує інструментарій, реалізований фор-

мальними знаковими системи, що дозволяє виявити закони побудови та функціонування, тобто утворення та перетворення систем знань. Питання про виділення об'єктивних засобів фіксації знань і виведення одних «одиниць» знань з інших за певними, об'єктивно значущими правилами, є одним з пріоритетних у сучасних дослідженнях з логіки [10].

Іншим важливим напрямом 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 певної предметної області. Основною компонентою таких ІС є підсистема *лінгвістичної обробки* ПМТ, завданням якої є розпізнавання і добування знань, що містяться в ПМТ, а також їх формалізація і синтез за формалізованим представленням цих знань на природній мові.

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 значною мірою зумовлена відсутністю єдиної теорії розуміння ПМТ людиною і неможливістю формалізації цього процесу стосовно комп'ютерних систем. Тому у процесі розробки ПМ-систем часто накладаються різні «спонтанні» обмеження, наслідки яких не до кінця усвідомлюються розробниками. Ці обмеження призводять до того, що чимало людино-машинних систем, на розроблення яких витрачаються величезні кошти, під час промислового використання не задовольняють вимоги 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 розробленням теоретичних основ і методології проектування, що охоплюють 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 тощо. Кожна з перелічених складових загальної методології проектування являє собою складну інформаційно-алгоритмічну структуру. Наприклад, розроблення онтології ПдО тісно пов'язано з концептуалізацією онтологічних категорій, розробленням 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* під час вирішення конкретних задач в прикладних областях.

У процесі проектування знання-орієнтованих інформаційних систем (і зокрема ОКІС) істотним є вибір як формально-логічного представлення знань, так і джерел придбання і поповнення знань. Наразі 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, у т. ч. NLP-технології.

Під час створення ОКІС варто виділити три аспекти досліджень – онтологічний, логічний і методологічний. Ці аспекти мають свої, в загальному випадку фіксовані об'єкти досліджень, відповідні процеси пізнання або розробці деякої ПС. У зв'язку з ними всі об'єкти може бути поділено на *три групи*: система сутностей (або об'єктів реального світу), система знань і система обробки сутностей відповідно до цієї системи знань. Перша група є предметом онтологічного дослідження, друга – логічного дослідження, а третя – методологічного. При цьому під методологією будемо розуміти сукупність прийомів, методів і механізмів з їх взаємодією, що застосовуються в процесі досліджень.

Значна частина досліджень належить до обробки ПМО або лінгвістичного корпусу текстів (ЛКТ) із заданої предметної області. Загальноприйнятого визначення для поняття «ПМО» невідомо, а для поняття «ЛКТ» змістовний огляд загальноприйнятих визначень виконано в [4]. ЛКТ – це будь-яка колекція текстів певної тематики (предметно-проблемної області), що доступні в електронній формі, а також результати їхнього спільного аналізу. Там само наведено методику формування довільного ЛКТ.

Варто уточнити смисли певних вживаних термінів, які не завжди збігаються з їх загальноприйнятим вживанням. Ці терміни належать до вхідної інформації для ЗОІС і згруповані в категорію «Природномовна знакова система». Нижче в онтологічній ієрархії знаходяться два терміни-синоніми – множина природномовних об'єктів і лінгвістичний корпус текстів. Нижче цих термінів також знаходяться два терміни-синоніми – природномовний текст і природномовна інформація. І на найнижчому рівні знаходяться два терміни-синоніми – природномовне висловлювання і природномовне речення. Останні терміни є синонімами, оскільки розглянуто лише стилі текстів «науково-технічний» і «ділова проза». З цього випливає, що усі речення таких текстів є розповідними і можуть розглядатися як висловлювання.

1.4. Формалізовані підходи добування та представлення знань

Проблема отримання знань з ПМО, їх представлення та обробка належать до сфери ШІ. Цю надзвичайно важливу проблему вивчали видатні учені та філософи (Арістотель, Лейбніц, Тарський та ін.). У загальній проблемі аналізу та розуміння ПМО, у всьому розмаїтті робіт, методів, методологій та термінології простежуються такі завдання:

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

Варто розглянути ці завдання більш детально.

Проблема аналізу ПМО 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.

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 простежується 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 є незручним, оскільки він спілкується (або, принаймні, повинен спілкуватись) з системою на природній мові.

Комп'ютерне представлення відіграє визначальну роль для системного аналітика, який використовує мову формалізації для побудови моделі предметної області, створює орієнтовані на цю ПдО словники понятійних одиниць і на їх основі – інтелектуальні системи, наповнені конкретними знаннями. Аналітик має справу з конкретною множиною термінів та ПМ-конструкцій, що дають тлумачення в термінах мови формалізації, встановлювати між ними інформаційно значущі зв'язки. Тому, у 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 прийнятними наразі є мови логічного типу, в яких поняття «зв'язок за змістом» формалізується з максимальною повнотою. Одним з головних питань є питання про рівень абстрактності зафіксованих категорій логічної мови. Якщо нас цікавить лише загальна схема логічного слідування, то достатньо розглядати лише числення предикатів першого порядку з правилом резолюцій у ролі правила виводу, не конкретизуючи нічого, окрім списку аксіом і логічних констант. Однак за допомогою більш багатих за змістом ПдО повинно бути вирішено питання про типи термових констант і вигляді нелогічних аксіом (власних аксіом), що використовуються у процесі формалізації знань про емпіричні об'єкти.

Викладене свідчить про актуальність проблеми розроблення мови логічного типу для представлення знань, що містяться в тексті. Вочевидь, що ця мова має бути розширюваною у бік здатності до більш повного опису об'єктів, а також здатності формувати передбачення (гіпотези), які враховують контекст.

1.5. 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, що є однією з форм декларативного програмування.

Нехай дана множина D , на якій визначено певну кінцеву 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 $L \in R \subseteq D$. Проблема виконання обмежень формулюється наступним 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 $\{C_p, \dots, C_q\}$.

Кожне обмеження $C_i \in C$ – це пара (s_p, R_i) , де

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

$R_i \in L$ – n -арне обмеження на D , що 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 такому разі 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 сфера проблеми, інтерпретується як множина об'єктів, добутих з вхідного тексту T , яке факторизовано за певними відношеннями еквівалентності R (це відношення названо відношенням синонімії), де «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 тощо).

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.

Під час аналізу ПМТ першочерговим завданням є побудова двох фундаментальних відношень, присутніх практично у кожному ПМТ. Це відношення еквівалентності і часткового порядку, що названі загальнозначущими. Перше відношення визначає класи синонімічних об'єктів, а друге – ієрархію підпорядкованості класів еквівалентності. Обидва ці відношення становлять 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 зв'язки («якщо-то»), атрибутом відношення і т. д.

Як зазначено у [26], найбільш перспективними представляються ЗОІС, функціонування яких спирається на автоматизацію процесу вилучення та формалізації змісту природномовних 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.

1.6. 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, у результаті якої будується інтерпретаційна модель предметних знань (у широкому сенсі) [27]. У процесі аналізу останні

діляться на інваріантні та прагматичні знання, концептуальні складові яких представляють онтологічні знання ПдО.

Певні ідеї з розроблення методології проектування онтології ПдО беруть свій початок в літературі з об'єктно-орієнтованого підходу (ООП), що виник як технологія програмування великих програмних продуктів [18]. 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 ухвалює проектні рішення, засновані на операторних властивостях класу, тоді як розробник онтології ухвалює ці рішення, ґрунтуючись на структурних властивостях класу. У результаті структура понять і відношення між поняттями в онтології відрізняються від структури класів об'єктів подібної ПдО в об'єктно-орієнтованій програмі [19]. Окрім того, під час розробки онтології внутрішній зміст поняття есплікується завжди, у той час як в об'єктно-орієнтованому програмуванні найчастіше застосовується метод інкапсуляції як спосіб обмеження доступу до внутрішнього вмісту об'єкта.

Системний підхід до пізнання орієнтує аналітика на розгляд будь-якої ПдО з позицій закономірностей системного цілого та взаємодії його складових частин. Системність знань ґрунтується на багаторівневій ієрархічній 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.

Період 1990-х років вважається етапом зародження парадигми комп'ютерних онтологій. Її було сформульовано як спробу знівелювати (і за можливістю усунути) різного роду суперечності у процесі функціонування та впровадження інтелектуальних систем з використанням БЗ предметних областей. Яскравим представником таких систем на той час були експертні системи (ЕС). У розробників постійно виникало запитання: «Як забезпечити їх просування до кінцевого користувача?». Було запропоновано розроблення оболонок – «пустих» ЕС і низки інших нововведень, проте вони не мали вирішального значення. Напрошувався висновок, що для ефективного функціонування

ЕС на найважливішому етапі «життєвого циклу» – функціонування у кінцевого користувача під час вирішення реальних завдань – за кожної ЕС необхідно закріпити експерта у відповідній ПдО. Необхідність присутності експерта пояснювалася, зокрема швидкоплинністю зміни знань у багатьох предметних областях і відповідно необхідністю оновлення БЗ ЕС в режимі реального часу. У тих же ПдО, де знання мали відносну інваріантність, ЕС продовжували ефективно функціонувати.

Вищесказане і низка 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 у семіотиці і мають висловлювати загальноприйняті смисли реальних об'єктів. Ці смисли вилучаються із загальноприйнятих визначень термінів (понять), зафіксованих у 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 та їх додаток має бути обґрунтованим, визначеним насамперед вимогами передбачуваної сукупності розв'язуваних завдань. Логічні аксіоми початкового набору понять повинні бути несуперечливими. Для цього має бути передбачено механізм логічного виведення, який також перевіряє на несуперечність аксіоми, що додаються і виведені в онтології твердження.

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

4. Мінімальний вплив кодування (Minimal encoding bias). В онтологічній системі (ОнС) має бути реалізовано принцип спільного використання онтологій, який передбачає: специфікацію онтології на рівні повноцінного представлення, а не символічного кодування; запис такої специфікації на загальноприйнятій і платформонезалеж-

ній мові опису онтологій можна передати для використання будь-якому програмному агенту.

5. Мінімальні онтологічні зобов'язання (Minimal ontological commitment). Цей принцип схожий на принципи обґрунтованості та розширюваності / обмеження. Важливо, щоб множина понять онтології відображала концептуальну структуру ПдО, відносно стабільну впродовж «життєвого циклу» ОнС. А остання має надавати можливість розширення або спеціалізації окремих гілок онтологічного графа. Відділення концептуальних знань від знань, виражених фактами, є стратегією побудови ОнС, а точніше – онтологічних БЗ.

Зупинимось 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 у процесі реалізації технології автоматизованої побудови онтології ПдО.

Наведені нижче 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 [30]: «*Онтологія – це формальна теорія, що обмежує можливі концептуалізації світу*». На основі цих визначень різні дослідники формулюють власні, окремі визначення онтології ПдО – відповідно до їх конкретної області професійних інтересів. Тематика і практична спрямованість наших досліджень (онтологія як засіб побудови БЗ трансдисциплінарних наукових досліджень) зумовлює таке визначення комп'ютерної онтології ПдО.

Визначення 1.5

Комп'ютерна онтологія ПдО – це:

- 1) ієрархічна структура скінченної множини понять, що описують задану предметну область;
- 2) структура є онтографом, вершинами якого є поняття, а дугами – семантичні відношення між ними;

3) поняття і відношення інтерпретуються відповідно до загально-значущих 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 підходящій формальній теорії.

Онтологію певної ПдО формально представляють впорядкованою трійкою [3, 5, 10, 30]:

$$O = \langle X, R, F \rangle, \quad (1.1)$$

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): $R = \emptyset$; $R \neq \emptyset$; $F = \emptyset$; $F \neq \emptyset$ в усіх чотирьох комбінаціях значень і дає різні варіанти онтологічних конструкцій, починаючи від простого словника і таксономії до формальної структури концептуальної БЗ для високоінтелектуальних знання-орієнтованих систем. Детальний розгляд різних зазначених комбінацій з їх змістовною інтерпретацією виконано в [21].

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

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

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 усіх компонент представлено на певній формальній мові, яка доступна для їх інтерпретації комп'ютером. Схема формальної моделі повної онтології описується таким чином:

$$O = \langle X, R, F, A(D, R_s) \rangle, \quad (1.2)$$

³Класифікації онтологій за різними підставами наведені в численній літературі [4].

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, що використовуються для запису завжди істинних висловлювань (визначень і обмежень);

D – множина додаткових визначень понять;

R_s – множина обмежень, що визначають сферу дії понятійних структур.

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

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

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

Така схема класифікації за функціональною ознакою узгоджується з описом: «Онтологія або концептуальна модель предметної області складається з ієрархії понять предметної області, зв'язків між ними і законів, які діють в рамках цієї моделі» [4].

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].

Головною ідеєю системно-онтологічного підходу (СОП) є розроблення онтологічних засобів підтримки рішення прикладних

задач – поліфункціональної онтологічної системи. Така система (точніше, її концептуальна частина) описується (1.3), що охоплює онтологію ПдО (складається з онтології об'єктів і онтології процесів) і онтологію задач [22, 33].

$$O_{HC} = \langle O^{P/O} (O^O, O^P), O^3 \rangle \quad (1.3)$$

На рисунку 1.7 зображено схему онтологій-компонент предметної області та *проблемного простору* (ПрП). ПрП – це модель всіх таких аспектів або компонент ПдО, з якими пов'язані (опосередковано або безпосередньо) знання, що необхідні для вирішення різних задач у цій ПдО.

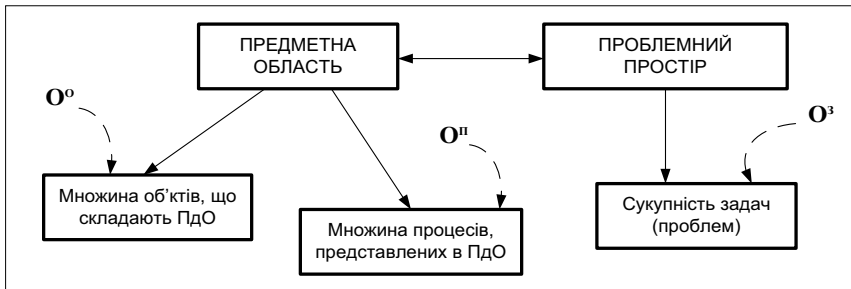


Рис. 1.7. Схема онтологій-компонент предметної області

Будь-який ПрП складається з двох блоків: інваріантної (відносно незмінної) частини і множини змінних частин, що відповідають 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 [18]. Типи компонент – суть поняття, що групуються в онтології об'єктів і процесів:

- O^O – онтологія множини об'єктів (понять, концептів) ПдО, що розглядається як ієрархічна структура класів, підкласів та елементів класів;
- O^P – онтологія множини процесів ПдО, що розглядається як ієрархічна структура процесів, підпроцесів, дій та операцій;

• O^3 – онтологія сукупності задач (типових наборів), що можуть бути поставлені і вирішені в ПдО. Розглядається як ієрархічна структура задач, підзадач, процедур і операторів.

1.7. 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 ці поняття можуть бути виражені в текстах. Процедура порівняння понять і мовних виразів може бути реалізована різними способами [21].

1. До понять вже розробленої онтології ПдО приписуються мовні одиниці. За складністю реалізації цей спосіб є найпростішим, проте йому властива низка очевидних недоліків:

- імена понять досить громіздкі, неприродні, ними важко оперувати як розробникам, так і можливим користувачам;
- надмірна багатозначність лексичної одиниці, що виникає через те, що одне і те ж слово може відповідати занадто великій кількості понять залежно від контексту;
- онтологічна гіпотеза Сепіра-Уорфа (те, що не описується словами, не може бути відображено в онтології).

2. Встановлення відповідності між лексичними ресурсами (наприклад, WordNet) і певною (концептуальною) онтологією. WordNet-ресурси описують лексичні відношення між значеннями слів, представлені у вигляді окремих одиниць в ієрархічній мережі – синсетів. Відношення між лексичними одиницями значною мірою 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, що проявляється у відношеннях між

поняттями. Тому лінгвістичні онтології можуть розглядатися як особливий вид лексичної БД і особливий тип онтології.

3. Розроблення єдиного ресурсу (змішаної онтології), в якому було б збалансовано обидві частини: система понять і система лексичних значень, що досягається розумним розподілом цих одиниць в створюваному ресурсі та 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, у т. ч. природномовними текстами) в різних предметних областях. Нижче наведено загальні відомості про лінгвістичної онтології WordNet.

WordNet. Лінгвістична онтологія для лексики англійської мови розроблялася в середині 1980-х років в лабораторії когнітології Принстонського університету (штат Нью-Джерсі, США) під керівництвом професора Дж. А. Міллера. Вільно поширюється в Інтернеті та являє собою лексикографічну БД, структуровану відповідно до лексико-семантичних відношень: синонімії, антонімії, гіпонімії, меронімії та морфологічних відношень [10, 11].

Найважливішим відношенням WordNet є подібність значень. В основу поняття «синонімія» покладено два слова, що є синонімічними в контексті *S*, якщо заміна одного на інше в *S* 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 один одного).

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 головним організуючим принципом для прикметників і прислівників.

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

Іншим типом відношень є «частина – ціле», що відоме в лінгвістиці як метонімія / голонімія. Поняття, представлено 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 декілька голонімів).

Морфологічні відношення між формами слів є важливою частиною лексичних відношень. У системі WordNet спочатку інтерес обмежувався лише семантичними відношеннями: не було проєктів, які б враховували морфологічні відношення. Проте з часом програми удосконалювалися і стало очевидним, що WordNet повинен служити будь-яким запитам, у т. ч. пов'язаних з морфологічними характеристиками [12].

БД 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, а 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 вищеперерахованих) задають типи ребер з R_2 . У WordNet виділено 14 базових типів таких відношень (окрім 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;
- дія, 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, що визначає широкий клас значень слів, пов'язаних між собою семантичними покажчиками. Логічну структуру WordNet показано на рисунку 1.8.

Розглянемо приклад запису синсету в БД:

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

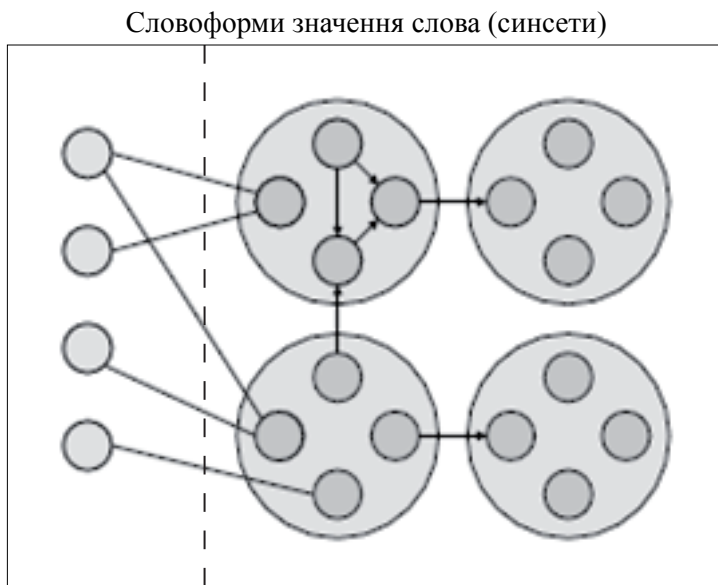


Рис. 1.8. Логічна структура WordNet

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 приклади вживання слів, що входять в синсет.

Онтологія WordNet популярна в наукових дослідженнях з комп'ютерної обробки природної мови та пошуку інформації в Інтернеті (Information Retrieval). Вона постійно оновлюється. Наразі для Windows-користувачів доступна версія 2.1, для UNIX-подібних ОС – версія 3.0 і в ній існує понад 82 000 синсетів іменників, більше 18 000 синсетів прикметників, більше 3 600 синсетів прислівників і майже 14 000 синсетів дієслів. Усього вона охоплює більше 147 000 англійських слів і фраз. Інші аспекти опису WordNet і EuroWordNet та велику кількість прикладів їх використання наведено в [11, 20, 32].

Відомі й інші (змішані та концептуальні) онтології верхнього рівня Sensus [13], Omega [14], Cys [15], DOLCE [16], PropBank [17, 18],

FrameNet [9–11]. Певні дані, що характеризують згадані онтології, наведено в таблиці 1.1. Аналіз публікацій свідчить про те, що найбільший інтерес (окрім лінгвістичної онтології WordNet) для розробників представляють онтології Mikrokosmos, SUMO і Cус.

Верхній рівень онтології Cус 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. Окрім того, Cус найбільш розроблена частина проекту. Вона є комерційною і недоступною для загального користування. Існує і доступна частина проекту онтології (OpenCус). Серед позитивних моментів Cус можна назвати наявність формальної мови опису CусL, БЗ і широку аксіоматизацію, що постійно розвиваються та доступні в режимі клієнт-сервер [15].

Онтологія Mikrokosmos має досить гнучку ієрархічну структуру з явним пріоритетом групи категорій I&R&O. Дві інші групи категорій розподілено за гілками ієрархічного дерева і не мають чітко виражених обрисів. Онтологія має практичну спрямованість, у якій не всі гілки дерева розроблено 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 онтології доступна як на веб-порталі розробників, так і в численних публікаціях [13, 14, 19–21, 24–31].

Численні фрагменти онтології SUMO є варіантами адаптації різних ОВР в одну онтологію верхнього рівня. У ній явно виражений пріоритет групи категорій M&A з подальшим переважанням групи категорій C&O. Варто 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 цю онтологію можна рекомендувати для адаптації з онтологіями предметних областей, де явно переважають матеріальні концепти.

Таблиця 1.1

Основні характеристики онтологій

№	Назва онтології	Тип онтології	Розробник	Браузер	Кількість концептів (лексем)	Формальна аксіоматизація	Інтеграція з іншими онтологіями
1	WordNet	Лінгвістична	Princeton University	WordNet	Приблизно 100 000	–	SUMO, FrameNet
2	Sensus	Змішана	ISI USC	Ontosaurus	Понад 70 000	–	–
3	Omega	Лінгвістична	ISI USC	Omega	Понад 120 000	–	WordNet, Mikro-kosmos
4	Mikro-kosmos	Лінгвістична	CLR UNMS	Mikro-kosmos	Понад 7 000	+	WordNet, Omega
5	OpenСус	Концептуальна	Сусорр	Opencус	Понад 100 000	+	WordNet
6	DOLCE	Концептуальна	LAO ICST	DOLCE	Понад 4 000	+	–
7	PropBank	Змішана	University PennState	–	Понад 4 300	–	FrameNet VerbNet
8	FrameNet	Змішана	ISI, Berkeley, CA	Frame Grapher	Приблизно 900	+	WordNet PropBank SUMO
9	SUMO	Концептуальна	Teknowledge Corporation, SUO WG	SUMO	Понад 1 000	+	FrameNet, WordNet, EMELD

Мовно-онтологічна картина світу

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

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

- відповідати певним 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 вирішення проблеми потенційної многозначності вираження змісту вхідного тексту;
- надавати можливості використання як семіотичної основи для фіксації змісту вхідної інформації;
- виконувати функції мовно-категоріальної надбудови БЗ у конкретних предметних областях.

Визначення 1.6

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 впливає з визначення, МКС є змішаною онтологією, лінгвістична частина якої названа мовно-онтологічною картиною світу. Остання є головною компонентою мовно-онтологічної інформаційної системи. У МОКС включені лише загальнонаукові, загальноживані (або енциклопедичні) поняття, тобто основні лексичні одиниці тлумачного 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 разом з іншими метаонтологіями загального характеру. Звідси одне з призначень МОКС – служити ланцюгом, що 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'язано з побудовою ієрархічних структур категорій верхнього рівня. Побудова системи категорій є необхідним етапом розроблення інтелектуальних систем, заснованих на знаннях, а також систем, які забезпечують автоматизацію розуміння змісту тексту. 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 передбачає побудову моделі семантичних відношень у лексиці як деякого смислового каркасу, на якому базується глобальна «сферична» мережа різнотипних семантичних відношень. Смісловий каркас будується на основі 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 та класифікації понять, а по-друге – неповнотою лексичного континууму і відношень між представленими лексичними одиницями (насамперед розуміється не їх теоретичний доробок, а комп'ютерна обробка, точніше – комп'ютерне моделювання). По-третє, вони відокремлені від сучасних інтелектуальних інформаційних технологій та їх додатків.

У цій роботі:

– за основу було взято ієрархічну структуру категорій верхнього рівня, синтезовано на логіко-філософських засадах;

– повнота лексичного континууму забезпечується відтворенням у відповідних таблицях лексикографічної БД повної множини лексем тлумачного словника;

– дослідження в своїй кінцевій меті орієнтовано на розробку апаратно-програмних засобів підтримки (граматичного процесора) на сучасній мікроелектронній базі.

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

Окрім того, потрібно 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 у мовно-онтологічній інформаційній системі. Його основною функцією є підтримка побудови формалізованого опису відображення вхідного ПМТ з відтворенням повної акторної складової семантики, наявній у ньому.

Таке відображення можна записати:

$$G : C_O(EAT) \rightarrow C_A(EAT), \quad (1.4)$$

де $C_O(EAT) = \cup_{i=1}^N C_{Li}$, де C_{Li} – формалізований опис i -ї 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 під час взаємодії лінгвістичного процесора (ЛП), СП і процесора інтерпретації БЗ і БЗ 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, сформульова-

них певною мовою на рівні здорового глузду, пов'язаних максимально повною системою відношень, що відображають навколишній світ з усію множиною його об'єктів і явищ, тобто представляє собою лінгвістичну проекцію буття, в якій зафіксовано досвід взаємодії людини з навколишньою дійсністю. МОКС – це складова частина прагматичної моделі мовної свідомості, що є головною компонентою сучасних інтелектуальних ІС з природномовним представленням, обробкою та актуалізацією знань.

Визначення 1.7

МОКС – це відкрита, експліцитно задана на лексико-смисловому 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

$$O = \langle X, R, F \rangle, \quad (1.5)$$

де X – кінцева множина понять, характеристик, ролей і атрибутів (або контент-слів), виражених лексичними засобами ПМ, насамперед такими повнозначними частинами мови, як іменник, дієслово, прикметник і прислівник. У логіці вони представляються 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 та логічного представлення вихідного ПМТ.

Розвиток абстрактної моделі МОКС

Варто уточнити зміст поняття ієрархічного структурування в онтології МОКС щодо лексичного континууму природної мови. Смісл будь-якого висловлювання, як такого, виникає, коли текстові компоненти

вступають у певні діючі відношення (об'єкт ↔ дія). Сфера дії визначається кількістю об'єктів, які вона охоплює.

Такий розгляд спрощено описує структуру відображення та більше орієнтовано на опис об'єктної складової семантики. Очевидно, що для повного відображення побудованих лінгвістичним процесором синтаксичних структур вхідного ПМТ, а також опису факторної складової тексту, необхідно враховувати базові повнозначні частини мови.

В узагальненій схемі взаємодії структурних компонент МОКС за основу беремо 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 узгоджується з відомою концепцією WordNet. Ієрархічна структура МОКС повинна являти собою деякий гіперграф, як композицію чотирьох ациклічних орієнтованих графів ієрархічних структур іменника, дієслова, прикметника та прислівника. Кожен із цих графів будується відповідно до своїх концептуальних та лінгвістичних відношень, властивих зазначеним частинам мови (відображення семантичних особливостей об'єднання іменника з прийменником враховується у графі для іменника).

Відповідно до вищенаведеного опису на рисунку 1.9 зображено абстрактну модель мовно-онтологічної картини світу. У моделі використано такі 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);

$КГ_{AO}$, $КГ_O$, $КГ_D$ та $КГ_{AD}$ – елементи семантичних структур першого ступеня (відповідно концептуальні графи атрибута об'єкта, об'єкта дії і атрибуту дії);

З*, 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 усіх повнозначних частин мови та відношень між ними. Таке спрощення є справедливим, тому що в цій моделі зроблено акцент саме на онтологічну складову.

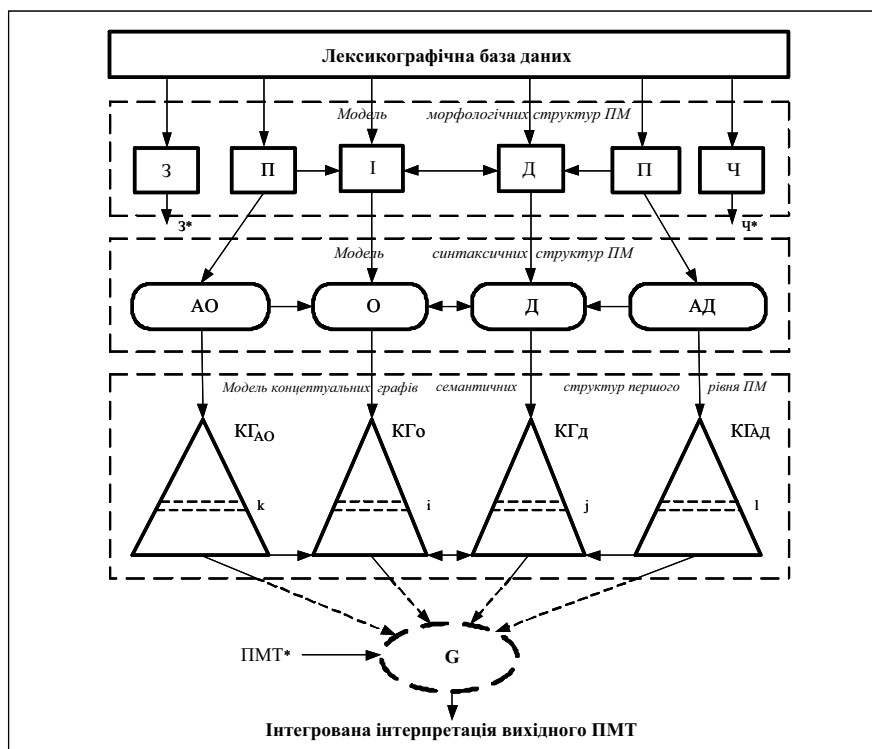


Рис. 1.9. Абстрактна модель МОКС

Як було визначено, онтологічний граф не охоплює поняття нижнього рівня. Проте формалізованого критерію визначення границі нижнього рівня не існує, оскільки він повністю залежить від інтуїції та професійних поглядів дослідника.

Одним із етапів розширення абстрактної моделі є включення в лексикографічну базу даних (ЛБД) додаткових таблиць, які охоплюють як повнозначні (дієприкметник і дієприслівник), так і службові частини мови (прийменник, сполучник і частка) і стійкі словосполучення. Таблицю дієприслівників доцільно 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 низку інших граматичних особливостей української (російської) мови.

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

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

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

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

КГ атрибуту об'єкта розділяється на три підграфа, відповідно якісних, відносних і присвійних атрибутів об'єкта.

КГ атрибуту дії пов'язується з відповідними групами семантичних тем дієслів.

Формальне відображення є семантичною моделлю другого рівня (або акторної семантики) і охоплює підмножини логіки першого порядку концептуальних графів, обчислення предикатів та мови представлення знань (KIF), спроектованих на вихідний ПМТ.

У таблиці 1.2 наведено належним чином структуровані формальні основи застосування правил, принципів, абстракцій та відношень у процесі розвитку метаонтології категорій верхнього рівня і МОКС (окремо для рівнів концептів і примітивів) [5, 26–28, 32], а також деякі правила-рекомендації, які необхідно враховувати під час 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 та є основними стандартними блоками для визначення більш складних відношень у концептуальних графах, БД або експертних системах. Наведемо деякі правила-рекомендації, яких необхідно дотримуватися у процесі побудови онтології предметної області [32].

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

Таблиця 1.2

Формальні основи структурування онтологій

I. Метаонтологія категорії верхнього рівня	
Види абстракцій	Класифікація, узагальнення, агрегація та асоціація
Принципи класифікації	Фундаментальні принципи дихотомії Арістотеля, трихотомії Пірса і решіток Лейбніца, природна класифікація
Типи класифікаційного поділу	Таксономічні, мереологічні, просторові та часові
Спадкування ознак	Множинне

Продовження таблиці 1.2

II. Мовно-онтологічна картина світу	
A. Рівень концептів	
Концептно-рольові відношення	Виробляють два репрезентативні примітиви, відповідно названі приєднанням і кореферентним зв'язком. На природну мову переводяться різними граматичними формами дієслів <i>мати</i> і <i>бути</i>
Семантичні відношення	<i>Класифікаційні</i> : «рід – вид», «ціле – частина», «клас – елемент», «вищерозташований – нижчерозташований», «клас – підклас»; <i>Ознакові</i> : «об'єкт – атрибут», «об'єкт – дія»; <i>Кількісні</i> : мати міру
Лінгвістичні відношення	Гіпонімія, метонімія, синонімія й антонімія
B. Рівень примітивів	
Предикативні відношення	Узагальнено-одиночного характеру, окремо-одиночного характеру і одиночно-одиночного характеру
Семантично-рольові відношення	<i>Класифікаційні</i> : мати ім'я, бути еталоном; <i>Ознакові</i> : мати значення атрибута, мати значення способу дії; <i>Кількісні</i> : мати значні міри; Порівняння: рівне, порівняльне, більше, більше або дорівнює, менше, менше або дорівнює, незрівняне; <i>Приналежності</i> : простої приналежності; <i>Прості часові</i> : бути одночасно, бути раніше, бути пізніше, збігтися в часі, перетинатися в часі, бути всередині в часі, починатися одночасно, закінчуватися у часі; <i>Прості просторові</i> : 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; <i>Каузальні</i> : бути метою, бути мотивом, причина-наслідок;

Продовження таблиці 1.2

Семантично-рольові відношення	<i>Інструментальні</i> : служити для, бути засобом для, сприяти, бути інструментом, бути допоміжним засобом; <i>Інформаційні</i> : бути відправником, бути одержувачем, бути джерелом інформації; <i>Порядкові</i> : бути наступним, бути поточним, бути найближчим; <i>Модальні</i> : можливості, дійсного втілення, необхідності; <i>Модифікаційні</i> ; <i>Квантифікаційні</i> : загальності, існування, унікальності, одиничності, можливості, необхідності; <i>Кореляційні</i>
-------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

2. *Процедура переходу повинна виконуватися за однією основою або характеристика, що 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.*

Одним з можливих варіантів перекладу ПМ-речень у формальні структури може бути переведення у концептуальні графи [22], оскільки вони для цього розроблялися. Вони забезпечують читабельність, наочний ітераційний процес виявлення помилок перекладу і в той же час є підмножиною логіки першого порядку. Окрім того, концептуальні графи мають можливість формування процедури вкладеності, що відповідає ланцюгу синтаксичних одиниць «фраза → речення → абзац → параграф → розділ → текст → том» або фрагментам дійсності «елементарна ситуація → ситуація → сценарій».

1.8. Системна інтеграція лінгвістичних і предметних онтологій

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

Загалом процес, що забезпечує системну інтеграцію множини онтологій, можна описати наступною формулою

$$\mathbb{O} + \bigcap_i \bigcup O_i, \quad i = \overline{1, N} \quad (1.6)$$

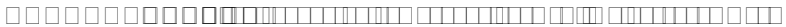
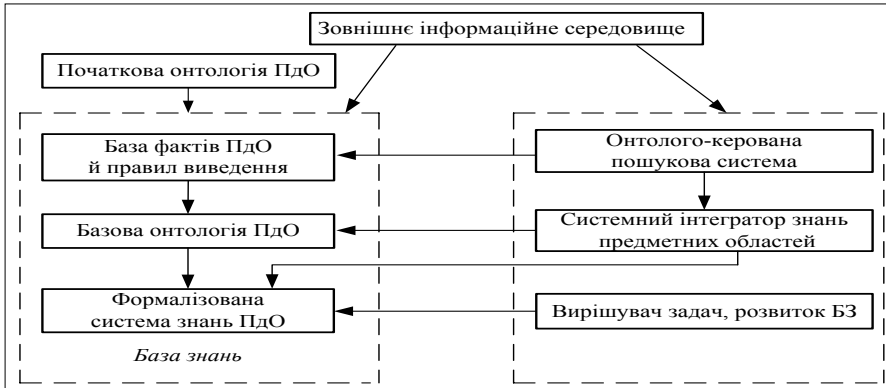
де \bigcap – знак концептуального об'єднання. Смісл цього знака полягає в системній інтеграції вихідних онтологічних графів з урахуванням сфер визначень $O_i, \quad i = \overline{1, N}$ і їх взаємодії.

Узагальнена архітектура знання-орієнтованої системи ефективно реалізує процедури обробки предметних знань, розвитку БЗ та системної інтеграції. Її зображено на рисунку 1.10.

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, що 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, у т. ч. за підтипами).

Пізнання починається з вивчення речей до вивчення процесів (у т. ч. взаємодії компонент, охоплюючи 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 у предметних областях можна оцінити через характеристики (параметри) їх формально-онтологічних представлень. Зокрема, під час представлення онтологічним графом (без урахування

типів відношень і складності функцій інтерпретації) величина W може характеризуватися числом вершин ОГ. У разі простої дерево-видної структури це число може бути виражене формулою

$$W = \sum_i \sum_h \sum_l O_i \cdot S_{h,l} \quad (1.7)$$

де O_i – онтограф i -ї 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.

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

$$W = \sum_i O_i \frac{1-S^h}{1-S} \quad (1.8)$$

Врахування типів відношень і складність функцій інтерпретації призводить до ОГ зі зваженими вершинами і ребрами. При цьому вираз (1.8) набуває вид

$$W = \sum_i \sum_h O_i \cdot \left[\alpha_i + \sum_j \beta_{i,j} \right], \quad (1.9)$$

де α_i і $\beta_{i,j}$ – значення вагових функцій відповідних відношень і функцій інтерпретації, приписані вершинам (α_i) і ребрам ($\beta_{i,j}$) ОГ. Вираз дає повну оцінку складності ОГ, а відношення $\omega = W^o/W$ 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'язано з її аналізом, концептуалізацією і побудовою формальної теорії. При цьому формалізація належить до чотирьох основних видів представлення інформації

$$I = I(V, A, T, G), \quad (1.10)$$

тобто до вербального (V), аналітичного (A), табличного (T) і графічного (G). Між ними існує взаємно-однозначна відповідність, що на практиці не завжди реалізується строго і повністю. Тому всі вони зна-

ходять своє, цілком певне місце при описі наукової теорії. У 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 Φ , 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 (WordNet і FrameNet) підтвердив 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 БД WordNet шляхом тегування відповідного концепту в SUMO. Види відношень між синсетами WordNet і концептами SUMO можуть бути різними:

- синсет еквівалентний концепту;
- синсет є більш вищим класом, ніж концепт;
- синсет є елементом концепту.

Розроблені файли відображення дозволяють співвіднести слова природної мови у терміни SUMO, використовуючи синсети WordNet як проміжний рівень [19].

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, що дозволяє користувачеві задавати терміни природної мови і бачити на екрані комп'ютера відповідні концепти онтології SUMO з відповідною формальною аксіоматизацією. Це дозволяє значно спростити процес інжинірингу знань і моделювання даних. Окрім того, відображення можуть слугувати важливим джерелом знань для застосувань NLP, зокрема під час реферування текстів і семантичному пошуку документів у 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 приблизно 40 семантичних типів (FT) як обмеження на заповнювачі FEs і мають відповідності з SUMO класами як FTs [12]. Фрагмент зв'язування SUMO і FrameNet представлений на рисунку 1.11.

Варто зазначити, що існують й інші проекти зв'язування лексикон-онтологій [21]. Але розглянуте зв'язування SUMO і FrameNet має низ-

ку переваг. FrameNet, на відміну від WordNet, моделює семантичні та синтаксичні валентності, охоплюючи високоякісні анотації. Фреймові семантики природно представляють крослінгвістичні абстракції та нормалізацію парафраз. Що стосується SUMO, то вона значно ширша DOLCE, а на відміну від Сус є відкритою для доступу.

Під час інтеграції знань варто враховувати прикладні аспекти. Принципово розрізнати сфери:

1) взаємодії з користувачем – когнітивізація представлення, пошук інформації, створення початкової онтології та початкової системи знань;

2) розвитку системи знань на основі наявних, генерація нових знань.

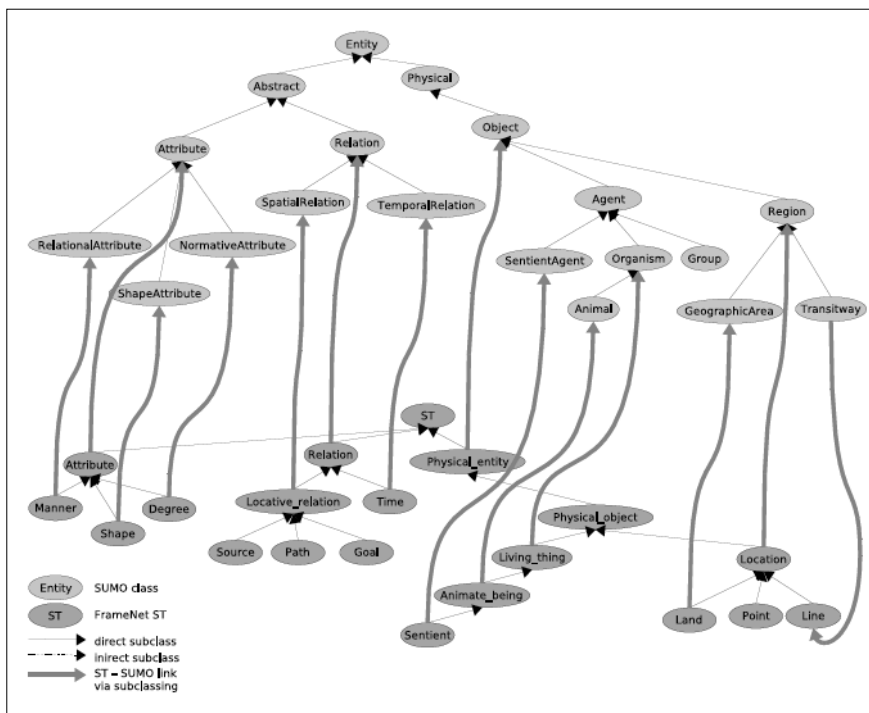


Рис. 1.11. Фрагмент зв'язування SUMO і FrameNet

1.9. Методологія розробки онтології ПдО

Загальний підхід до проектування

Відома методологія структурного аналізу і проектування (SADT) складних систем у довільній предметній області [22]. Ця методологія породила сімейство методик (і відповідних стандартів) IDEF (Integrated DEFinition), орієнтованих на розробку моделей ПдО й акцентуючих увагу на певному конкретному аспекті проектування [23, 24]. Зокрема:

– методика IDEF0 рекомендована для змістовного аналізу та функціонального проектування складних систем управління, у т. ч. і програмного забезпечення. Опис об'єктів і процесів у 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 відправною точкою 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 запитання «Що робить система?», то в IDEF3 деталізуються IDEF0-функції. Ця модель відповідає на запитання «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 для опису 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 розроблено 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 вони (функціональні моделі та описи) є змістовним технічним завданням на проектування ОнС, яке охоплює 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 – це функціонально-блокова модель, а для системно-онтологічного підходу – онтолого-сміслова.

Вищевиконаний аналіз дозволяє стверджувати, що методики IDEF і СОП вирішують різні завдання (з певним «перекриттям» деяких функцій). Це проявляється для пари «онтологія процесів ↔ поведінкова модель ПдО».

Етапи проектування онтологічних систем ґрунтовані на прийнятих в методології SADT фазах проектування складних систем: 1) аналіз – визначення того, що система буде робити; 2) проектування – визначення підсистем та їхню 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 ПдО

У всіх методологіях наявний етап попереднього аналізу ПдО або складання змістовного ТЗ на проектування [2; 3; 5, 22–27]. Цей етап (як і процес проектування БЗ ПдО в цілому) має складний аналітичний характер і полягає в багаторазовому абстрагуванні, у результаті якого з усього різноманіття сторін і властивостей сутностей предметної області виділяються найбільш істотні, релевантні конкретним задачам. Знання предметної області, розуміння суті відповідних у ній процесів, законів, правил і обмежень, які керують її розвитком, є необхідною умовою успішного вирішення завдань, що стоять перед

дослідником. Більше того, наявність таких знань є необхідною умовою постановки, формулювання цих задач, без чого неможливе саме рішення [9]. Онтологічні системи створені для того, щоб зробити знання колективним надбанням широкого кола осіб, дати потужний інструмент для фіксації, придбання та обробки знань, перевірки їх на несуперечність, повноту тощо. Окрім 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 охоплює:

- обґрунтований вибір точного (і достатнього) фрагмента ПрП, щодо якого будуть ставитися і вирішуватися задачі користувача;
- вибір методів і процедур системно-онтологічного аналізу, якими 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) представлена складною системою, то варто розглянути питання про попередній етап проектування на основі методик IDEF, які доповнюють вищеописані кроки проектування. Тим більше, що відомі інструментальні засоби аналізу ПдО і побудови БЗ (наприклад, KADS-системи, SIMER+MIR) за певних причин не можуть бути використані [3].

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

Онтологія об'єктів ПдО

Визначення 1.8

Під онтологією об'єктів предметної області розуміється четвірка [2]:

$$O^o = \langle X, R, F, A(D, Rs) \rangle, \quad (1.11)$$

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

$R = \{R_1, R_2, \dots, R_p, \dots, R_n\}$, $R \subseteq X_1 \times X_2 \times \dots \times X_n$, $k = \overline{1, m}$, $m = \text{Card } R$, – скінченна множина семантично значущих відношень між поняттями-об'єктами ПдО. Вони визначають тип взаємозв'язку між поняттями. Відношення ділять на загальнозначущі (з яких виділяють 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^t і множини обмежень Rs_i^t для поняття X_i . Визначення записуються як тотожно істинні висловлювання, які, наприклад, можуть бути взяті з тлумачних словників ПдО. У них може бути зазначено додаткові взаємозв'язки понять X_i з поняттями X_j . У множині обмежень Rs_i можуть бути задані обмеження на інтерпретацію відповідних понять X_i .

Онтологія визначає загальноновживані, семантично значимі «понятійні одиниці знань», якими оперують дослідники і розробники знання-орієнтованих інформаційних систем. Вона відокремлює «статичні» та «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 істотною компонентою концептуальної моделі ПдО є множина понять заданої предметної області. Нагадаємо деякі твердження, що безпосередньо пов'язані з побудовою онтографа ПдО [28].

Усі поняття (або концепти) поділяються на низку класів (за семантичною залежністю).

- Залежно від відображення виду або роду предметів – на *видові* та *родові* поняття.

- Залежно від відображення частини або цілого предметів – на *поняття-частини* і *поняття-цілі*.

- Залежно від кількості відображуваних предметів – на *одиночні* і *загальні* поняття.

- Залежно від відображення предмета або властивості, абстрагованого від предмета, – на *конкретні* і *абстрактні* поняття.

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

Побудова множини X вважається найбільш важливим моментом під час розроблення 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 разі варто 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 лише одне поняття, в аксіомах якого може бути вказаний синонімічний ряд термінів. Іншими словами, синоніми одного і того ж поняття не представляють різні класи.

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

Потім, можливо, необхідно повторити певні фрагменти процесу аналізу ПдО (з прив'язкою до складеного списку понять), виконані на попередньому етапі. Варто наголосити, що до числа вищезазначених «контрольних точок» (точок входу ітерації) повинно бути включено завершення розроблення будь-якої компоненти онтології.

У результаті має бути отримано 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 $R_k \subseteq X^k$, тобто необхідно побудувати множину ребер, що зв'язують вузли *орієнтованого графа*. У ролі 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 у заданій ПдО.

На практиці множину R на початковому етапі подають деяким узагальненим відношенням «*вище–нижче*». Існує декілька підходів для розроблення ієрархії класів: процес *низхідного, зростаючого* розроблення, а також *комбінований*. Останній найчастіше використовується розробниками, оскільки він є природнішим, спочатку оперує поняттями середнього рівня, до яких найчастіше звертаються розробники. Потім ці поняття узагальнюються і обмежуються.

Наприкінці цього підетапу варто співвіднести розроблені класи та їх ієрархії з результатами попереднього аналізу ПдО, зокрема уточнюються залежності для конкретних пар (x_i, x_j) . У процесі співвіднесення (і побудови ієрархії) варто враховувати, що [8, 9]:

- 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 лише один прямий підклас, то, можливо, під час моделювання допущено помилку або онтологія неповна;
- якщо у цього 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.

Варто 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 проєктованої онтології множини F і A можуть інтерпретуватися по-різному [12, 25–27 та ін.]:

1) $A = F$ – множина аксіом тотожна множині функцій інтерпретації. У такому разі встановлюються істотні зв'язки між компонентами онтології та 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 $F: x_1 \times x_2 \times \dots \times x_{n-1} \Rightarrow x_n$. У такому разі встановлюються зв'язки між уже розробленими компонентами онтології та сукупністю задач передбачуваного додатку (додатків). Онтології з таким поданням використовуються в питально-відповідних системах, у яких результатом є одне зі значень двоелементної множини або ім'я предиката.

Насамкінець незалежно від того, яке з цих формулювань буде прийнято, ефективність розробленої онтології буде визначатися кінцевими результатами додатків.

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

Властивості понять мають певні значення: *тип* значень, *потужність* значень, *дозволені* значення (для даного класу) тощо. Наприклад, значення бувають з одиначною потужністю, потужністю без обмежень і потужністю з певним 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. Коли у науково-технічній літературі обговорюється онтологія ПдО, то мається на увазі її статична складова. Саме компоненти останньої найбільш розроблені, як в літературі з філософії, так і в конкретних описах предметних областей. Поведінковий опис сутностей-процесів найчастіше виконується графічними діаграмами і природномовними описами. Розроблення ж БЗ не є безпосередньою метою зазначених методик. Тому методики розроблення онтології процесів практично невідомі, хоча у деяких відомих онтологіях верхнього рівня [9] сутність поняття «Процес» розглянуто досить детально.

На рисунку 1.12 представлено синтезований онтограф, який представляє схему початкової ділянки поняття «Процес», а саме тієї її частини, що 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]. «Процес» насамперед розглядається як залежна від часу категорія і потім поділяється за видами змін, наявністю початкових і кінцевих точок тощо. Далі «Процес» поділяється на безперервний і дискретний. Перший з них характеризується наявністю експліцитних початкової та кінцевої точок або без явної вказівки цих точок. Другий вид процесу вказує, що зміни відбуваються дискретними

кроками, названими подіями, що чергуються з періодами спокою, названими станами. Наведена схема початкового розвитку онтології процесів не відображає всі характеристики (підстав розгалужень в онтографі) категорії «Процес».

На рисунку 1.13 подано загальну схему онтології процесу ПдО, в якій категорія «Процес» представлена онтографом з p рівнями і pr підпроцесами (Пп) на кожному рівні. Передостанній ($p-1$) рівень передано множиною дій (Д), на які розбивається кожен Пп попереднього рівня. 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 – певною змішаною формою організації з'єднань. Окремий випадок такої організації – паралельний. Подальший розвиток (конкретизація) онтології процесів можливий, коли задано конкретну предметну область і відповідний проблемний простір, а в більш вузькому сенсі – конкретні ознаки розгалужень (умови ініціювання Пп, умови закінчення Пп та обмеження) в онтографі.

Онтологія задач проблемного простору ПдО

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

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

Відомі різні схеми конструювання (синтезу) онтологічних структур з предметно-проблемних знань [13, 31]. У цій роботі в онтологію ПдО об'єднано об'єкти і процеси (статистичні знання), що згруповані в окремі онтології. Онтологія задач охоплює класи задач, методи їх

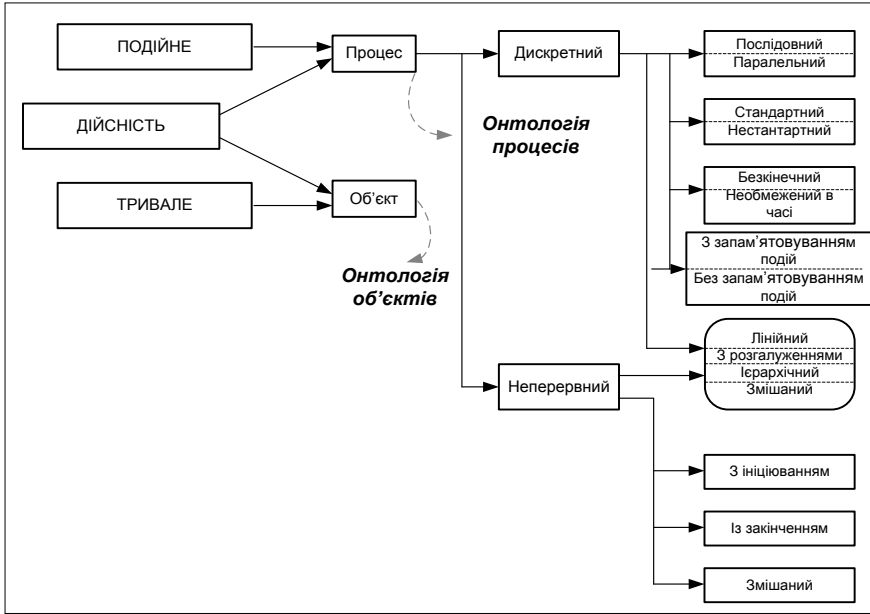


Рис. 1.12. Початкова ділянка онтології процесів

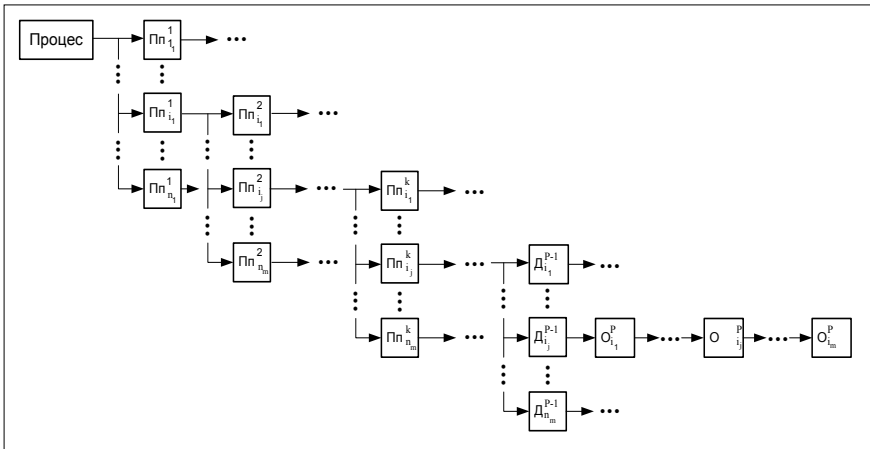


Рис. 1.13. Загальна схема онтології процесу

вирішення та відповідні алгоритми. За такої схеми групування предметно-проблемних знань, останні строго розділені на статичні та динамічні. Якщо змінюється клас розв'язуваних задач, то онтології об'єктів і процесів вже готові до повторного використання, перепроєктується лише онтологія задач [22, 23].

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

Онтологічні знання, що описують певний ПрП, можна розділити на такі компонентні знання: типи вхідних і вихідних даних, інструменти, оператори (людина або комп'ютерна програма) та операції (дії оператора або *Вирішувача задач*) [28].

Для реалізації підходу необхідно розробити уніфіковану мову представлення онтологічних знань та інструментальне середовище як набір спеціалізованих і універсальних базових операцій, що 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

$$O^3 = \langle OZ^{PrP}, M, P^3 \rangle, \quad (1.12)$$

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

$$3^p = \langle D_{in}^p, R^p, C^p, D_{out}^p \rangle \quad (1.13)$$

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, компоненти якого визначені вище.

$$R^p : (D_m^p, R^p, C^p) \rightarrow D_{out}^p, \quad (1.14)$$

де B^3 – вирішувач задач.

Однією з переваг онтологічного підходу, зокрема ієрархічного подання є те, 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 зазвичай специфікують декомпозицію задач на підзадачі.

Алгоритм проектування онтологій

На рисунку 1.14 подано блок-схему алгоритму проектування онтології ПдО і онтології задач. Онтологія ПдО представлена онтологією об'єктів і онтологією процесів, для яких у блок-схемі передбачено окремі гілки. Передбачається, що множина функцій інтерпретації в моделях онтологій тотожна множині аксіом [2]. На рисунку прийнято наступні скорочення:

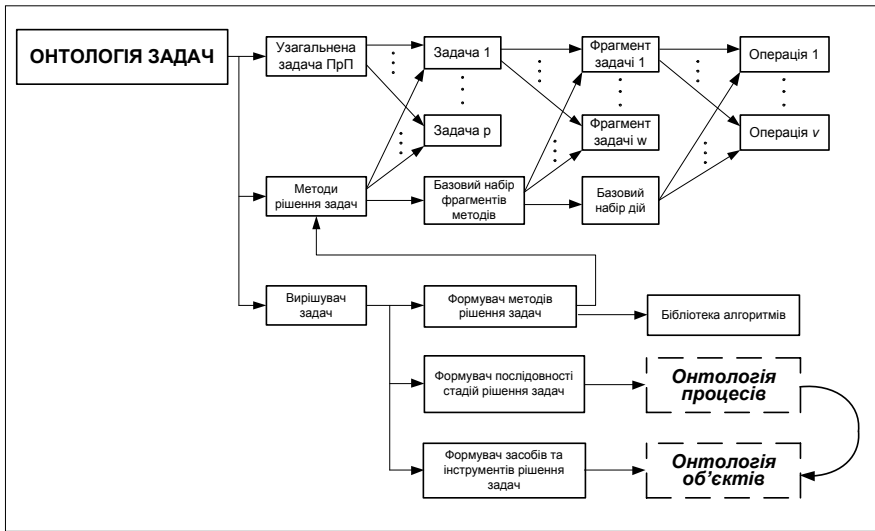


Рис. 1.14. Схема онтології задач

- 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^O, R^O, F^O – скінченні множини понять, відношень і функцій інтерпретації онтології процесів;

– X^3, R^3, F^3 – скінченні множини понять, відношень і функцій інтерпретації онтології задач;

– OG^O, OG^O, OG^3 – онтографи відповідно об'єктів, процесів і задач.

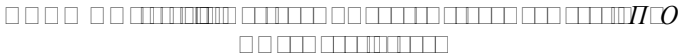
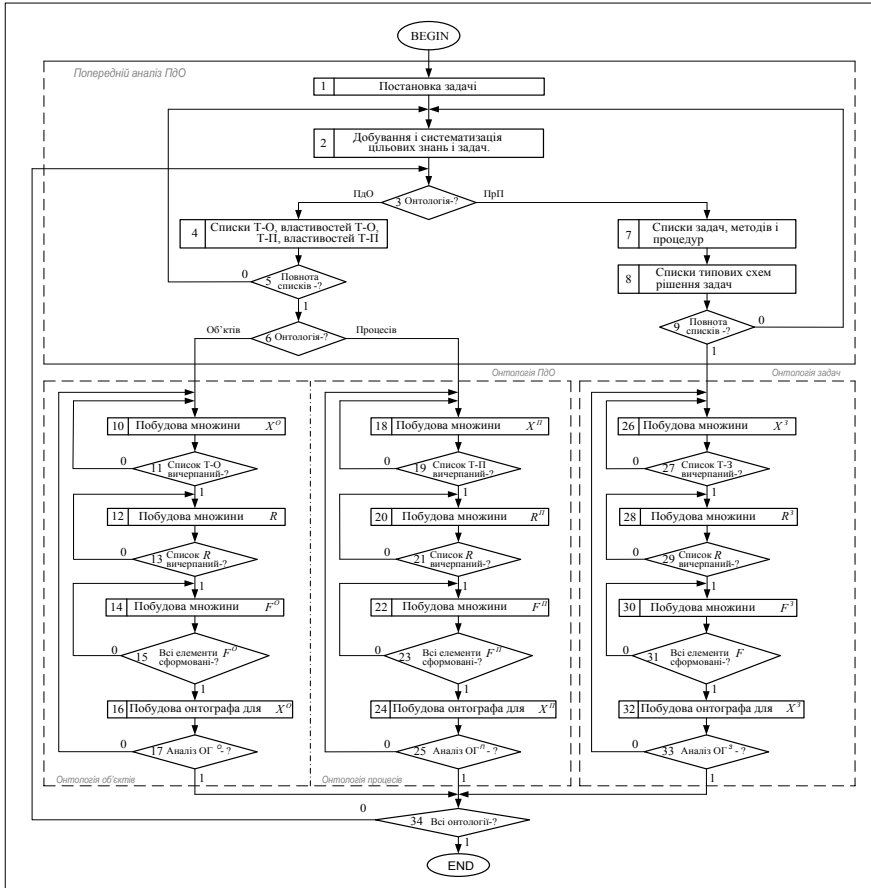
Нижче наведено опис алгоритму проектування онтології об'єктів, онтології процесів і онтології задач з посиланнями на пункти (номери блоків) на рисунку 1.15.

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.

У результаті формуються списки Т-О, Т-П і їх властивостей з урахуванням синонімії списки задач, процедур і методів їх вирішення. Варто 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 (O, P, Z): 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, і, відповідно до відношення «вище – нижче», аналізуються і встановлюються взаємозв'язки $x_i^O R_k^O x_j^O$, $x_i^P R_k^P x_j^P$ і $x_i^Z R_k^Z x_j^Z$ відповідно і будуються попередні онтографи для X^O, X^P і X^Z . При побудові множини F приписуються властивості для кожного X^O, X^P 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 , $\{f_i^P\}$, для кожного X^P 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^P і OG^T за елементами множин понять X^O , X^P і X^T відповідно; аналізуються на повноту і несуперечність онтографи OG^O , OG^P , OG^T 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^0 , O^I і O^3 . 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.

Безсумнівну 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 декілька тисяч записів. Складання вручну таких списків представляється проблемним. У Розділі 2 описується інструментальний комплекс, програмні модулі якого реалізують, зокрема, представлений на рисунку 1.15 алгоритм.

Приклад створення онтології на основі семантичного аналізу документів та формування онтографів наведено у додатку А.

РОЗДІЛ 2

Інструкція для користувача програмою TextTermin: автоматичне виділення термінів на основі лінгвістичного аналізу

(Складено за текстом книги
Наталії Полонської-Василенко «Історія України»)

КРОК 1. Стисла характеристика програми

Програма **TextTermin** призначена для семантичного (змістовного) аналізу тексту. У результаті автоматичного виділення (розпізнавання) термінів на основі лінгвістичного аналізу можна:

- визначити частотні характеристики термінів (скільки разів цей термін зустрічається в тексті);
- виконати повнотекстовий пошук і вибрати необхідні терміни й фрагменти тексту (конспекти), у яких ці терміни використовуються;
- розпізнати терміни як частини мови (іменник, прикметник тощо);
- проаналізувати порядок слів у виділених словосполученнях;
- виконати пошук словосполучень з урахуванням схеми їхньої побудови із заданими видами зв'язку (наприклад: «українська мова» – «прикметник – іменник»);
- сформувати список термінів за заданими параметрами для подальшого аналізу;
- побудувати «дерево» відносин між термінами («мова» – «українська мова» – «вивчення української мови»).

Можливості програми дозволяють використати її як аналітичний інструментарій для виконання науково-дослідницької роботи з україн-

ської мови й літератури.

Нижче наведено технологію роботи з програмою у вигляді послідовних кроків.

КРОК 2. Підготовка до роботи

2.1. Вимоги до робочого місця (комп'ютера)

Існують такі основні вимоги до робочого місця (комп'ютера):

- комп'ютер: особливих вимог немає – сучасний домашній або офісний комп'ютер;
- операційна система: Windows XP, Windows 7 з установленим MS Office.

2.2. Вимоги до тексту

Програма працює з текстом у форматі «Блокнот» (текстовий формат з розширенням **.txt**). Якщо текст для аналізу створений в іншому форматі (наприклад, **.doc** або **.rtf**), його необхідно конвертувати у текстовий формат з розширенням **.txt** і кодуванням **ANSI**. Приклад тексту у необхідному для роботи форматі (фрагмент) наведено на рисунку 2.1.

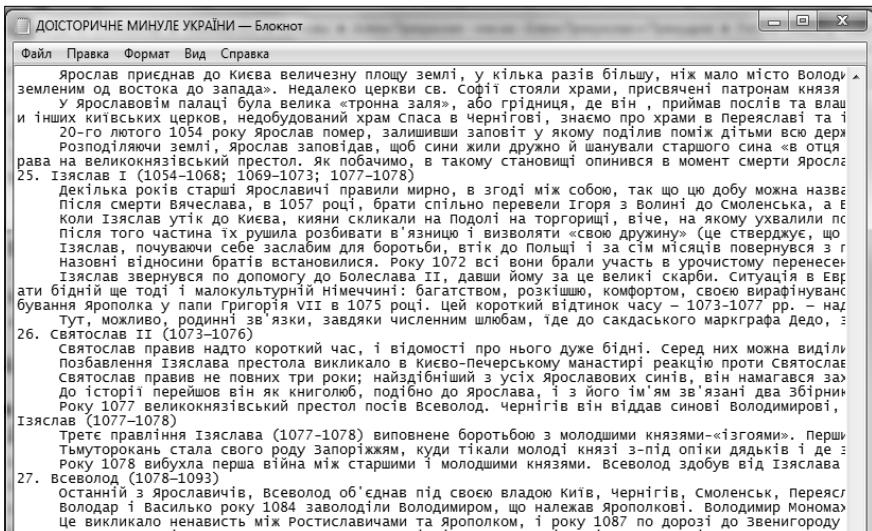


Рис. 2.1. Документ із розширенням .txt (Блокнот)

2.3. Налаштування

Увага! Усі налаштування (рос. «настройки») виконують відповідно до вашої операційної системи.

Перед установкою програми TextTermin необхідно виконати такі налаштування (для Windows XP, рис. 2.2):

- відкрити «Провідник»;
- у рядку меню вибрати «Відкрити закладку “Сервіс”»;
- відкрити «Властивості папки»;
- у вікні, що відкрилося, відкрити закладку «Вид»;
- у вікні «Додаткові параметри» знайти опцію «Приховувати розширення для зареєстрованих типів файлів» і зняти галочку! (рис. 2.2);
- клікнути кнопку «Застосувати до всіх папок», «Застосувати» й «ОК».

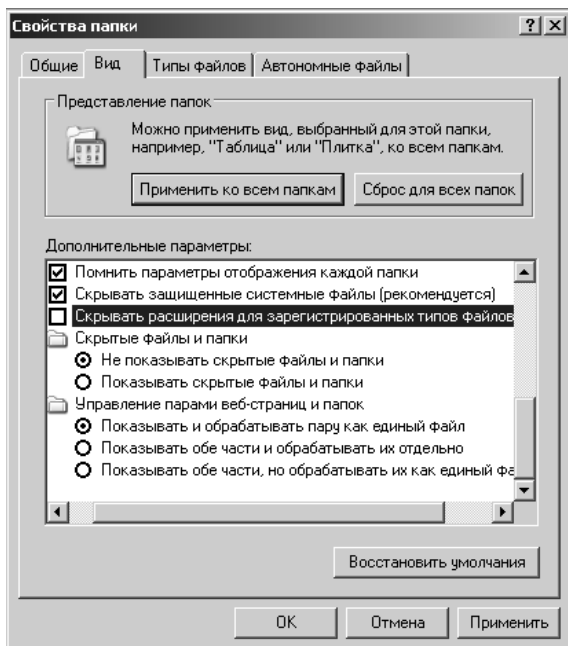


Рис. 2.2. Налаштування параметрів для Windows XP

В ОС Windows 7 перед установкою програми TextTermin необхідно зробити такі налаштування:

- клікнути кнопку «Пуск»;
- у відкритому вікні обрати кнопку «Панель управління» та її натиснути;
- у меню «Налаштування параметрів комп'ютера» обрати параметр «Оформлення та персоналізація» (рис. 2.3).

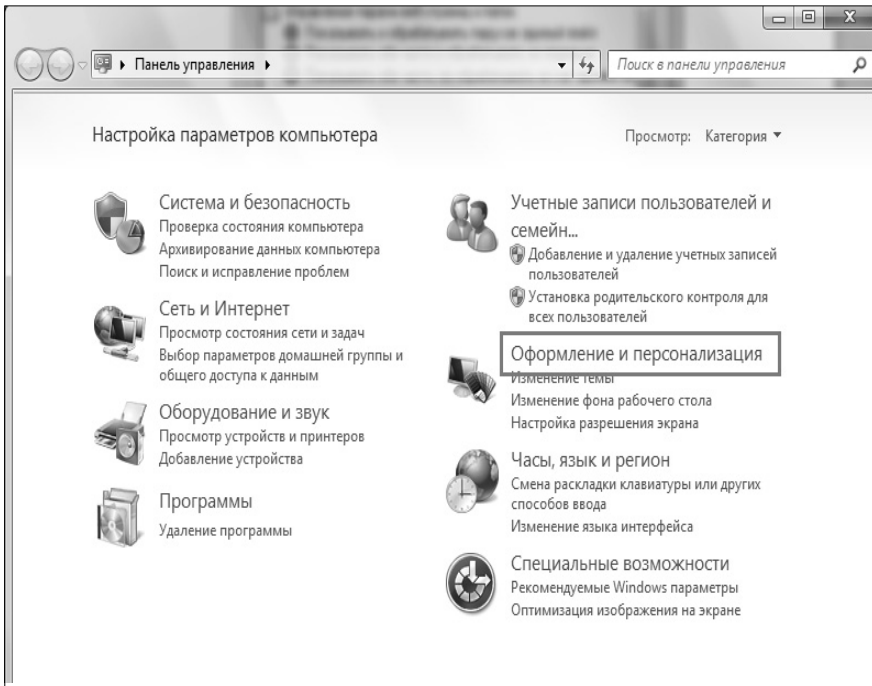


Рис. 2.3. Налаштування параметрів комп'ютера

- обрати рядок «Параметри папок» та у вікні, що з'явилося, обрати закладку «Вид», а потім у додаткових параметрах знайти рядок «Сховати розширення для зареєстрованих типів файлів» та зняти галочку, яка там стоїть; після цього натиснути кнопку «Застосувати», потім – «ОК» (рис. 2.4).

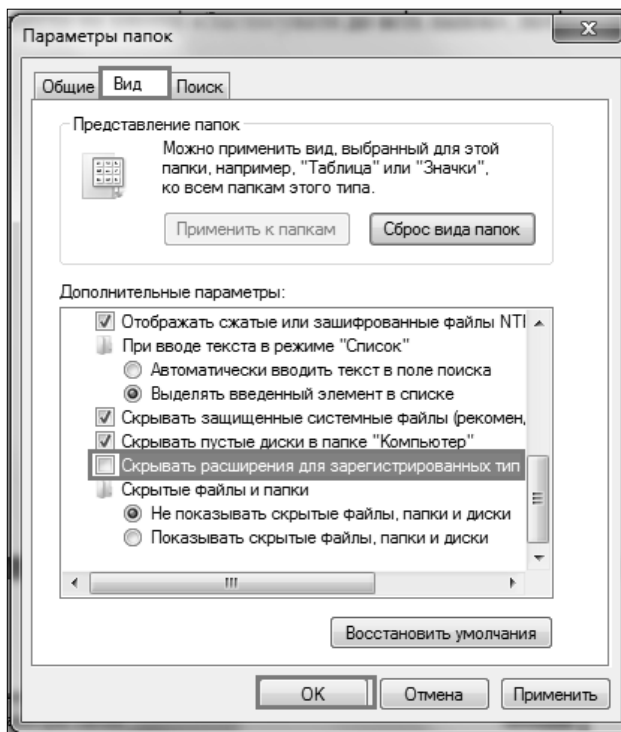


Рис. 2.4. Налаштування параметрів для Windows 7

Отже, Ваш комп'ютер готовий до установки програми TextTermin!

2.4. Установлення програми

Файл із програмою має бути заархівовано. За допомогою програми WinRAR (WinZIP, 7z або інших архіваторів) необхідно розархівувати програму в обрану вами робочу папку.

Далі необхідно відкрити папку TextTermin, створити ярлик файлу запуску програми й відправити його на робочий стіл (рис. 2.5).

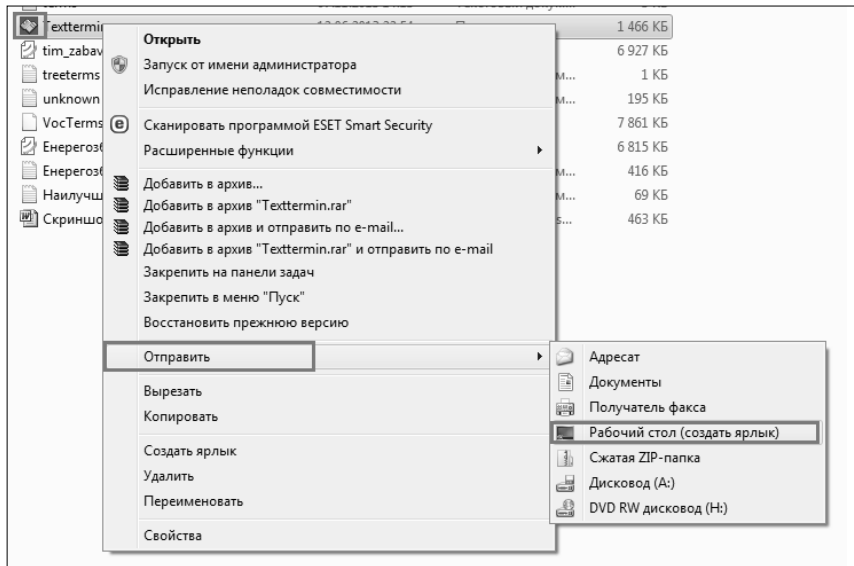


Рис. 2.5. Створення ярлика

Після запуску програми на вашому екрані з'явиться інтерфейс завантаження файлів (рис. 2.6).

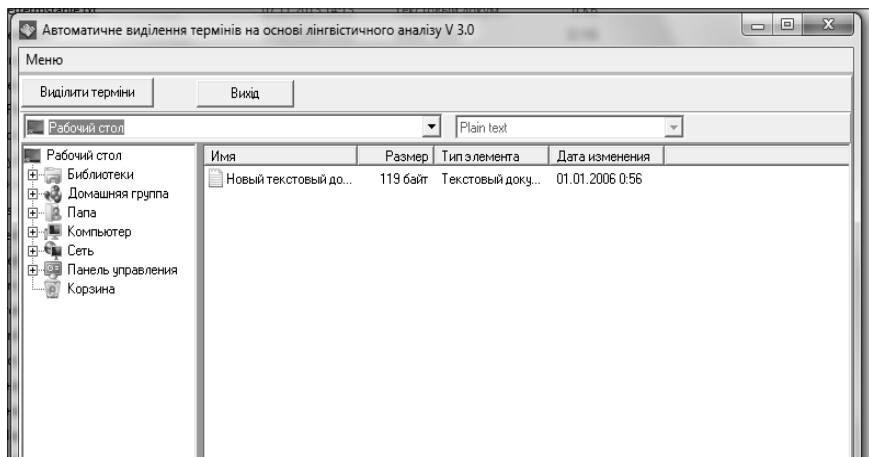


Рис. 2.6. Завантаження файлів

Увага! Переконайтеся, що необхідні для роботи файли збережені (перетворені) у форматі «Блокнот» з розширенням .txt та кодуванням ANSI!

Для завантаження необхідного текстового документа, використовуючи «Провідник», оберіть необхідний документ та виділіть його (рис. 2.7).

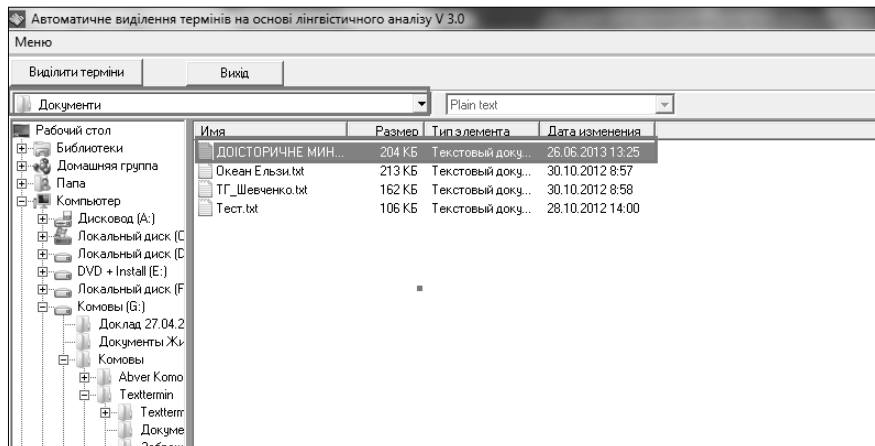


Рис. 2.7. Пошук необхідного документа

Після цього необхідно натиснути кнопку «Виділити терміни» (рис. 2.8).

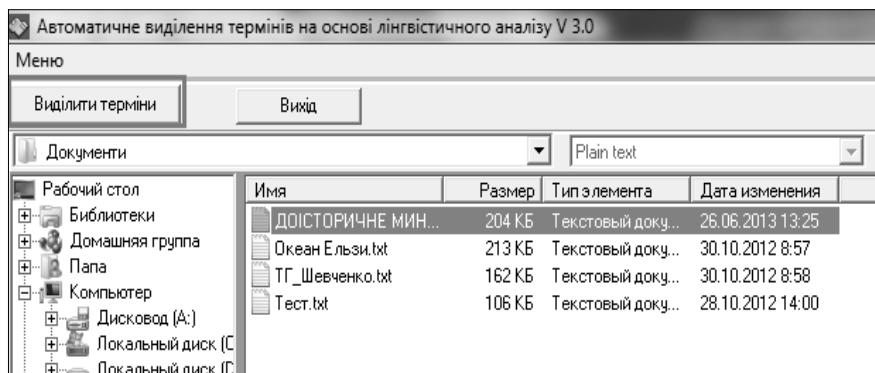


Рис. 2.8. Виділення термінів

Результати аналізу обраного вами файлу з'являться в основному вікні програми «**Terms' Extraction**».

Увага! При натисканні кнопки основне вікно «Terms' Extraction» буде закрито з результатами аналізу. Але ви зможете знову обрати файл для дослідження й продовжити роботу – дані не будуть загублені.

КРОК 3. Автоматичне виділення термінів на основі лінгвістичного аналізу

2.5. Опис основного вікна «Terms' Extraction»

У цьому вікні здійснюються всі основні операції з аналізу тексту (рис. 2.9).

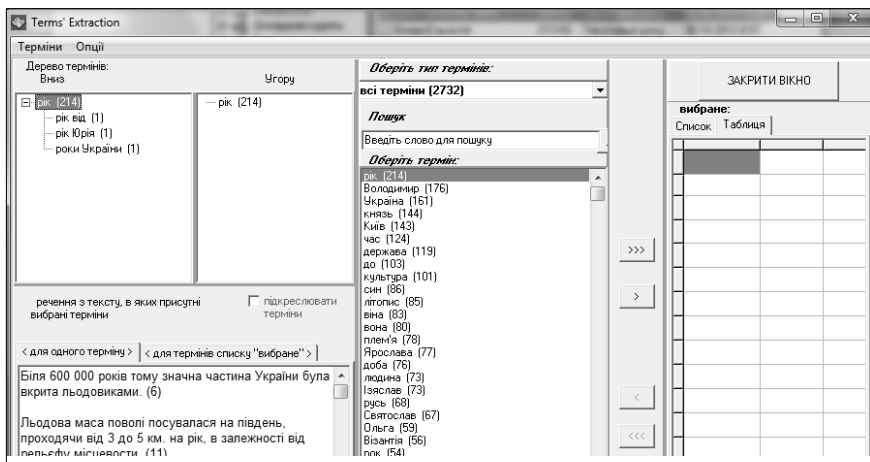


Рис. 2.9. Вікно «Terms' Extraction»

Вікно «Terms' Extraction» можна розділити на чотири основні зони пошуку й дослідження термінів, словосполучень й елементів тексту.

2.6. Опис вікна «Дерево термінів»

У цьому вікні відображаються обрані терміни й словосполучення з ними. Окрім цього, можна здійснювати навігацію (перегляд) слова, що цікавило, або словосполучень у «конспекті», рухаючись по «Дереву

термінів». У разі, коли потрібно швидко перейти до початку «конспекту» потрібно натиснути у вікні на слово «Угору» (рис. 2.10).

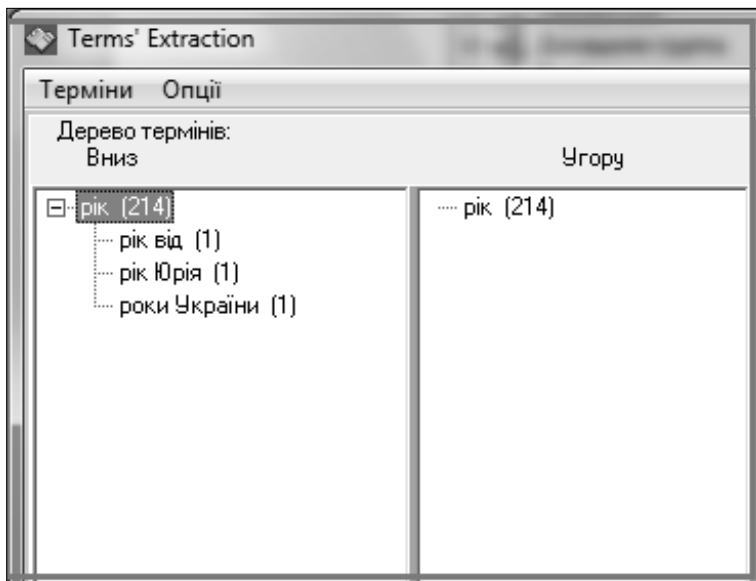


Рис. 2.10. Вікно «Дерево термінів»

2.7. Опис вікна «Речення з тексту, у яких присутні вибрані терміни»

Закладки «Для одного терміну» й «Для термінів списку “вибране”» відображають конспекти (речення з тексту, у яких присутні виділені терміни).

Вікно «Для одного терміну» потрібно використовувати для роботи з одним терміном, що вибирається зі списку у вікні з частотними характеристиками термінів «Оберіть термін».

Вікно «Для термінів списку “вибране”» варто використовувати в разі, коли необхідно проаналізувати один або декілька термінів одночасно.

Якщо термін необхідно виділити у тексті, потрібно поставити галочку у рядку «Підкреслювати терміни» (див. рис. 2.11).

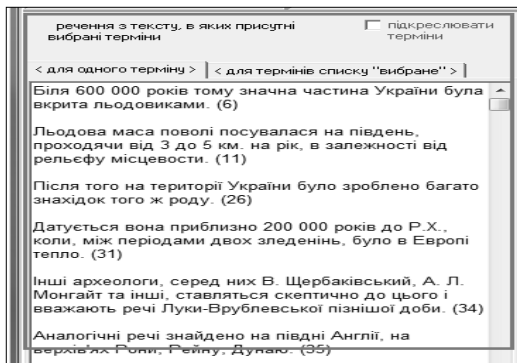


Рис. 2.11. Вікно «Речення з тексту, в яких присутні вибрані терміни»

2.8. Список термінів

Список термінів відображає терміни та частоту їхнього згадування у тексті (рис. 2.12).

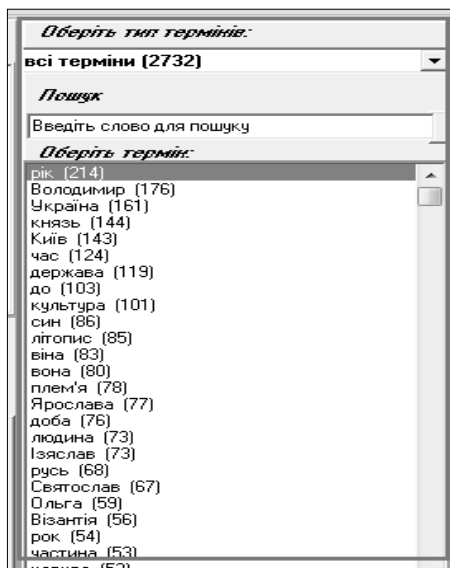


Рис. 2.12. Список термінів

КРОК 4. Робота з матеріалом

2.9. Визначення типу термінів

Опція «Оберіть тип термінів» (рис. 2.13) дозволяє обрати варіанти вживання термінів у тексті (як частин мови, так й словосполучень), які необхідно відібрати й відобразити у вікні «Оберіть терміни».

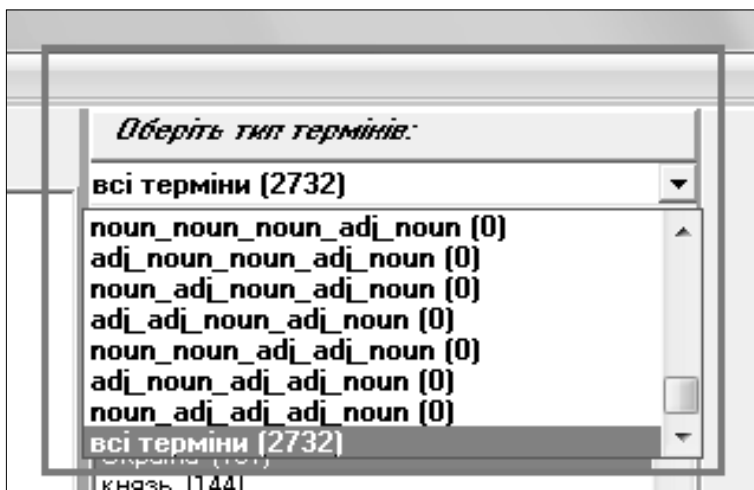


Рис. 2.13. Вікно «Тип термінів»

Відбір термінів та словосполучень здійснюється залежно від варіантів їх використання в тексті. Наприклад:

- іменник – **noun**;
- іменник + іменник – **noun_noun**;
- прикметник + іменник – **adj_noun**;
- прикметник + прикметник + іменник – **adj_adj_noun** і т. д.

Результатом вибірки буде список словосполучень із заданими параметрами.

У вікні «Оберіть терміни» наведено список відібраних словосполучень, з якими можна продовжити роботу.

У наведеному прикладі кількість словосполучень **adj_noun** (прикметник + іменник) дорівнює 264 (рис. 2.14).

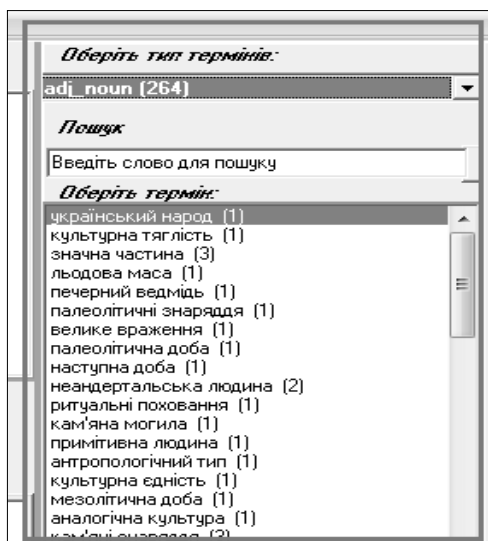


Рис. 2.14. Список словосполучень з відношенням «прикметник + іменник»

Якщо мета виконання завдання полягає у тому, щоб зробити повнотекстовий аналіз, у діалоговому вікні «Оберіть тип термінів» необхідно вибрати варіант «Всі терміни» (у дужках – кількість усіх розпізнаваних термінів). Приклад наведено на рисунку 2.15.

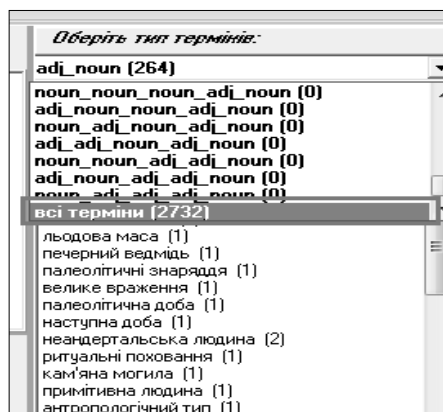


Рис. 2.15. Вибір термінів

2.10. Пошук

Далі відбувається робота у діалоговому вікні «Пошук». У поле «Введіть слово для пошуку» введіть слова, які необхідно знайти у списку.

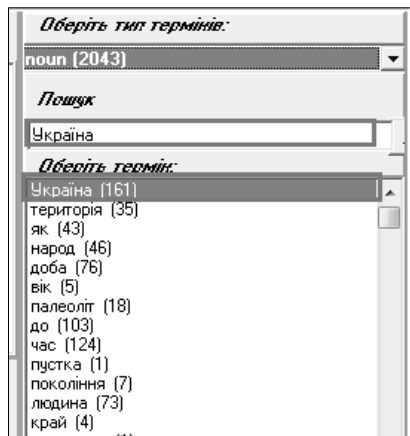


Рис. 2.16. Пошук у тексті

Для прикладу проведемо пошук слова «Україна». Якщо воно наявне в тексті, то буде позначено в списку слів у вікні «Оберіть термін». У дужках навпроти обраного слова буде вказана його частотна характеристика, тобто скільки разів це слово зустрічається в тексті (у наведеному прикладі слово «Україна» зустрічається 161 раз), що відображено на рисунку 2.16.

Робота з діалоговим вікном «Оберіть термін» полягає у наступному. На початку роботи (під час вибору файлу для аналізу й натискання кнопки «Виділити терміни») вікно «Оберіть термін» показує результати аналізу всього тексту. У вікні «Оберіть тип термінів» відображається загальна кількість термінів (у наведеному прикладі – «всі терміни (2732)»), що відображено на рисунку 2.17.

Також у цьому вікні відображаються результати вибору типу термінів або словосполучень із ними, і результати пошуку залежно від заданих параметрів.

Основне завдання, що вирішується за допомогою цієї опції – це вибір (виділення) терміна або групи термінів для подальшої побудови дерева термінів.

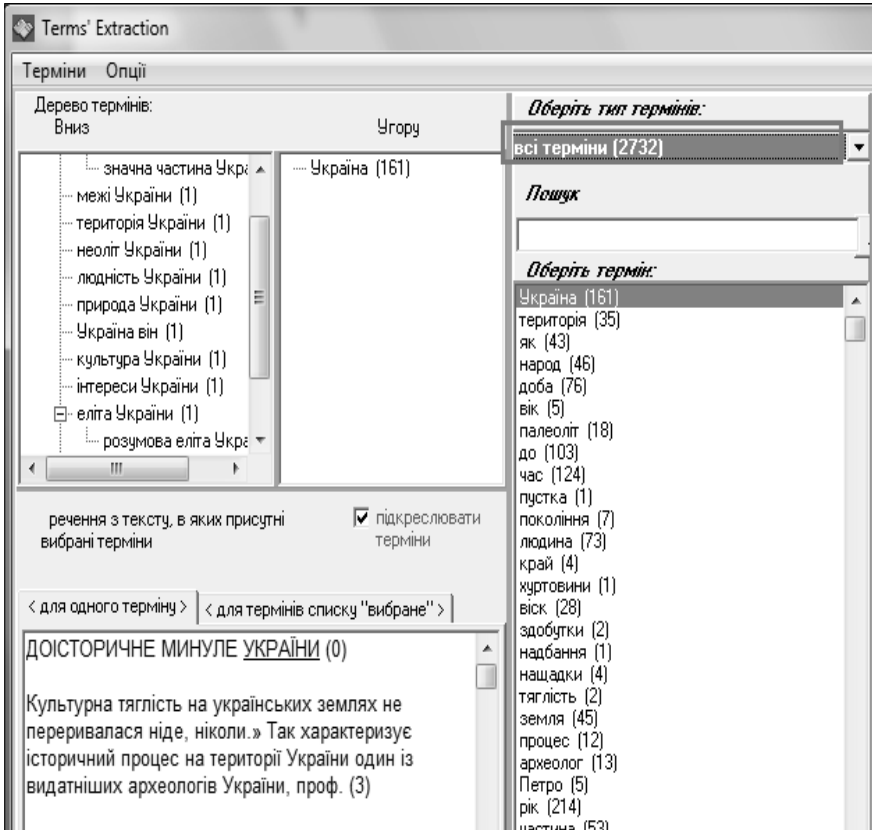


Рис. 2.17. Опція «Всі терміни»

Приклад. Робота з усім списком термінів (рис. 2.18). Для початку аналізу візьмемо слово «палеоліт» і виберемо його в списку термінів. У вікнах «Дерево термінів» й «Речення з тексту, у яких присутні вибрані терміни» одразу відображається, у яких словосполученнях / реченнях використовується термін / слово «палеоліт».

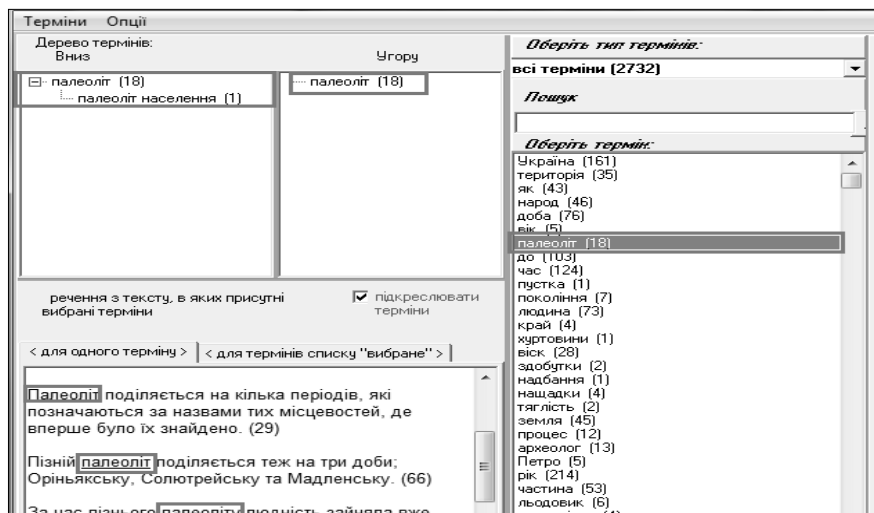


Рис. 2.18. Побудова дерева термінів за конкретним терміном (словом)

КРОК 5. Збереження даних

2.11. Збереження даних

У разі, якщо потрібно зберегти дані після лінгвістичного аналізу тексту, то список термінів необхідно перенести у вікно «Вибране» і натиснути кнопку «Зберегти». TextTermin пропонує зберегти додаткову інформацію про текст, що був проаналізований. У результаті аналізу система формує два файли з розширенням .txt: (terms.txt; unknown.txt) і автоматично один файл із розширенням .xml (allterms.xml). У файлі terms.txt збережено список усіх термінів та їхніх статистичних характеристик.

У файлі unknown.txt збережено список слів, не розпізнаних системою (відсутніх в завантажених словниках).

У файлі allterms.xml збережено список усіх термінів зв'язків між ними, їхніх статистичних характеристик і позицій, де зустрічаються виділені терміни у форматі .xml.

КРОК 6. Розділ «Побудова таблиці»

2.12. Перенесення термінів у таблицю

Для перенесення термінів у таблицю потрібно виконати наступні дії. Необхідний термін, що знаходиться у вікні «Дерево термінів», потрібно скопіювати та перенести в таблицю.

Увага! Перший термін має зайняти дві клітинки поруч у таблиці. За цим принципом потрібно переносити усі термини.

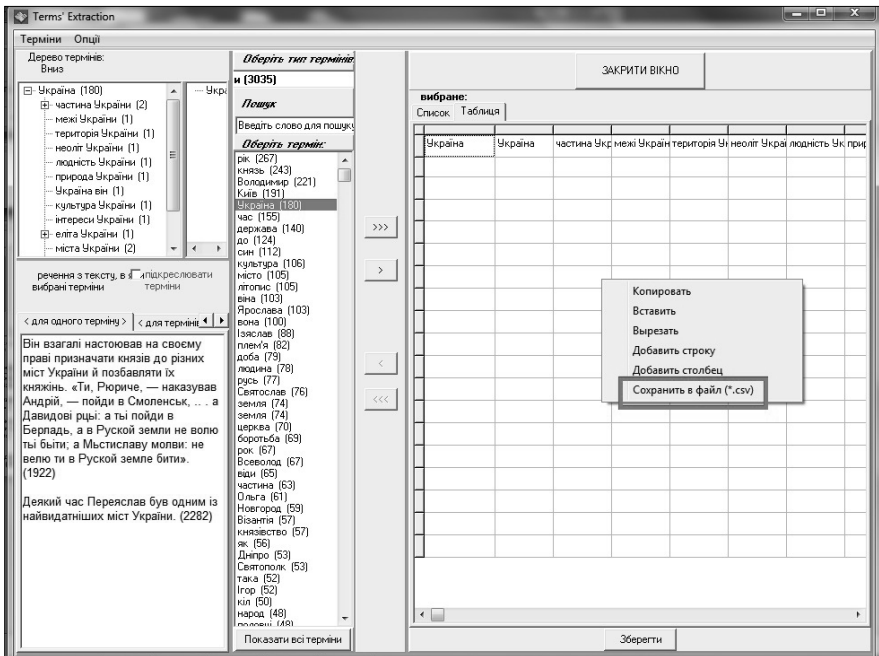


Рис. 2.19. Перенесення термінів у таблицю

Увага! Список термінів може бути відредагований.

Для подальшої роботи дані із таблиці, що була побудована у програмі Texttermin, необхідно перенести у таблицю Excel.

Спочатку потрібно відкрити таблицю Excel, а у таблиці відкрити закладку «Дані», та натиснути кнопку «Текст по стовпчиках» (рис. 2.20).

РОЗДІЛ 2. ІНСТРУКЦІЯ ДЛЯ КОРИСТУВАЧА ПРОГРАМОЮ TEXTTERMIN: АВТОМАТИЧНЕ ВИДІЛЕННЯ ТЕРМІНІВ НА ОСНОВІ ЛІНГВІСТИЧНОГО АНАЛІЗУ

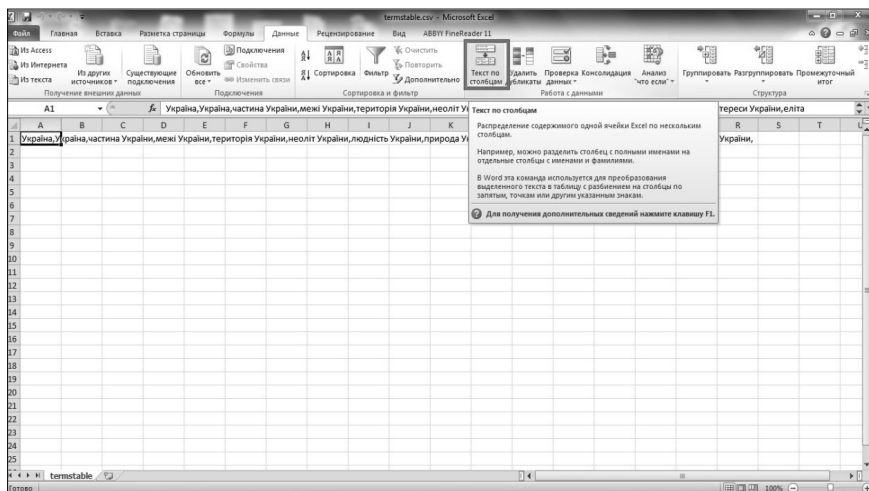


Рис. 2.20. Работа у таблиці Excel

Після успішного виконання цих дій потрібно перенести дані в таблицю. Через особливості програми Texttermin терміни переносяться одним масивом, тому необхідно розподілити кожен термін в окрему клітинку, використовуючи інструменти таблиці Excel (рис. 2.21).

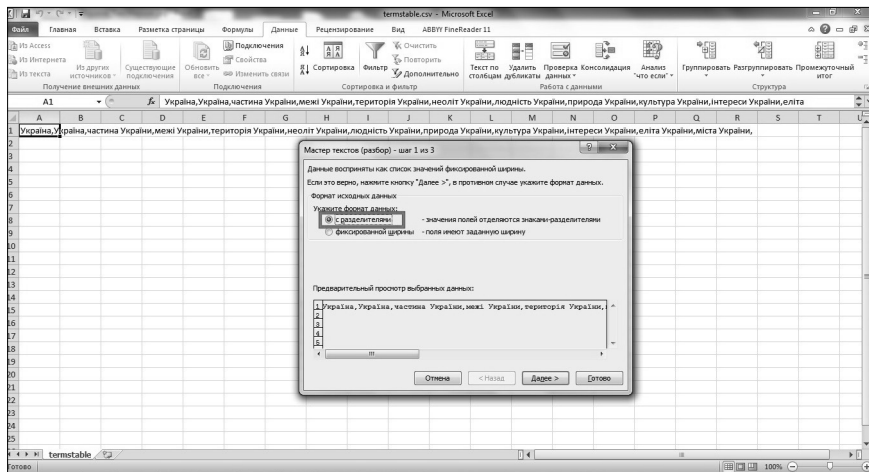


Рис. 2.21. Розподілення термінів за окремими клітинками (крок 1)

Використовуючи інструкції, наведені у вікні «Майстер текстів (розбір)», потрібно покроково розподілити терміни (рис. 2.22).

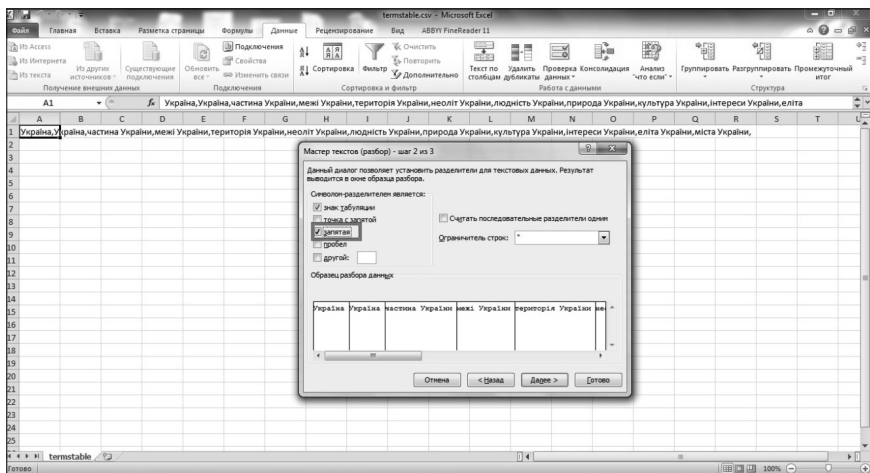


Рис. 2.22. Розподілення термінів за окремими клітинками (крок 2)

Останнім кроком на цьому етапі є форматування даних у таблиці (рис. 2.23).

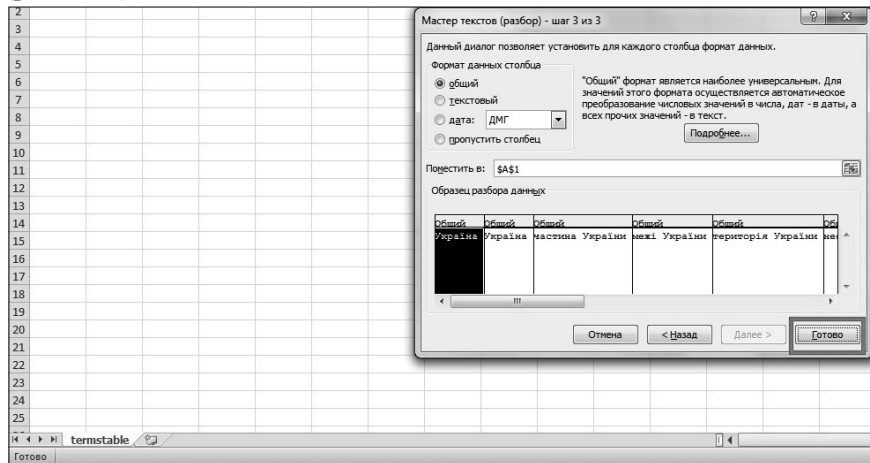


Рис. 2.23. Розподілення термінів за окремими клітинками (крок 3)

2.13. Збереження таблиці

Заповнену та відформатовану таблицю необхідно зберегти для подальшої роботи з нею. Для цього у закладці «Файл» потрібно натиснути кнопку «Зберегти як» (рис. 2.24).

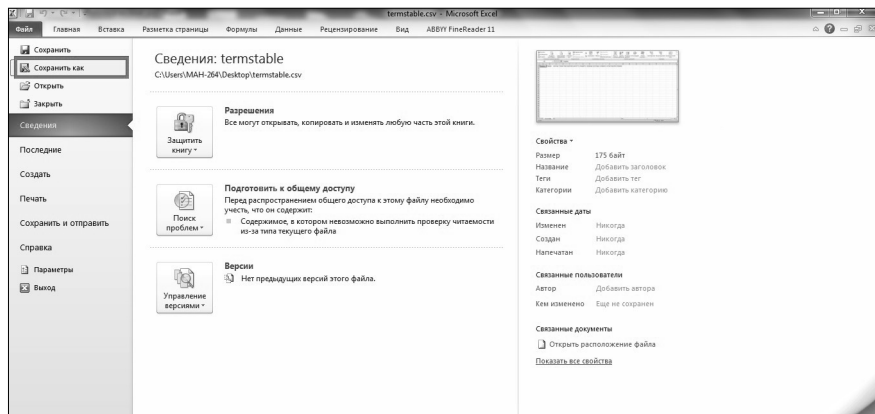


Рис. 2.24. Збереження таблиці

У вікні «Збереження документа» необхідно вибрати «Ім'я файлу» – Книга файлу Excel 97-2003 (* .xls) та тип файлу – CSV (роздільники – коми) (* .csv), а потім натиснути кнопку «Зберегти» (рис. 2.25).

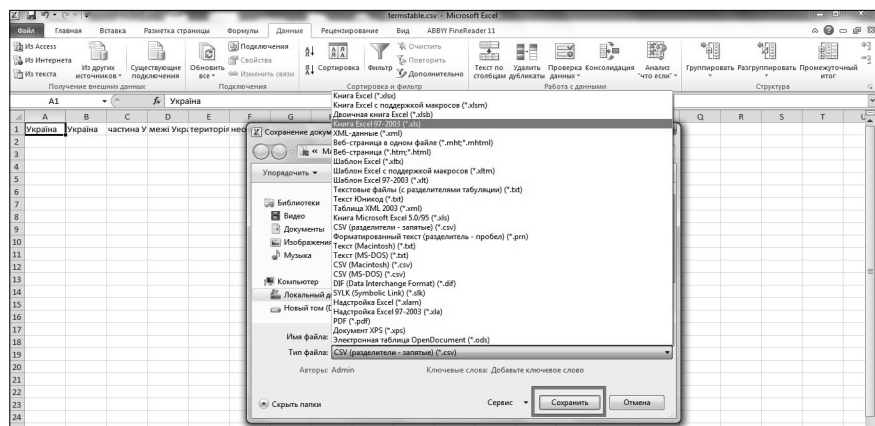


Рис. 2.25. Збереження таблиці Excel

2.14. Конвертація даних за допомогою програми Confor

Для успішного виконання конвертації даних за допомогою програми Confor у командному рядку браузера потрібно надрукувати `test1.inhost.com.ua/` (рис. 2.26).



Рис. 2.26. Введення адреси у командний рядок

Далі потрібно обрати мову, натиснувши на кнопку «Українська» (рис. 2.27).

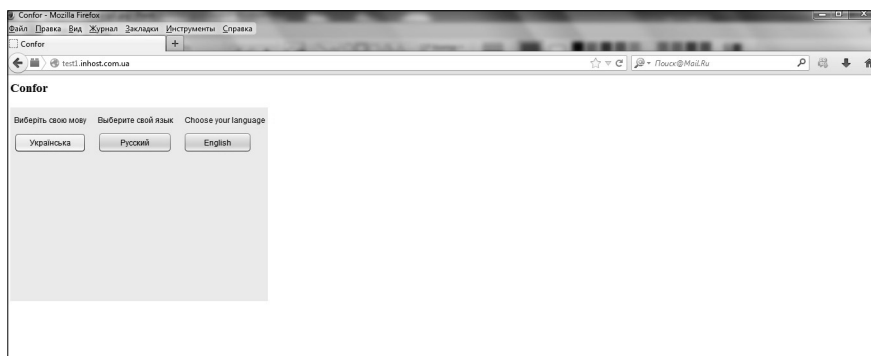


Рис. 2.27. Вибір мови

На наступному етапі потрібно обрати потрібний для роботи файл. Для цього в лівій частині екрану необхідно натиснути кнопку «Відкрити Excel-файл», а потім у вікні, що з'явилось, обрати файл, з яким ви працюєте (рис. 2.28). Файл повинен мати розширення .xls.

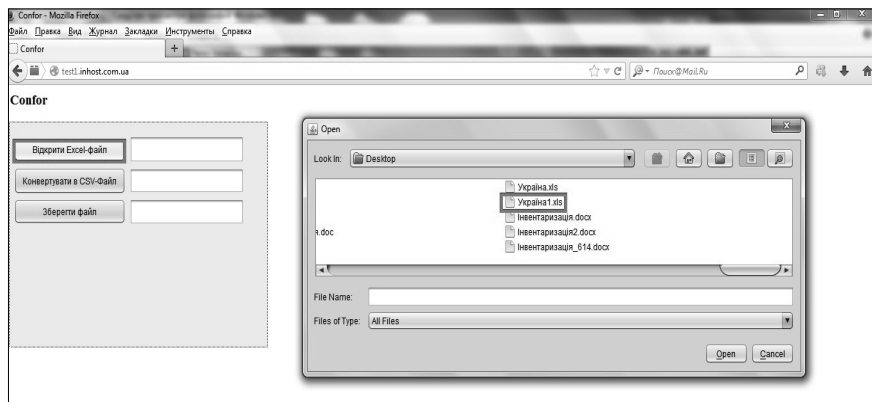


Рис. 2.28. Вибір файлу

Потім потрібно конвертувати цей файл у CSV-Файл, що можна зробити натиснувши кнопку «Конвертувати в CSV-Файл» (рис. 2.29).

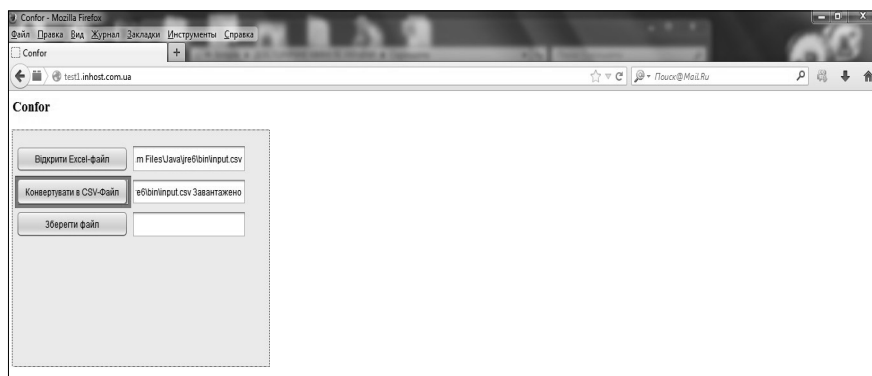


Рис. 2.29. Конвертація у CSV-Файл

Після усіх виконаних попередньо дій потрібно зберегти файл. Для цього натисніть кнопку «Зберегти файл».

РОЗДІЛ 3

Зразки виконання дослідницької роботи з використанням програми TextTermin

(Матеріал дослідження – листи Т. Г. Шевченка
до рідних та друзів)

3.1. Завдання I. За допомогою програми TextTermin побудувати дерево термінів та провести статистичний (кількісний) контент-аналіз: які слова та словосполучення найчастіше використовує поет у своєму листуванні

1. Потрібно відшукати надійне джерело в Інтернеті та скачати необхідні тексти. У нашому випадку – це листи Т. Г. Шевченка різних років до різних осіб (листи до А. І. Лизогуба, Г. Ф. Квітки-Основ'яненка, М. Г. Шевченка; джерело – <http://www.t-shevchenko.name/uk/Corresp/1852-55/1852-07-16ail.html>).

2. Зібравши необхідні тексти та скопіювавши їх в документ Word, потрібно зберегти цей документ у текстовому форматі (з розширенням *.txt).

3. Для подальшої роботи потрібно відкрити програму TextTermin відповідно до Інструкції, наведеної у Розділі 2, та вибрати необхідний файл, а потім натиснути кнопку «Виділити терміни».

4. У результаті автоматичного виділення термінів на основі лінгвістичного аналізу отримуємо «Дерево термінів». Якщо поставити галочку біля рядка «Підкреслювати терміни», вибраний термін буде підкреслено у реченнях з тексту.

5. На основі отриманих даних можна робити висновки щодо частоти використання тих або інших слів та словосполучень у листуванні поета.

3.2. Завдання II. Скориставшись даними програми TextTermin, побудувати онтологічний граф з вершиною «Лист» та прикріпити відповідні дані та гіперпосилання

Для здійснення подальших досліджень необхідно продовжити роботу у програмі TextTermin та заповнити таблицю.

1. Для заповнення таблиці необхідно скопіювати потрібні терміни та вставити їх у комірки таблиці. Необхідно звернути увагу, що перший термін дублюється у першому та другому стовпчику.

2. Для подальших досліджень потрібне використання отриманих даних. Це може бути статистичний аналіз, контент-аналіз тощо. Один із видів дослідження ґрунтується на побудові онтології. Для цього матеріали таблиці підлягають подальшому опрацюванню.

3. Таблицю, побудовану в програмі TextTermin, необхідно скопіювати у середовище Microsoft Office Excel.

Увага! Під час копіювання необхідно переносити кожну клітинку окремо (рис. 3.1).

	A	B	C	D	E	F	G	H	I	J	K	L
1	Бог	Бог	море	Оренбург	Петербур	Яготин	фортеця	Одеса				
2	степ	кобзар	Урал	нудьга	листи	коментар	брат	Микита				
3												
4												
5												

Рис. 3.1. Таблиця Excel

4. Після заповнення таблиці структури онтології (назвемо її «Шевченко_структура.xls»), конвертуємо її Web-додатком Confortest1.inhost.com.ua. (рис. 3.2).

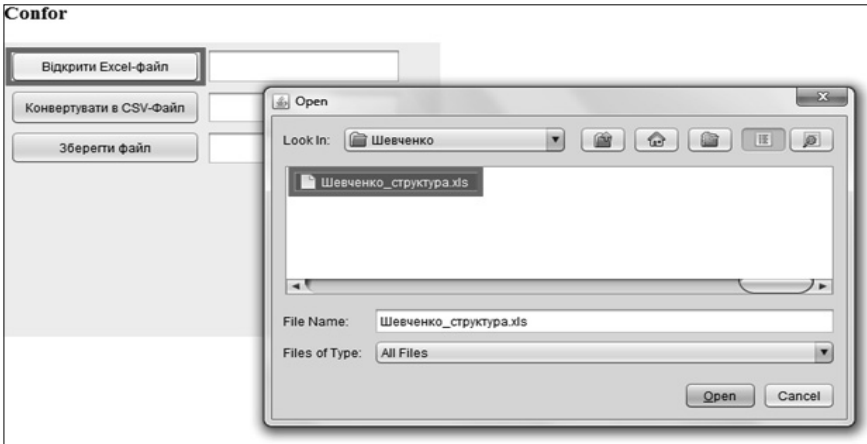


Рис. 3.2. Конвертація даних таблиці (файлу Excel) в CSV-Файл

Таким чином, структура онтології, описана в таблиці MSEXcel «Шевченко_структура.xls», перетворюється на файл, формат якого зчитує Web-додаток GraphEditor.

5. Для редагування та наповнення онтології необхідно відкрити відповідний електронний ресурс за адресою: <http://editor3.inhost.com.ua/> (рис. 3.3).

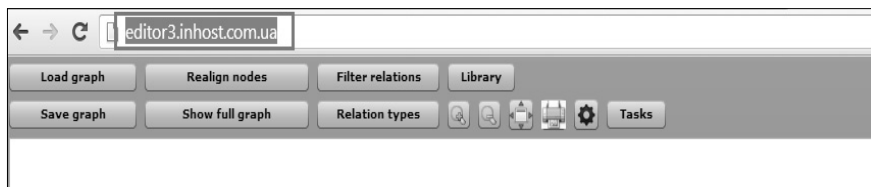


Рис. 3.3. Web-додаток GraphEditor

6. Далі на панелі інструментів потрібно натиснути кнопку Settings (рис. 3.4).



Рис. 3.4. Панель інструментів

7. У вікні, що відкрилося, потрібно вибрати мову (рис. 3.5).

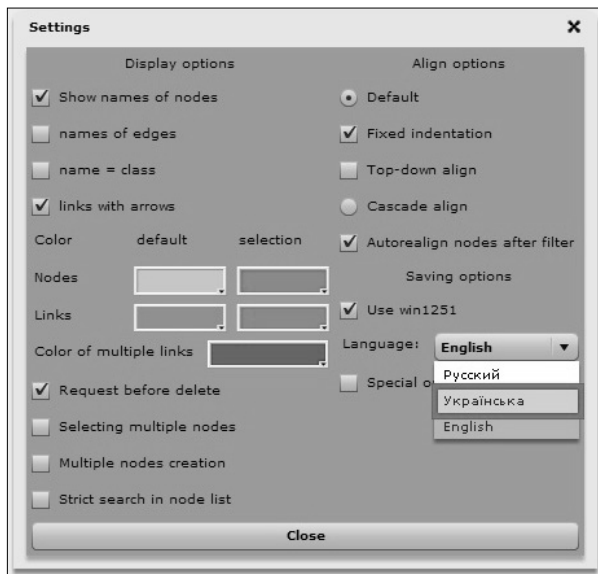


Рис. 3.5. Опції для вибору мови

8. Далі на панелі інструментів необхідно натиснути кнопку «Зчитати граф» (рис. 3.6).

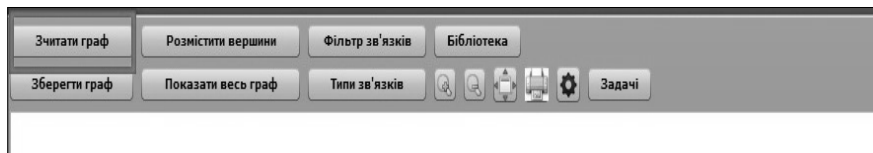


Рис. 3.6. Кнопка «Зчитати граф»

9. У вікні, що відкрилося, обрати сконвертований Confor файл та натиснути кнопку «Відкрити» (рис. 3.7).

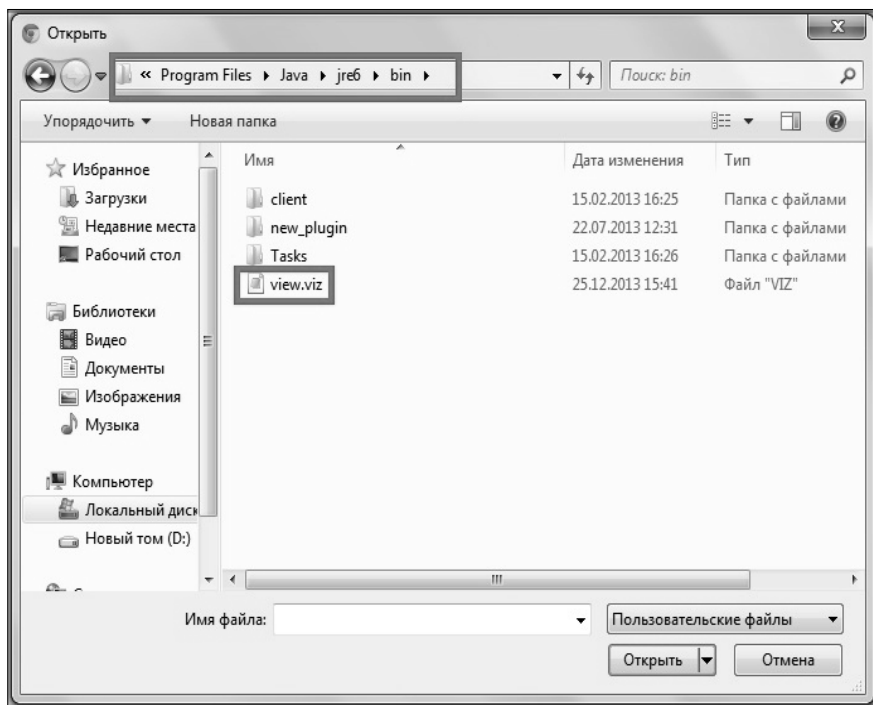


Рис. 3.7. Вибір файлу VIZ

10. В основному вікні з'явиться граф (рис. 3.8).

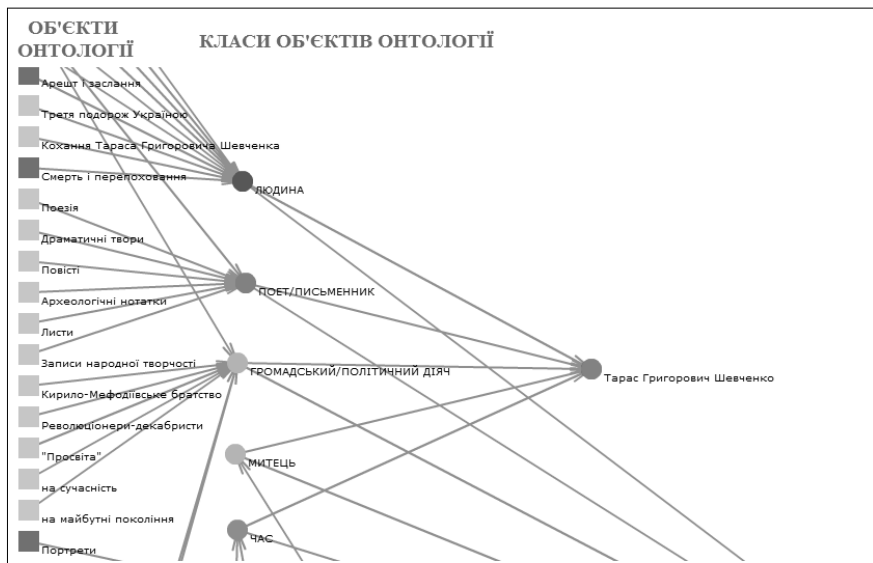


Рис. 3.8. Граф

11. Наведені у графі дані можна редагувати. Для цього необхідно натиснути кнопку «Редагувати» (рис. 3.9).

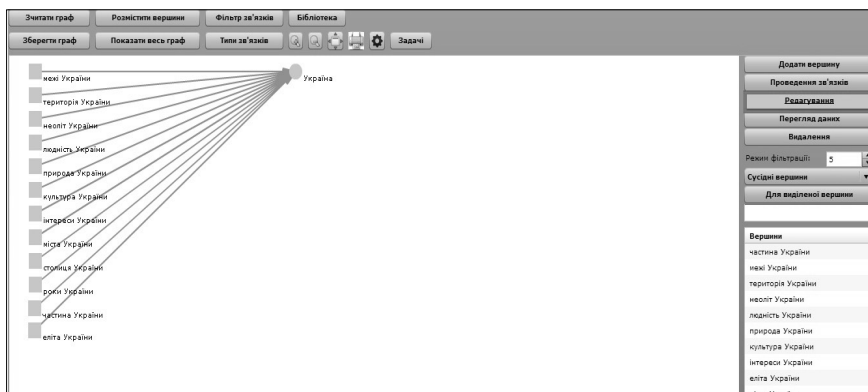


Рис. 3.9. Редагування графу

12. Далі на панелі, що з'явиться, необхідно відредагувати дані, використавши відповідні кнопки та стрілочки (рис. 3.10).

Редагування вершин

Ім'я: Україна

клас:

Форма: **Круг**

колір:

X: 440 Y: 20

Шрифт: Verdana 10

Додати дані (Л)

Редагувати дані

Видалити дані (V)

Прийняти Відмінити

Дані	Посилання	Тип	Клас

Рис. 3.10. Редагування вершин графу

13. Тепер до вершин графу можна прикріпити гіперпосилання на Інтернет-ресурси за допомогою відповідних кнопок (рис. 3.11).

Додати/змінити данс

Карта України онлайн

Посилання: <http://map.meta.ua/ua/#zoom=6&lat=48.5&lon=31.2&base=B00>

Гіперпосилання

Прийняти Відмінити

Рис. 3.11. Додавання / зміна даних; додавання гіперпосилань

14. Також можна додати текстові дані. Для цього необхідно набрати текст та натиснути кнопки «Текст» та «Прийняти» (рис. 3.12).



Рис. 3.12. Додавання текстових даних

15. Усі додані дані відображаються у вікні «Редагування вершини». Після їх перегляду необхідно натиснути кнопку «Прийняти» (рис. 3.13).

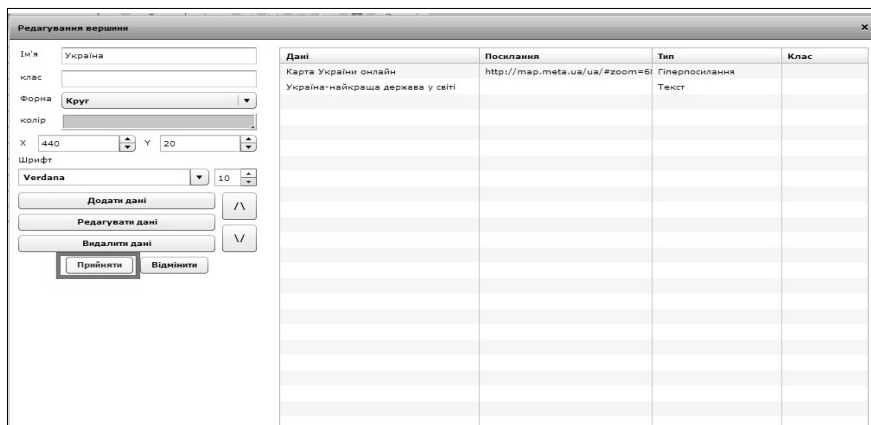


Рис. 3.13. Вікно «Редагування вершини» з доданими даними

16. Усі прикріплені дані можна переглянути. Для цього необхідно закрити попереднє вікно та натиснути кнопку «Перегляд даних» (рис. 3.14).

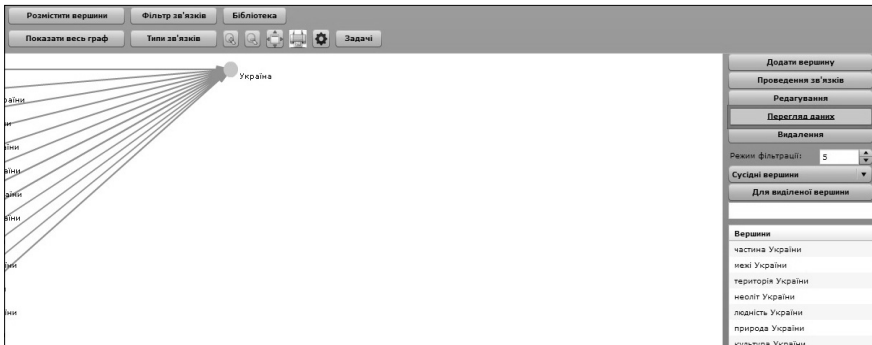


Рис. 3.14. Перегляд даних

17. У вікні, що з'явилось, можна побачити дані щодо гіперпосилань для вершини.

Нижче наведено приклади даних та гіперпосилань до вершини онтології «Листи». Вершина «Листи» містить інформацію про короткий зміст листа Тараса Григоровича до брата Микити та бібліографічні відомості (рис. 3.15).

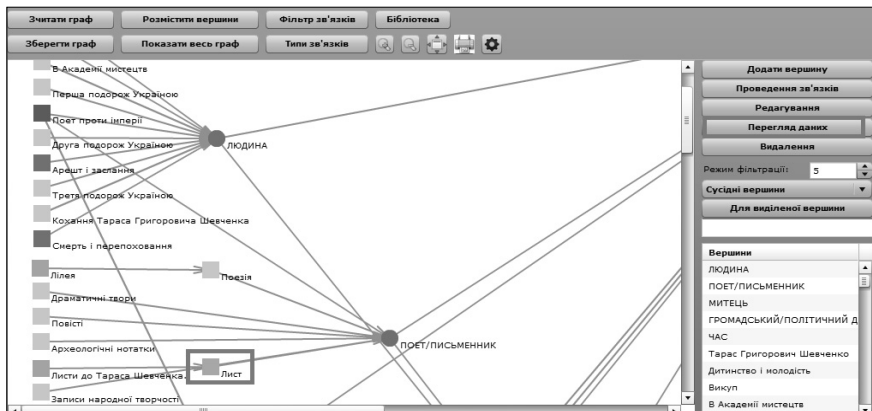


Рис. 3.15. Вершина онтології «Лист»

Якщо натиснути на вершину «Лист», відкривається наступне вікно з переліком даних та інформації щодо листування Т. Г. Шевченка (рис. 3.16).

РОЗДІЛ 3. ЗРАЗКИ ВИКОНАННЯ ДОСЛІДНИЦЬКОЇ РОБОТИ З ВИКОРИСТАННЯМ ПРОГРАМИ TEXTTERMIN

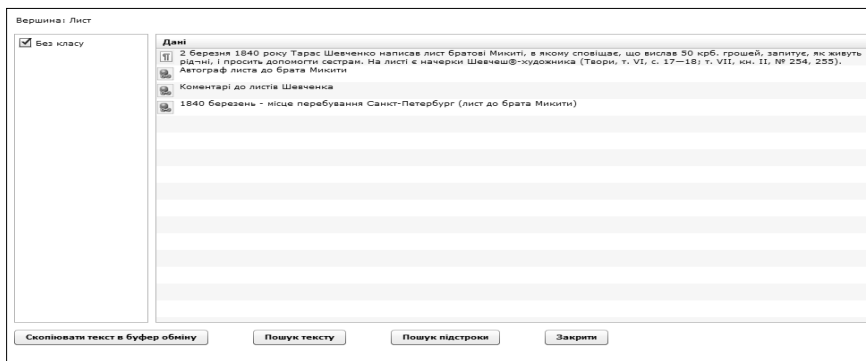


Рис. 3.16. Дані для вершини «Лист»

Після натискання гіперпосилання «Автограф листа до брата Микити» відкривається вікно з зображенням відсканованого оригіналу листа Т. Г. Шевченка до М. Г. Шевченка (рис. 3.17).

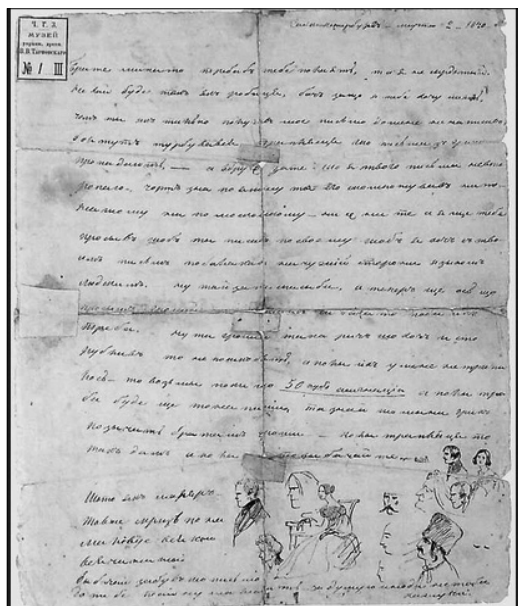


Рис. 3.17. Відсканований оригінал листа

Після натискання гіперпосилання «Коментарі до листів Шевченка» відкривається вікно з відповідними матеріалами (рис. 3.18).

КОМЕНТАРИ

У шостому томі академічного зібрання творів Шевченка вміщено листи, ділові папери, дарчі та власницькі написи, документи, складені Шевченком або за його участю.

Подано листи майже 22-річного періоду: від 15 листопада 1839 р. (дата найранішого відомого листа, адресованого братові Микиті Шевченку) до 24 лютого 1861 р. (дата буквально передсмертного листа І. М. Мокрицькому). Вміщено 237 приватних листів, 15 офіційних листів та ділових паперів, 4 колективних листи, підписані Шевченком. В одному хронологічному ряду з листами до індивідуальних адресатів подано листи до редакцій газет та приписки поета в листах інших осіб. Вперше введено у видання листи адресовані В. М. Репніній (червень 1844 р.), О. І. Бутакову (1850), О. С. Уварову (1859), В. В. Тарновському-старшому (1859), приписку Шевченка в листі П. О. Куліша до О. М. Бодяньського від 10 січня 1847 р., колективний лист представників петербурзької громадськості від 26 грудня

1858 р. В. Ф. Адлербергу. Розширено розділ «Офіційні листи», який нині налічує 15 одиниць. Уперше до академічного видання введено розділи «Дарчі та власницькі написи» (на книжках, малюнках та офортах, автографах поетичних творів та фотографіях), «Документи, складені Т. Шевченком або за його участю».

Тексти подаються за автографами, а в разі їх відсутності — за першодруками. Вперше в академічному виданні за автографами подано 16 листів Шевченка до Бр. Залеського, що дозволило внести в текст суттєві виправлення. За іншим джерелом — публікаціями в журналі «Основа» — подаються листи поета до В. Г.

Рис. 3.18. Коментарі до листів Шевченка

Обравши гіперпосилання «1840 – березень місце перебування Санкт-Петербург лист до брата Микити» ми можемо зробити перехід на інтерактивну мапу (рис. 3.19).

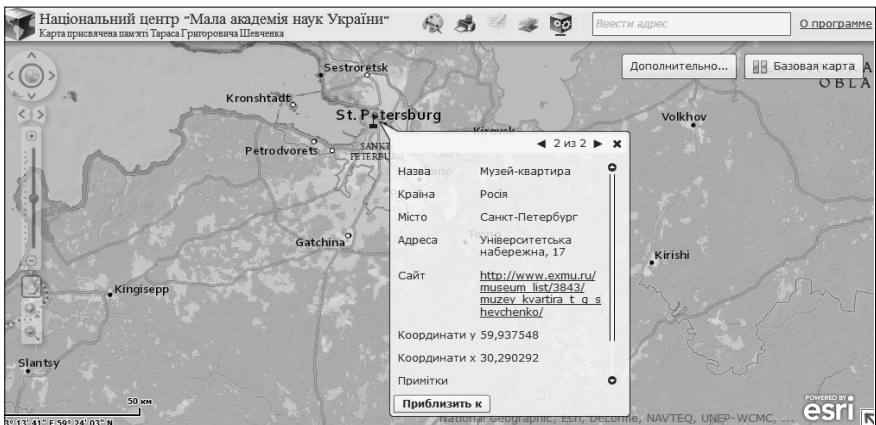


Рис. 3.19. Інтерактивна мапа, присвячена пам'яті Т. Г. Шевченка, на якій зазначено місце перебування поета під час написання листа

Також завдяки вбудованій пошуковій машині можна знайти додаткові відомості. Для цього потрібно натиснути «Пошук підстроки», виділити фрагмент тексту, який нас цікавить, натиснути «Пошук тексту».

Результати пошуку відображаються в інтерфейсі пошукової системи як список документів, що ранжуються залежно від релевантності запиту. Для перегляду контенту потрібно вибрати відповідний документ та натиснути на нього, завдяки чому відкриється нове вікно. У новому вікні відкриється весь текст документу (рис. 3.20), в якому різними кольорами виділено пошукові слова.

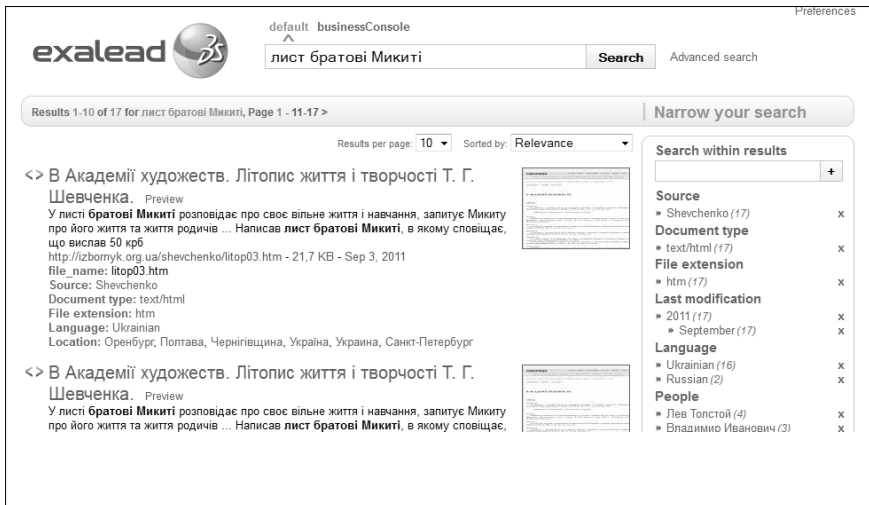


Рис. 3.20. Результати пошуку за пошуковим текстом «лист братові Микиті»

РОЗДІЛ 4

Технологія створення онтологічних моделей та профільних баз знань

(Обробка текстів фахової літератури для студентів педагогічних спеціальностей університетів)

За роки навчання у більшості студентів (майбутніх учителів) поповнюється активний словник, формуються навички роботи з інформаційними джерелами (наприклад, зі словниками та довідниками). У цілому студенти оволодівають фаховою та загальнонауковою термінологією. Проте недоліком професійної мовної підготовки можна назвати відсутність логічних смислових зв'язків між компонентами терміносистеми: слова-терміни зазвичай використовуються неусвідомлено, їхні дефініції «зазубрюються». Майбутні вчителі можуть взагалі уникати термінів, а про слова-дублети забувають взагалі. З огляду на це, створення онтологічних моделей і профільних БЗ є оптимальним способом подолання зазначених суперечностей.

У результаті роботи програми TextTermin побудовано і збережено таблиці термінів, кожна з яких є основою для створення «дерева термінів». При цьому всі слова «активізовані», тобто містять відповідний контекст-супровід (словникові дефініції, уривок тексту посібника, підручника, схеми тощо). Ці терміни систематизовані як лінгвістичні, математичні, хімічні, психологічні, а тому відповідно призначені для формування систематизованих знань з теорії того чи іншого фаху. Отримані під час обробки текстів спеціальної літератури для студентів (майбутніх учителів-предметників) матеріали оптимізують процес формування і розвитку професійної компетентності студента, педагога-

фахівця, а також допомагають учням зорієнтуватися у проблемі науково-дослідницької роботи з певної галузі знань.

4.1. Лінгвістична термінологія

Климова К. Я., Романюк П. Ф. Практична стилістика сучасної української мови: навчальний посібник для студентів ВНЗ III–IV рівнів акредитації та викладачів-словесників закладів освіти. – Житомир: Вид-во ЖДУ, 2005. – 144 с.

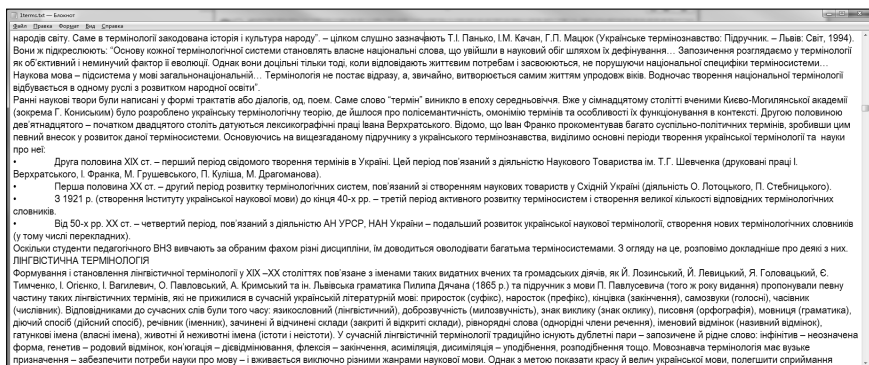


Рис. 4.1. Текст джерела 1

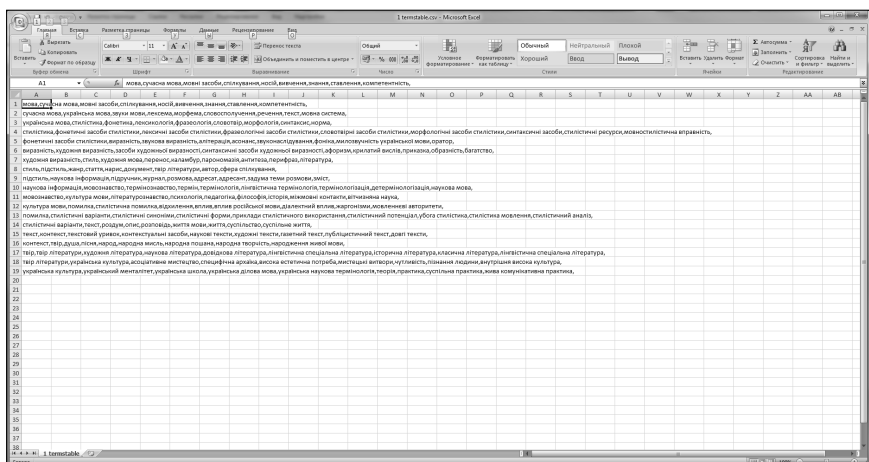


Рис. 4.2. Таблиця лінгвістичних термінів

4.4. Психологічна термінологія

Максименко С. Д., Соловієнко В. О. Загальна психологія: навчальний посібник. – К.: МАУП, 2000. – 256 с.

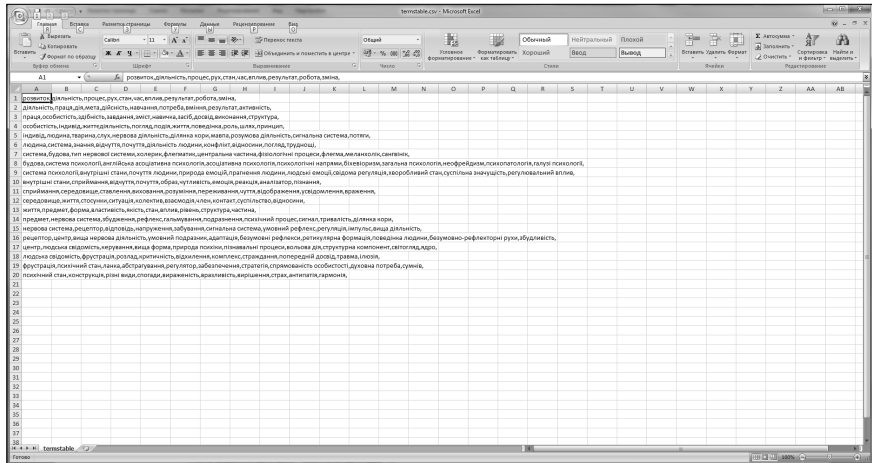


Рис. 4.7. Текст джерела 4

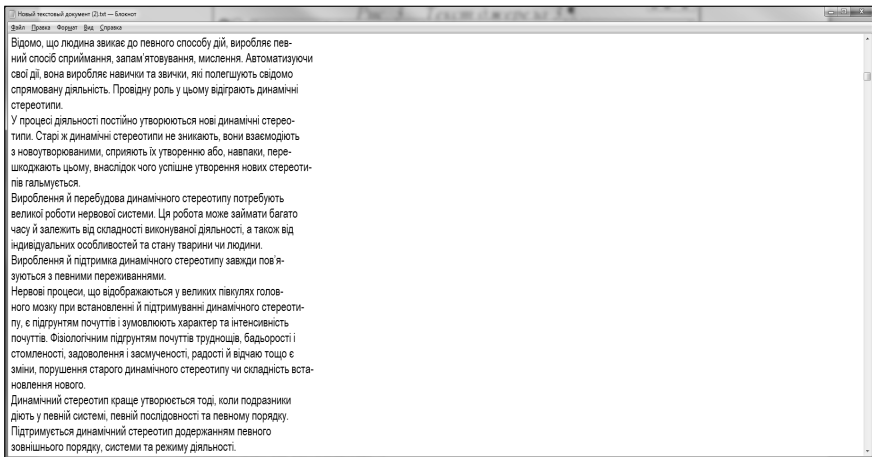


Рис. 4.8. Таблиця психологічних термінів

РОЗДІЛ 5

Технологія контент-аналізу лексики художнього твору

(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 Вкраїні вірші приймали
як благовісне, пророче слово.*

Панас Мирний

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 способів вирішення цікавої для нього проблеми. Якщо йдеться про мовознавчу площину, то такий учень досконало володіє засобами літературної мови, має сформовані читацькі інтереси, є цікавим співрозмовником. Юні мовознавці у студентські роки 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 наразі так само, як і в ХХ, і в ХІХ ст. викликає інтерес і у відомих учених, і в юних дослідників. «Кобзар» читають і перечитують, кожного разу знаходячи у цій невичерпній скарбниці нові досі непомічені перлини. Для прикладу наведемо 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» (Мацько Л. І. Тарас Шевченко – основоположник нової української літературної мови // Українська мова в освітньому просторі [Текст]: навчальний посібник для студентів-філологів освітньо-кваліфікаційного рівня «магістр». – К.: Вид-во НПУ імені М. П. Драгоманова, 2009. – С. 185–186, С. 191).

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, наданих 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 та практичної стилістики, літературознавства та інших філологічних галузей.

Користуючись наведеною вище «Інструкцією для користувача програмою 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 та студентської молоді сприяє створенню полісуб'єктного освітнього середовища.

Нижченаведені матеріали є орієнтиром для студентів, школярів та вчителів / керівників науково-дослідницької роботи, – у процесі поетапного вирішення будь-якої проблеми наукового філологічного дослідження.

5.1. Зміст технології: дослідження концепту «Україна» у текстах збірки Т. Г. Шевченка «Кобзар» за допомогою програми TextTermin

Підготовчий етап

• З'ясування та обґрунтування позицій

Наукова галузь дослідження

Дослідження лексеми «Україна» здійснюється у площині стилістики української мови як галузі конкретного мовознавства. Під час роботи використовуються Інтернет-ресурси 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 концепту «Україна». Практична цінність полягатиме у:

- встановленні семантичних відтінків поняття «Україна» на основі кількісно-якісного аналізу масиву текстів творів Тараса Шевченка;
- доборі в реченнях прикладів вживання слова «Україна» та похідних від нього слів і виразів.

Матеріали роботи зможуть використати не лише школярі, але й мовознавці, літературознавці, викладачі-філологи, учителі-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 Інтернет-мережі, які підтверджують висловлене. Наприклад: *концепт* – «зміст поняття, смислове значення імені».

- **Формулювання мети дослідження, якої можна досягти завдяки вирішенню завдань (відповідно до мікротем у межах теми) – кроки 1–4**

Мета роботи полягає у тому, щоб окреслити філософсько-педагогічні, психологічні, лінгвістичні засади концепту «Україна» та встановити семантичні відтінки цього слова і похідних від нього слів на основі кількісно-якісного аналізу масиву текстів творів Тараса Шевченка засобом програми TextTermin.

5.2. Основний етап: вирішення завдань дослідження

1. Ознайомлення з «Інструкцією для користувача» програми TextTermin.

2. Пошук якісної електронної версії збірки «Кобзар» в Інтернет-мережі (бажано у форматі Microsoft Office Word).

3. За допомогою інструкції переведення тексту «Кобзаря» у формат Блокнот. Збереження джерела у папці (рис. 5.1).

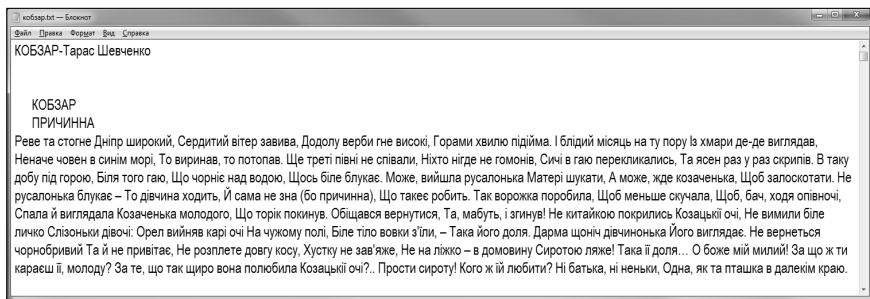


Рис. 5.1. Текст джерела

4. Запуск програми TextTermin. Виділення джерела («Кобзар»), лінгвістичний аналіз якого здійснить програма (рис. 5.2).

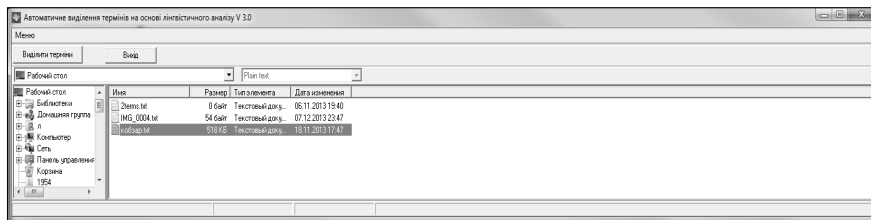


Рис. 5.2. Вибір джерела

5. Ознайомлення з результатами лінгвістичного аналізу (рис. 5.3).

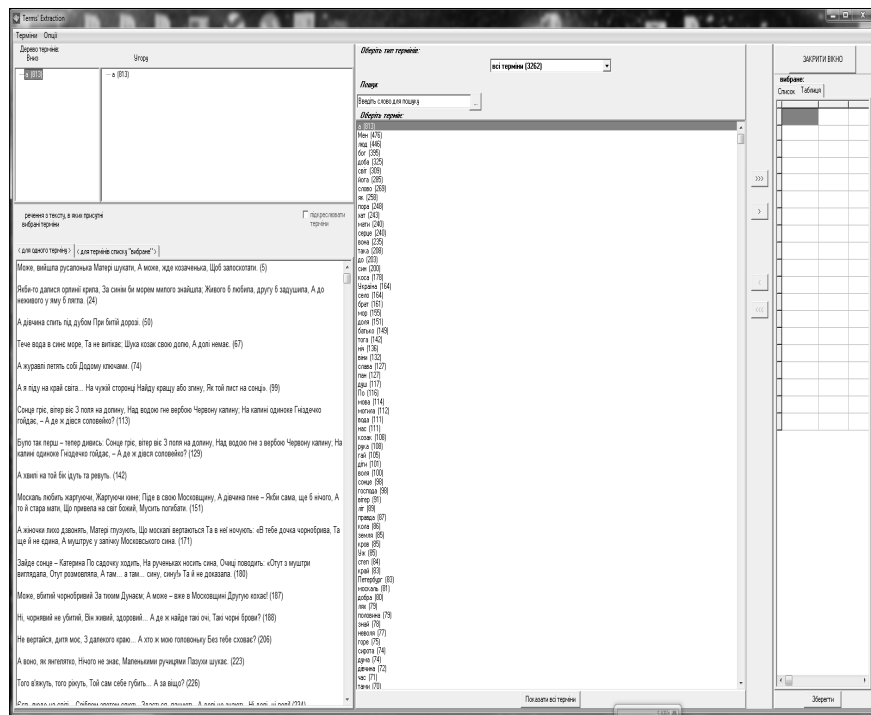


Рис. 5.3. Створення програмою списку термінів

Як видно з рисунку (екран монітора), загальна кількість лексем, виділених програмою – 3262. Слово «Україна» повторюється 164 рази.

6. Виділимо слово «Україна» у списку термінів (рис. 5.4). За бажанням можна задати опцію «підкреслювати терміни». Тоді у контекстах зліва слово «Україна» буде підкреслюватися.

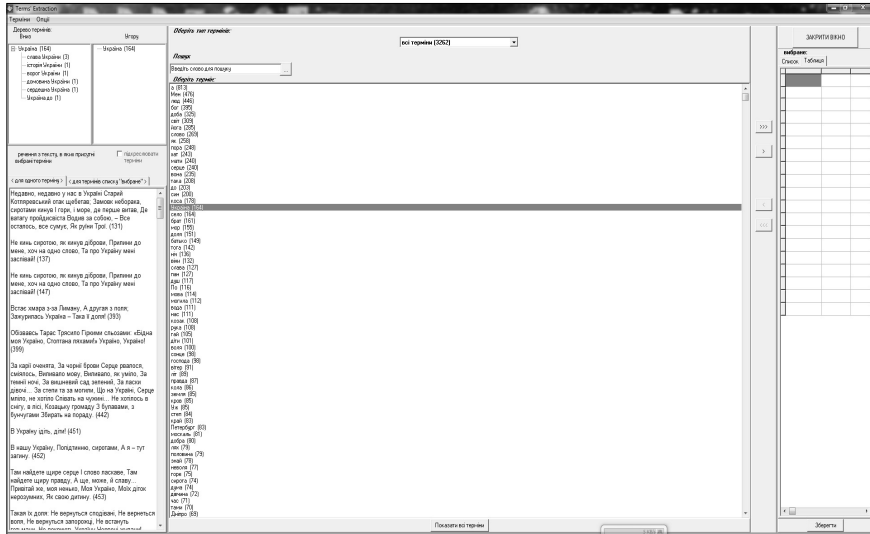


Рис. 5.4. Робота програми із заданим словом

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 щецбетав; Замокв неборака, сиротами кинув І гори, і море, де перше витав, Де ватагу пройдисвіта Водив за собою, – Все осталоь, 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)*
- *Обізвавсь 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 Збирать 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 щиру правду, А ще, може, й славу ... Привітай же, моя ньенько, Моя Україно, Моїх діток нерозумних, Як свою дитину. (453)*

Аналіз кожного з контекстів допомагає визначити семантичні відтінки слова «Україна»: територія, населення, країна, ньенька; зафіксувати перенос значення (Зажурилась Україна – метонімія); встановити граматичні відмінкові форми слова: в Україну, на Україні, про Україну, Україно тощо.

8. Для подальшого створення «дерева» відносин між словом «Україна» та іншими словами у творах збірки «Кобзар» у напівручному режимі створюємо таблицю, копіюючи слова у списку термінів та вставляючи їх послідовно у рядки таблиці (рис. 5.5).

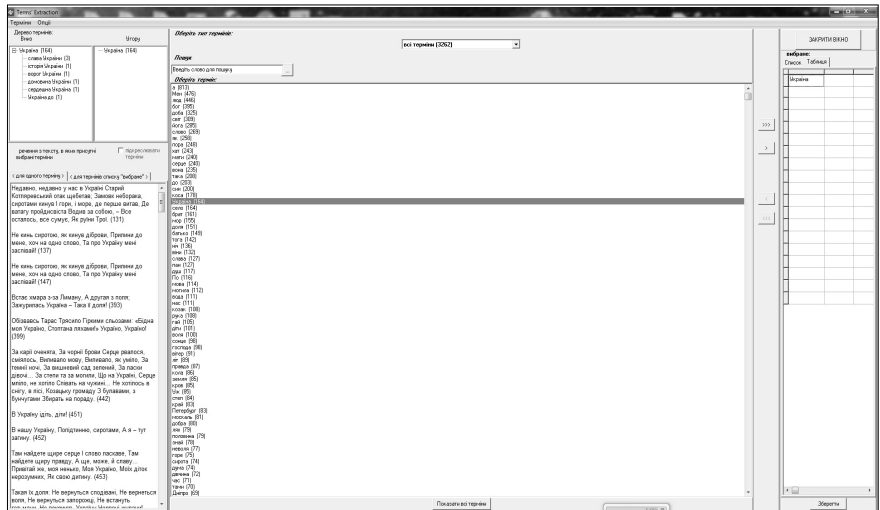


Рис. 5.5. Заповнення таблиці

Збережена таблиця має наступний вигляд (рис. 5.6).

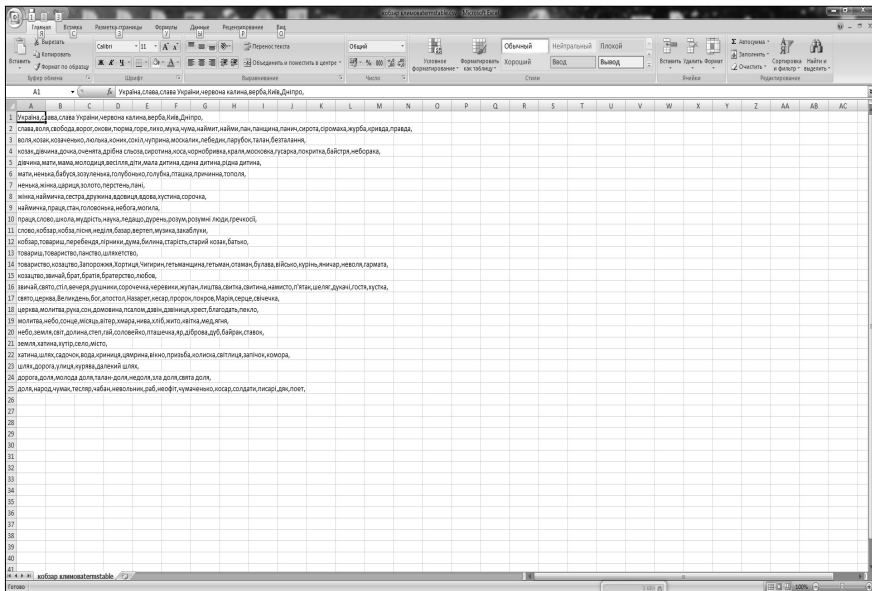


Рис. 5.6. Таблиця – основа для створення «дерева» відносин між словами

9. Наступним етапом є оформлення результатів практичної частини дослідження і повного тексту науково-дослідницької роботи.

Отже, наведена 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 програми TextTermin та відкритих джерел Інтернету досягти поставленої мети.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. *Андон Ф. И.* Логические модели интеллектуальных информационных систем [Текст] / Ф. И. Андон, Л. Е. Яшунин, В. И. Резниченко. – К.: Наукова думка, 1999. – 397 с.
2. *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. – [Электронный ресурс]. – Режим доступа: www.inteltec.ru/publish/articles/textan/RCDL2004.shtml
3. *Вернадський В. І.* Вибрані твори по історії науки [Текст] / В. І. Вернадський. – М.: Наука, 1981. – С. 214–232.
4. Використання розподілених інформаційних ресурсів в навчальному процесі: методичні рекомендації [Текст] / О. В. Лісовий, М. А. Попова, Н. І. Поліхун та ін.; за ред. канд. техн. наук В. В. Камішина і канд. техн. наук О. Є. Стрижака. – К.: Інформаційні системи, 2010. – 228 с.
5. *Вонсович В. П.* Використання імітаційних технологій і прийомів у навчально-професійній діяльності студентів / В. П. Вонсович. – [Електронний ресурс]. – Режим доступу: http://www.nbu.gov.ua/portal/Soc_Gum/Vchu
6. *Гаврилова Т. А.* Базы знаний интеллектуальных систем [Текст] / Т. А. Гаврилова, В. Ф. Хорошевский. – СПб.: Питер, 2001. – 384 с.
7. *Гладун В. П.* Конспектирование естественно-языковых текстов / В. П. Гладун, В. Ю. Величко // XI-th International Conference «Knowledge-Dialogue-Solution»(KDS'2005). – Varna, Bulgaria, 2005. – vol. 2 – P. 344–347.

8. *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, O Corcho (2004). – Springer, 2004. – 403 с.
9. *Gruber T. R.* A translation approach to portable ontology specifications / T. R. Gruber– Knowledge Acquisition, 5 (2), 1993. – P. 199–220.
10. *Gruber T. R.* Toward Principles for the Design of Ontologies Used for Knowledge Sharing / T. R. Gruber. – International Journal of Human and Computer Studies, 43(5/6), 1995. – P. 907–928.
11. *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. – P. 3–15.
12. *Di Battista G., Eades P., Tamassia R., Tollis I.G.* Algorithms for Drawing graphs: an Annotated Bibliography // Computational Geometry, Theory and Applications. – 1994. – № 4. – P. 235–282.
13. *Зайцева Е. Н.* Информационно-обучающая среда как способ развития самостоятельной работы студентов при изучении иностранному языку: автореф. дис. ... канд. пед. наук [Текст] / Е. Н. Зайцева. – Ярославль, 2003. – 23 с.
14. *Клещев А. С.* Отношения между онтологиями предметных областей [Текст] / А. Клещев, И. Артемьева. – Информационный анализ. – Ч. 1. – Вып. 1. – 2002. – С. 4–9.
15. *Марка Д. А.* Методология структурного анализа и проектирования [Текст] / Д. А. Марка, К. Мак Гоуэн. – М.: «МетаТехнология», 1993. – 239 с.
16. Методология функционального моделирования IDEF0. РД IDEF0 – 2000 // Госстандарт России. – Москва. – 2000. – 75 с. – [Электронный ресурс]. – Режим доступа: www.nsu.ru/smk/files/idef.pdf
17. Методы и средства автоматизированного проектирования прикладной онтологии [Текст] / Б. В. Добров, Н. В. Лукашевич, О. А. Невзорова, Б. Е. Федунев. – Изв. РАН. Теория и системы управления. – М., 2004. – № 2. – С. 58–68.
18. Наказ Президента України № 344/2013 «Про Національну стратегію розвитку освіти в Україні на період до 2021 року» від 25.06.2013.

19. *Непеїна Г. В.* Освіта для сталого розвитку: витоки та перспективи [Текст] / Г. В. Непеїна // Наукові праці. – Вип. 99. – 2009. – С. 20–24.
20. *Палагін А. В.* Архитектура онтолого-управляемых компьютерных систем [Текст] / А. В. Палагін. – Кибернетика и системный анализ, 2006. – № 2. – С. 111–124.
21. *Палагін О.* Міждисциплінарні наукові дослідження: оптимізація системно-інформаційної підтримки [Текст] / О. Палагін, О. Кургаєв. – Вісник НАН України, 2009. – № 3. – С. 14–25.
22. *Палагін А. В.* Системная интеграция средств компьютерной техники [Текст] / А. В. Палагін, Ю. С. Яковлев. – Вінниця: УНІВЕРСУМ, 2005. – 680 с.
23. *Палагін А. В.* Системно-онтологический анализ предметной области / А. В. Палагін, Н. Г. Петренко. – УСИМ, 2009. – № 4. – С. 3–14.
24. Педагогика. Большая современная энциклопедия [Текст] / Сост. Е. С. Рапацевич. – М.: Соврем. слово, 2005. – 720 с.
25. *Попова М. А.* Онтологический интерфейс как средство представления информационных ресурсов в ГИС-среде [Текст] / М. А. Попова, А. Е. Стрижак // Ученые записки Таврического национального университета имени В. И. Вернадского. – 2013. – Т. 26 (65). – № 1. – С. 127–135. – (Серия: «География»).
26. Сталий розвиток суспільства: роль освіти [Текст] / В. Підліснюк, І. Рудик, В. Кириленко та ін. – К.: Вид-во СПД «Ковальчук», 2005. – 88 с.
27. *Ставрова О. Б.* Современный урок технологии с применением компьютера [Текст]. – М.: Школьная пресса. – 2004.
28. *Стрижак О. Є.* Засоби онтологічної інтеграції і супроводу розподілених просторових та семантичних інформаційних ресурсів. – Екологічна безпека та природокористування [Текст]: 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; ред. кол.: О. С. Волошкіна, О. М. Трофимчук (голов. ред.) та ін. – К., 2013. – Вип. 12. – 1988 с.
29. Теоретичні основи проектування інформаційних середовищ, як педагогічних систем, спрямованих на підтримку творчої діяльності

учнів: монографія [Текст] / за ред. канд. техн. наук В. В. Камишина і канд. техн. наук О. Є. Стрижака. – К.: Інформаційні системи, 2010. – 194 с.

30. Толковый словарь по искусственному интеллекту [Текст] / авт.-сост. А. Н. Аверкин, М. Г. Гаазе-Рапопорт, Д. А. Поспелов. – М.: Радио и связь, 1992. – 256 с.

31. *Уваров А. Ю.* Компьютерная коммуникация в учебном процессе [Текст] // Пед. информатика. – 1993. – № 1. – С. 34–37.

ДОДАТКИ

ДОДАТОК А

Зразки завдань для дослідження життя та творчості Т. Г. Шевченка з використанням онтологій, побудованих за допомогою програми TextTermin

Існує безліч інформаційних джерел про життя та діяльність Т. Г. Шевченка.

Процес обробки біографій програмою TextTermin складається з декількох етапів.

На першому етапі з біографії Т. Г. Шевченка виділяються поняття-концепти онтології, виражені іменниками. Поряд з кожним поняттям в дужках вказано кількість згадувань цього терміну в тексті.

Зі списку термінів обирається той, що нас цікавить з точки зору відображення в онтології життя та творчості Т. Г. Шевченка (наприклад, «поет»). У лівому вікні програми TextTermin відображається список термінів пов'язаних з уже обраним (дерево термінів).

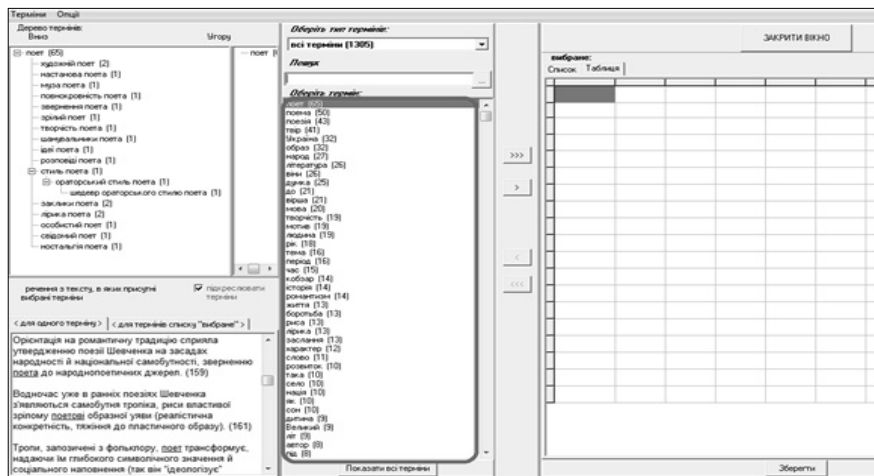


Рис. 1. Вікна системи TextTermin

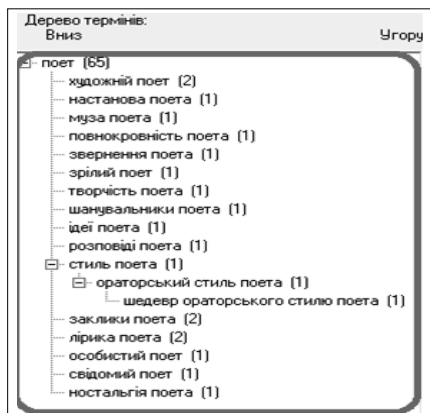


Рис. 2. Фрагмент вікна термінів

Після натискання на слово чи словосполучення у вікні «Дерево термінів», активується режим перегляду контексту, в якому зустрічається обраний вислів.

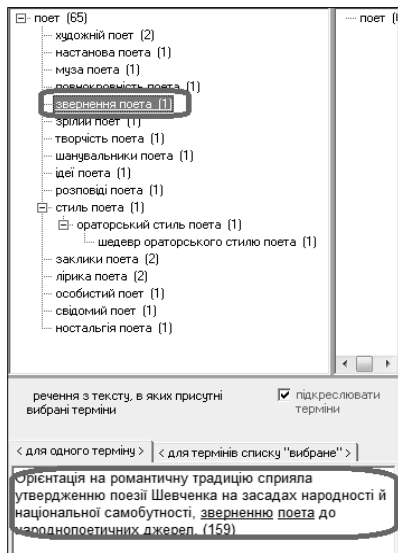


Рис. 3. Відображення контексту терміна

За таким алгоритмом заповнюються інші рядки таблиці.

поет	поет	настанова поета	звернення поета	творчість поета	ідеї поета	стиль поета	заклики поета
поезія	поезія	політична поезія	невільницька поезія	інтелектуальна поезія			
мотив	мотив	сентиментальний	мотив кохання				
заслання	заслання	період заслання	роки заслання				

Рис. 6. Приклад таблиці класів

Оскільки нами в програмі TextTermin розглядається та обробляється декілька інформаційних джерел про біографію Т. Г. Шевченка, то у результаті цього ми отримуємо декілька таблиць структур онтології. Для об'єднання їх в єдину таблицю структури онтології можна скористатись таблицею MSExcel, в яку переносимо дані з усіх отриманих попередньо таблиць. Для зручності кожен клас виділяється кольором (цей прийом стане у нагоді під час заповнення таблиці наповнення онтології).

Клас об'єктів онтології			Об'єкти онтології				
	A	B	C	D	E	F	G
1	Тарас Григорович Шевченко	Тарас Григорович Шевченко	ЛЮДИНА	ПОЕТ ПИСЬМЕННИК	МИТЕЦЬ	ГРОМАДСЬКИЙ ПОЛІТИЧНИЙ ДІЯЧ	ЧАС
2	ЛЮДИНА	ЛЮДИНА	Дитинство і молодість	Вишук	В Академії мистецтв	Перша подорож Україною	Поет проти
3	ПОЕТ ПИСЬМЕННИК	ПОЕТ ПИСЬМЕННИК	Поезія	Драматичні твори	Повісті	Археологічні нотатки	Лик
4	МИТЕЦЬ	МИТЕЦЬ	Художня творчість				
5	ГРОМАДСЬКИЙ ПОЛІТИЧНИЙ ДІЯЧ	ГРОМАДСЬКИЙ ПОЛІТИЧНИЙ ДІЯЧ	Кирило-Мефодіївське братство	Революціонери-декабристи	"Просвіта"	Поет проти імперії	на сучас
6	Художня творчість	Художня творчість	За жанром	За часом	За тематикою	Незавершені роботи	
7	За жанром	За жанром	Портрети	Композиції	Архітектурні пейзажі й краєвиди		
8	За тематикою	За тематикою	Опія	Ажурель	Сеня	Туш	Опія
9	Незавершені роботи	Незавершені роботи	Епоси	Начерки	Ескізи		
10	За часом	За часом	1830 — 1846	1847 — 1850	1851 — 1857	1857 — 1861	
11	ЧАС	ЧАС	Ушумування пам'яті	Вітани творчості Т. Г. Шевченка			

Рис. 7. Приклад сформованої таблиці класів

Варто звернути увагу, що «Художня творчість» є одночасно об'єктом онтології, що належить до класу «МИТЕЦЬ», та класом об'єктів онтології, що охоплює об'єкти «За жанром», «За часом», «За технікою» та «Незавершені роботи». В такому разі термін з об'єктів онтології копіюється до класів об'єктів та описується так само у рядок, як було наведено вище.

Після завершення заповнення таблиці структури онтології життя та творчості Т. Г. Шевченка (її можна назвати «Шевченко_структура.xls»), її необхідно конвертувати Web-додатком Confor.

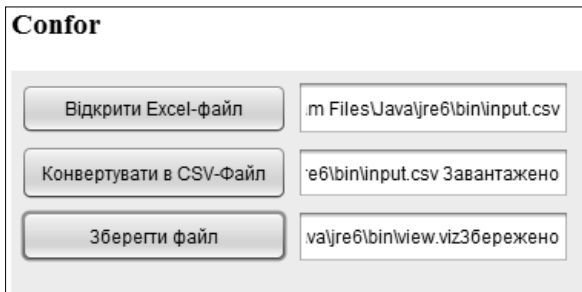


Рис. 8. Підготовка до конвертації таблиці у таксономію

Таким чином, структура онтології життя та творчості Т. Г. Шевченка, що описана в таблиці MSExcel «Шевченко_структура.xls», візуалізується як граф в Web-додатку GraphEditor «Шевченко_структура.xml».

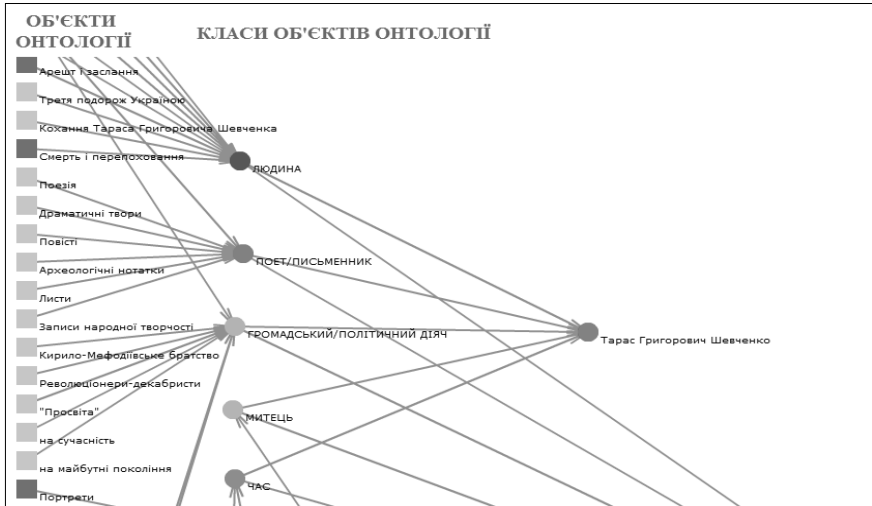


Рис. 9. Приклад таксономії

Вершинами графу є поняття, що пов'язані з біографією та творчістю Т. Г. Шевченка і описані в таблиці структури. Вершини з назвами іпостасей Шевченка відображені тим самим кольором, що і в таблиці (з метою полегшення візуального сприйняття). Вершини з назвами термінів, що розкривають іпостасі, відображені сірим кольором за замовчуванням. Залежно від рівня заглибленості у вивчення питання життя та творчості Тараса Григоровича досліджені вершини будуть відображатись різними кольорами (за допомогою меню «Редагування» в GraphEditor).

Ребрами графу, що об'єднують вершини, є зв'язки між поняттями та класами, які об'єднують ці поняття. Напрямок стрілок вказує до якого підкласу (а далі – класу) належить поняття.

Для роботи із пошуковою машиною необхідно виділити текст, прикріплений до вершини онтології, клікнувши лівою клавішею миші,

і натиснути кнопку «Пошук тексту». Graph Editor автоматично здійснить пошук виділеного тексту в електронній бібліотеці інформаційно-аналітичної платформи Exalead.

Завдання 1. Заповніть вершини онтології прикладами-ілюстраціями та текстовими описами, використовуючи функцію редагування вершин GraphEditor.

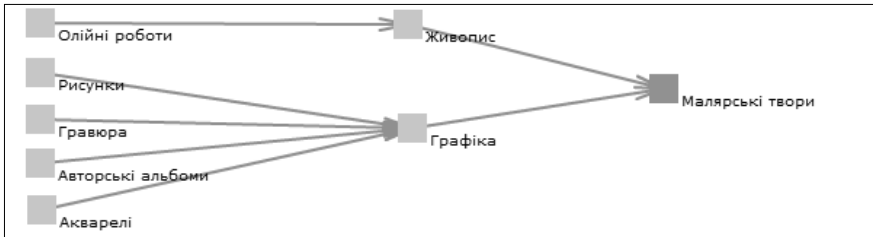


Рис. 10. Фрагмент таксономії за класом «Малярські твори Шевченка»

Завдання 2. Порівняйте офорти А. Дюрера «Велика пушка» та Т. Шевченка «Мангишлацький сад».

Намагаючись досягнути феномен Шевченка-художника, його твори часто порівнюють з роботами А. Дюрера, Рембрандта, шукають спільні риси творчості з російськими художниками О. Кипренським, П. Соколовим. З ім'ям Т. Г. Шевченка пов'язують початковий етап вивчення історії новітньої російської гравюри, а також проводять паралелі з творчістю засновників національних шкіл Східної Європи.



Рис. 11. Означення певних вершин онтології

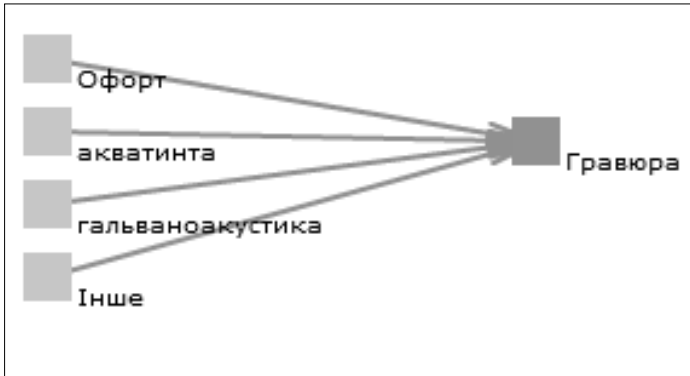


Рис. 12. Фрагмент онтології за класом «гравюри»

Тричі за художні успіхи Шевченко нагороджувався срібними медалями (1839, 1840, 1841), а після закінчення Академії мистецтв (1838–1845) він отримав звання некласного художника.

Завдання 3. Скориставшись пошуковою машиною, дізнайтеся, за які твори Шевченко було нагороджено срібними медалями. Назвіть дату отримання Шевченком звання некласного художника.

Рисунок «Циганка-ворожка» відзначений 26 вересня 1841 р. Радою Академії мистецтв срібною медаллю другого ступеня (див.: Тарас Шевченко: Документи та матеріали до біографії. – С. 35).

Звання некласного художника історичного й портретного живопису Шевченко отримав пізніше – 22 березня 1845 р. (див.: Ковальов В. В. Спогади про Т. Г. Шевченка).

Більша частина його малярства (приблизно 800 творів) зберігаються у Національному музеї Тараса Шевченка у м. Києві. Його твори є також у збірках музеїв Києва, Львова, Харкова, Одеси, Канева, у музейних і приватних колекціях Москви, Санкт-Петербурга, Казахстану, Польщі, США.

The screenshot shows the Exalead search engine interface. At the top, there is a search bar with the text 'Пetersburgska akademija misteciv' and a search button. Below the search bar, there are navigation links and a list of search results. One result is highlighted, showing a document preview with the following text:

Document preview 70 occurrences < previous Пetersburgska akademija misteciv modala next >

рибу». Ці скульптори за свої роботи тоді одержали перші золоті медалі.
 ...для одной картини Карла Павловича стоить прихвтити из Китая, а не только из Мавороссии. — На цей же виставці експонувалася картина К. П. Брюллова «Розп'яття» (1837 — 1838, олія), тепер зберігається в Російському музеї (Санкт-Петербург).

Получил первую серебряную медаль за этюд с натуры. Еще написал небольшую картину масляными красками «Сиротка мальчик делится мякотью с собакой под забором». — За рисунок з натуры «Два натур» 544 ниниз Шевченко одержав срібну медаль другого ступеня 29 квітня 1839 р.; за картину, названу в документі «Хлопчик жабрак дає хліб собаці», також одержав срібну медаль другого ступеня 27 вересня 1840 р. (див.: Шевченко Тарас. Документи та матеріали до біографії. — С. 20, 27). Обидва твори не знайдені.
 ...рано по утрам ходили с Иохимом на Смоленские кладбище лопухи и деревья рисовать. — У НМТШ нині зберігаються рисунок Шевченка «Куток Смоленського кладовища» (1840, олівець, сепія) та етюди рослин

Рис. 13. Текстовий супровід онтології «життя та творчості Шевченка»

Завдання 4. Роздивіться карту, присвячену Т. Г. Шевченку. Знайдіть умовні позначення, передивіться додаткову інформацію про перелічені музеї в різних країнах світу.

Скориставшись функцією редагування, розмістіть власні доробки, фотографії з екскурсій, опис музеїв тощо.

The screenshot shows a GIS environment with a map of Ukraine. A search window is open, displaying the following information:

Назва	Национальный музей Тараса Шевченка
Країна	Україна
Місто	Київ
Адреса	бул. Шевченка, 12
Сайт	http://museumshevchenko.org.ua
Координати y	50,443357
Координати x	30,515752
Принтіти	

At the bottom of the search window, there is a button labeled 'Поблизувати к'.

Рис. 14. Відображення значень вершин онтології в ГІС-середовищі

Завдання 5. Розгляньте онтологію. Чому, на Вашу думку, виділено чотири періоди? Нановніть вершини онтології власними міркуваннями.

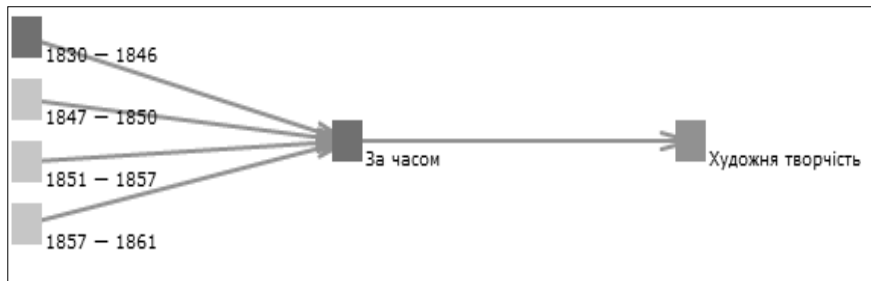


Рис. 15. Фрагмент онтології за відношенням «час»

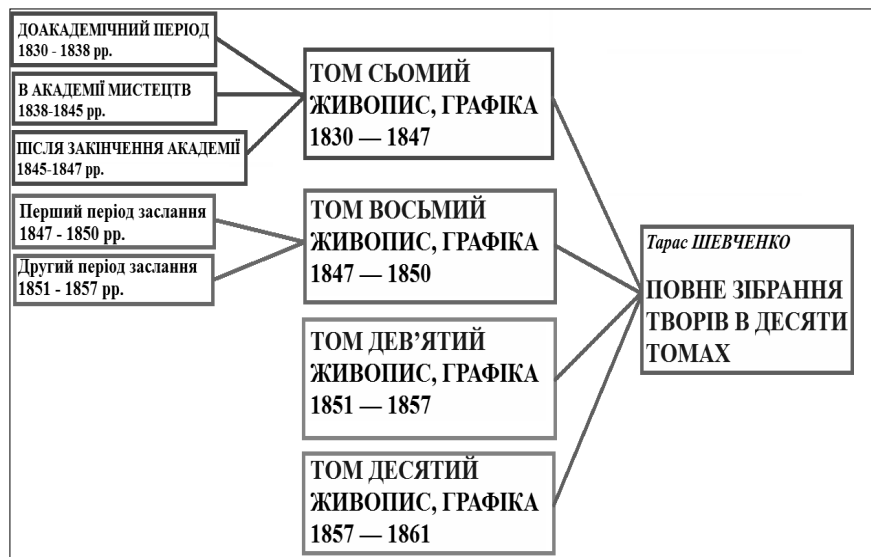



Рис. 16. Фрагмент онтологічного відображення творчості Шевченка за часом

Завдання 6. Скориставшись пошуковою машиною, дізнайтеся про інші автопортрети Т. Г. Шевченка.

exalead  default businessConsole

Автопортрет Advanced search

Results 1-10 of 205 for Автопортрет, Page 1 - 11-20 >

Results per page: 10 Sorted by: Relevance

<> 28. Автопортрет. Тарас Шевченко. Повне зібрання творів в десяти томах. Том. 7. Кн. 1 Preview

Автопортрет
<http://izbornyk.org.ua/shevchenko/shev7028.htm> - 4,5 KB - Sep 2, 2011
 file_name: shev7028.htm
 Source: Shevchenko
 Document type: text/html
 File extension: htm
 Language: Russian, Ukrainian

<> Сумнівні. 191. Автопортрет у білому кашкеті. Тарас Шевченко. Повне зібрання творів в десяти томах. Том. 8. Preview

Автопортрет у білому кашкеті ... За спільністю деталей з оригінальним автопортретом — серією 1849 — 1850 рр. (див. № 47) — можна визнати, що цей автопортрет, як і автопортрет № 192, є авторським повторенням, що, однак, вимагає подальшого дослідження.
<http://izbornyk.org.ua/shevchenko/shev8191.htm> - 4,3 KB - Sep 2, 2011
 file_name: shev8191.htm
 Source: Shevchenko
 Document type: text/html
 File extension: htm
 Language: Ukrainian
 Location: Львів, Оренбург, Москва, Україна

<> 192. Автопортрет у білому кашкеті. Тарас Шевченко. Повне зібрання творів в десяти томах. Том. 8. Preview

Автопортрет у білому кашкеті
<http://izbornyk.org.ua/shevchenko/shev8192.htm> - 3,9 KB - Sep 2, 2011
 file_name: shev8192.htm
 Source: Shevchenko
 Document type: text/html
 File extension: htm
 Language: Ukrainian
 Location: Оренбург

<> 193. Автопортрет у чорному кашкеті. Тарас Шевченко. Повне зібрання творів в десяти томах. Том. 8. Preview

Автопортрет у чорному кашкеті
<http://izbornyk.org.ua/shevchenko/shev8193.htm> - 3,8 KB - Sep 2, 2011
 file_name: shev8193.htm
 Source: Shevchenko
 Document type: text/html
 File extension: htm
 Language: Ukrainian
 Location: Оренбург

<> 60. Автопортрет за свічкою. Тарас Шевченко. Повне зібрання творів в десяти томах. Том. 10. Preview

Автопортрет за свічкою
<http://izbornyk.org.ua/shevchenko/shev10060.htm> - 4,1 KB - Sep 2, 2011
 file_name: shev10060.htm
 Source: Shevchenko
 Document type: text/html
 File extension: htm
 Language: Ukrainian

<> 61. Автопортрет у світлому костюмі. Тарас Шевченко. Повне зібрання творів в десяти томах. Том. 10. Preview

Автопортрет у світлому костюмі
<http://izbornyk.org.ua/shevchenko/shev10061.htm> - 3,8 KB - Sep 2, 2011
 file_name: shev10061.htm
 Source: Shevchenko
 Document type: text/html
 File extension: htm
 Language: Ukrainian

<> 62. Автопортрет у шапці та кожусі. Тарас Шевченко. Повне зібрання творів в десяти томах. Том. 10. Preview

Автопортрет у шапці та кожусі
<http://izbornyk.org.ua/shevchenko/shev10062.htm> - 4,5 KB - Sep 2, 2011
 file_name: shev10062.htm
 Source: Shevchenko
 Document type: text/html
 File extension: htm
 Language: Ukrainian
 Location: Львів, Москва

Search within results

Keywords

- які на неї спіралися x
- Якщо помітили помилку набору х на зціді
- виділіть її мишкою x
- натисніть Ctrl x
- найважливіших напрямів x
- аристотелізму належать концепція
- периплатеїми і середньовічних x араб
- також погляди і концепції x
- Вибрані сторіна x
- філософська доктрина x
- Аристотеля
- Київська старина x

Source

- Shevchenko (205) x

Document type

- text/html (205) x

File extension

- htm (201) x
- html (2) x

Last modification

- 2013 (2) x
- November (2) x
- 2011 (203) x
- September (203) x

Language

- Ukrainian (182) x
- Russian (32) x

Event

- Нью-Йорк 1956 (5) x
- Прага 1926 (5) x
- Нью-Йорк 1961 (3) x
- Париж 1956 (3) x
- Прага 1926 (3) x
- Нью-Йорк 1963 (2) x
- Нью-Йорк 1956 (2) x
- Нью-Йорк 1957 (2) x
- documenta (2) x
- Париж 1959 (2) x

People

- Николай Иванович (16) x
- Андрей Иванович (11) x
- Александр Македонский (9) x
- Свєтлїи Снегїи (8) x
- Эльа Гуса (8) x
- Лев Михайлович (8) x
- Михаил Семенович (7) x
- Лев Толстой (6) x
- Николай Дмитриевич (6) x
- Карл Павлович Ернolloв (6) x

Location

- Україна (113) x
- Львів (109) x
- Київ (100) x
- Україна (80) x
- Кавказ (77) x
- Москва (74) x
- Париж (56) x
- Оренбург (35) x
- Прага (29) x
- Київ (48) x

Organization

- Рус (5) x
- Augsburger Allgemeine Zeitung (4) x
- Kurier (3) x
- Canadian Tribune (2) x
- Fischer (2) x
- Neues Deutschland (2) x
- Tai (2) x
- Natura (2) x
- UNESCO (2) x
- Alaska Herald (2) x

Рис. 17. Знаходження відповідних документів за онтологією «автопортрети»

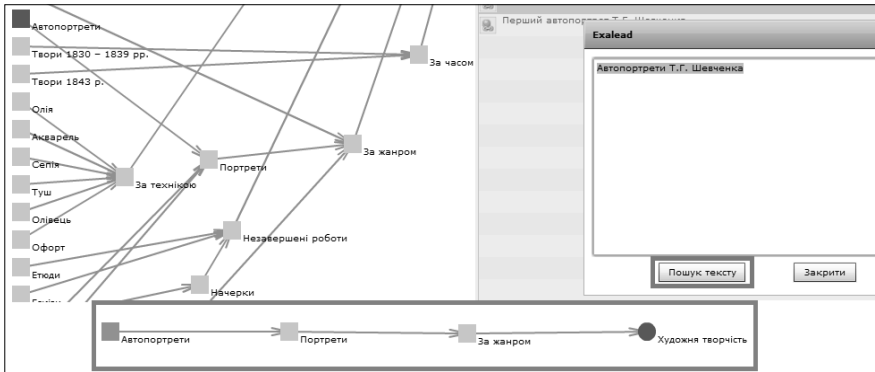


Рис. 18. Відображення типів об'єктів онтології «художня творчість»

Завдання 7. Дослідіть зв'язки творів «Катерина» поеми та картини, часу та місця їх написання.

КАТЕРИНА

Василію Андреевичу Жуковському на пам'ять 22 апреля 1838 года

I
Кохайтесь, чорнобриві,
Та не з москалями,
Бо москалі — чужі люде,
Роблять лихо з вами.
Москаль любить жартуючі,
Жартуючі кизне;
Піде в свою Московшину,
А дівчина гине...
Якби сама, ше б нічого,
А то й стара мати,
Що привела на світ Божий,
Мусить погибати.
Серце в'яне співаючи,
Коли знає, за що;
Люде серця не побачать,
А скажуть — ледшо!
Кохайтесь ж, чорнобриві,
Та не з москалями,
Бо москалі — чужі люде,
Згнушаються вами.

"Катерина"
1842 р., Санкт-Петербург

Рис. 19. Композиційне відображення різних типів об'єктів онтології («художня творчість», «поет»)

Завдання 8. Використовуючи пошукову машину, дослідіть твір «На пасіці».

На пасіці. Етюди та інші ескізи і начерк. Папір, олівець (14,8 × 11,9). [Н. р V 1843].

Датовано відповідно до малюнка на лицьовій стороні аркуша «Сліпий» з авторською датою 1843 р.

У центрі – етюд до картини «На пасіці», справа – частина такого ж етюда (обрізана), зліва та вгорі частини ескізів до «Судньої ради» (?), «Сліпого» та «Селянської родини» (?). Зліва вгорі начерк постаті, що стоїть.

На пасіці. Полотно, олія (53 × 41). [Н. р. V 1843].

До реставрації картини на звороті був напис, що невірно пояснював зміст картини: *Dziewczyna rozmawiająca z kozakiem. Obrazek malowany przez Szewczemke* (Дівчина розмовляє з козаком. Картина мальована Шевченком). Нижче іншою рукою: *Szewczenko*. Зараз полотно дубльоване.

Картину датовано за етюдом до неї, що знаходиться на звороті малюнка «Сліпий» («Невольник») з авторською датою 1843 р.

The screenshot shows the Exalead search engine interface. The search results for 'На пасіці' by Taras Shevchenko are displayed. The results include two entries: one for a set of studies and sketches (295 items) and another for a set of sketches (76 items). The diagram below the search results illustrates the ontology of the artist's work, showing a flow from 'Етюди' (Studies), 'Начерки' (Sketches), and 'Ескізи' (Sketches) to 'Незавершені роботи' (Unfinished works), which then leads to 'Художня творчість' (Artistic creation) and finally to 'МИТЕЦЬ' (The Artist). A specific work, 'На пасіці' (On the Beehive), is highlighted as an example of unfinished work.

Рис. 20. Відображення різних типів об'єктів онтології «митець»

Завдання 9. Ознайомтеся з портретом Єлизавети Василівни Кейкуатової. Знайдіть на карті місце перебування Т. Г. Шевченка під час написання портрету.

Портрет Єлизавети Василівни Кейкуатової. Полотно, олія (66,7 × 56,5). [Бігач]. [Кінець III – 4.IV] 1847.

Зліва внизу олійною фарбою дата і підпис автора: 1847, Т. Шевченко.

Дата уточнюється часом перебування Шевченка в с. Бігач.

The image is a composite of three parts illustrating the relationship between a portrait and its context. On the left is a reproduction of the oil painting 'Portrait of Elizaveta Vasilevna Keikutsova' by Taras Shevchenko. In the center is a map of Europe with a red dot and label 'Місце перебування Т. Г. Шевченка' (Location of T. G. Shevchenko) in the village of Bigach, Ukraine. On the right is a screenshot of a digital document viewer. The viewer shows a text block with the following content:

Екран

Зліва внизу олійною фарбою дата і підпис автора: 1847, Т. Шевченко.

Дата уточнюється часом перебування Шевченка в с. Бігач.

Про час перебування в с. Бігач, Чернівецької пал. (нині Березняківського р-ну, Чернівецької обл.), Шевченко згадує в листі до А. П. Лосєва від 18.11. 1822 р. (там же, 12 листі видання, стор. 68; спочив В. Данич — «Русская старина», т. 1, 1821, стор. 430, в тому разі час чернівецького перебування см. М. А. Дотурову — «Т. Г. Шевченко в депортації», т. 1990, стор. 102).

Єлизавета Василівна Кейкуатова, нар. 1813 р., похована в парку Юзівщини. Діє, Ф. Г. Мозаляев, Малороссийской Работы, т. III, К., 1906, стор. 202; Про неї також сказав Ф. Рачин в листі до М. Біляка (там же, «Русские рассказы», т. II, М., 1916, стор. 211, 221). В літературі портретувана небагато згадується як

Получить текст

Закрыть

Document preview: 9 словесних < review Єлизавета Василівна Кейкуатова > next >


¹⁴¹ Кейкуатова, Єлизавета Василівна (народилася у 1813 р.) — знатна, знайома зи. Ретиних, Шевченко познайомилася з нею в Юзівці у 1841 р. Після божої в м. Ізюм у Кейкуатових в с. Бігач, Чернівецької губ. (за сучасним адміністративним поділом — с. Бігач, Березняківського району, Чернівецької обл.). Зберігав портрет Кейкуатової, написаний Шевченком у 1847 р.

¹⁴² Про віз саме перекази тут йдеться — невідомо, В. М. Ретинга здебільшого робила для Шевченка

Below the screenshot is a diagram showing a path from 'Портрет Єлизаветы Васильевны Кейкуатовой' to 'Портрети' and then to 'За жанром' and 'Художня творчість'.

Рис. 21. Композиційне відображення об'єктів онтології «художня творчість» (портрет, тексти)

Завдання 10. Дізнайтеся дату та місце створення портрету «Казашка Катя». До якого періоду життя та творчості Т. Г. Шевченка він належить?



Национальный центр «Мала академія наук України»
Карта району північ Терек. Г. район міста Шевченка

Назва	Меморіальний комплекс
Храна	Казахстан
Місто	Форт-Шевченко
Адреса	25 км на схід від міста
Сайт	http://www.kazanka.org.uk/

Увійти/Вийти: 44 517107

<> Лис-Льб. Шевченківський словник. У двох томах. Том 1.
<http://lzboryk.org.ua/shechenko/slov22.htm> - 45.3 KB - Sep 2, 2011

Back to results

Document preview 1 occurrence Усього next >

ЛІТЕРАТУРНО-МЕМОРАЛЬНИЙ МУЗЕЙ Т. Г. ШЕВЧЕНКА у Форті Шевченка — державний музей Шевченка в Кат. РСР Відкритий 1.V 1932 за рішенням РНК Казахстану від 12.X 1925 у м. Форт Шевченка на п-ві Мангішлак, де поет відбував заслання 1850 — 57. Розміщений у кол. будівлю (Уссухів), де брав Шевченко. У її ти залах музею зібрано матеріали про життя, творчість і революц. діяльність Шевченка численні фотоматеріали, листи Тараса Шевченка, картини рад. художників, документи про уміщення пам'яті поета в рад. час тощо. Перед музеєм на високому постаменті — мармурове погруддя Шевченка (скульптор В. Бекетський), поряд де один пам'ятник поетові. В Шевченковому саду збираються сім'я і земляни, де працював поет. Над ними зроблено записку споруду Музей проводить велику культ.освітню роботу.

Лис. Пива в І.Музей великого Кобзаря у Форті Шевченка К., 1975.
Є. Укрбібл.

«САД БІЛІХ НОВОПЕТРОВСЬКОГО УКРІПЛЕННЯ» — малюнок, ескіз й інтер'єр Шевченка, виконані 1853 — 57 в Новопетровському укріпленні. В 1853 за наказом команданта І Усового біля укріплення почали садити сад. Дірева для цього привезено з Астраля, Гур'єва та Хагга-Баби. Садити сад допомагали і Шевченко. В саду було споруджено ліній будинок для розвідки команданта алтануку де Усових обшари поставлено юрті в який малював Шевченко, збудовано землянуку для шпигунку Усова, в який поет зостав свої малюнки. Цей сад малював на нь Шевченка. Сад Шевченко відгородив у ала-фокла, рисунках олівцем, ескізи і інтер'єр. Рисунок олівцем «Місце майбутнього саду біля Новопетровського укріплення» (12.1 x 32.5 створено влітку 1853, коли на місці майбутнього саду були лише вологопийні споруди й ліній будинок команданта. З трьох основних малюнків «Сад біля Новопетровського укріплення» одну автарель (12.4 x 29.8) виконано воском 1853, коли в саду ще було надсаджено дірева й збудовано алтануку. На малюнку зображено й казах юрті, розташовані поблизу саду. Другу автарель (17.1 x 29) Шевченко виконав улітку 1854. Проворчу уміну — монограма автора й дата «Ш. 1854». Обидві автарелі виконані поларуаром лиса (Усових). Рисунок олівцем, на високу зображено сад і тильного боку на глізі укріплення, виконаний між 1854 і 1855. Один з інтер'єр олівцем — «Біля Новопетровського укріплення» (11.4 x 28.7) — між 1854 і липнем 1857; другий — «Садикова споруда побудув багача № 2» (12.9 x 28.7) — між 1855 і 1857. На певному плава тут

Рис. 22. Композиційне відображення об'єктів онтології «музеї»

Науково-виробниче видання

**Програмно-інформаційні
засоби формування систем знань
навчального призначення**

Посібник

За редакцією О. Є. Стрижака

Комп'ютерна верстка *Нікіфоров С. П.*
Редактор *Ласкова А. О.*

Підписано до друку 21.11.2014 р. Формат 60x84 ¹/₁₆.
Папір офс. 80 г/м². Друк цифровий.
Умов. др. арк. 8,37. Наклад 300 прим.
Замовлення № 114.

Видано за рахунок державних коштів.
Продаж заборонено

Інститут обдарованої дитини НАПН України
04053, Україна, м. Київ, вул. Артема, 52-Д
тел./факс: (044) 481-27-27
E-mail: iod@iod.gov.ua

Свідоцтво про внесення в Державний реєстр
суб'єктів видавничої справи
серія ДК № 3366 від 13.01.2009 р.