

МАГІСТЕРСЬКА РОБОТА

МР. ШМ - 09.00.00.000 ПЗ

Група ШМ-24-1

Габіцький Роман

2025

Івано-Франківський національний технічний університет нафти і газу

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Габіцький Роман Ігорович

(прізвище, ім'я, по батькові)

УДК 004.9
(індекс)

МАГІСТЕРСЬКА РОБОТА

Онтологічні моделі карт видобутих знань в технологіях

цифрової агрокультури

(назва роботи)

Інженерія програмного забезпечення

(назва освітньої програми)

121 - Інженерія програмного забезпечення

(шифр і назва спеціальності)

Габіцький Р.І.

(підпис, ініціали та прізвище здобувача освітнього ступеня)

Науковий керівник Михайлюк Ірина Романівна, к.п.н., доцент

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Допущено до захисту

Завідувач кафедри

доц. Бандура В.В.

(посада) (підпис) (дата) (ініціали та прізвище)

Нормоконтроль

доц. Вовк Р.Б.

(посада) (підпис) (дата) (ініціали та прізвище)

Робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

Івано-Франківськ – 2025

Івано-Франківський національний технічний університет нафти і газу

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Освітній рівень магістр

Спеціальність 121 – Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:

Зав. кафедрою

ІІЗ

доц.

В.В. Бандура

“ 04 ” вересня 2025 р.

ЗАВДАННЯ

НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТУ

Габіцькому Роману Ігоровичу

(прізвище, ім'я, по-батькові)

1. Тема магістерської роботи “**Онтологічні моделі карт видобутих знань в технологіях цифрової агрокультури**”

керівник проекту (роботи) Михайлюк І.Р., к.п.н., доцент

затверджені наказом закладу вищої освіти від “ 05 ” листопада 2025 р. № 695/7

2. Строк подання студентом проекту (роботи) 15 грудня 2025 р.

3. Вихідні дані до проекту (роботи) Теоретичні концепції та моделі побудови інформаційних та програмних технологій видобування даних і знань

4. Зміст розрахунково - пояснювальної записки(перелік питань, які потрібно розробити)

1. Аналіз предметної області видобування знань в технологіях цифрової агрокультури

2. Огляд методологій побудови карт знань

3. Моделі та методології обробки знань в технологіях цифрової агрокультури

4. Побудова онтологічних моделей карт видобутих знань в домені цифрової агрокультури

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

1. Архітектура системи управління знаннями для сектору цифрової агрокультури (рис. 1.1)

2. П'ять ключових функціональних ролей у системах ІІІ (рис. 1.2)

3. Приклад карт компетенцій (рис. 1.3)

4. Приклад концептуальних карт (рис. 1.4)

5. Триетапна методологія побудови карт знань (рис. 1.5)

6. Консультанти розділів проекту (роботи)

Розділ	Консультант	Підпис, дата
Перевірка на плагіат	доц., к.т.н. Вовк Р.Б.	

7. Дата видачі завдання 04 вересня 2025 р.

Керівник _____

(підпис)

Завдання прийняв до виконання _____

(підпис)

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назви етапів магістерської роботи	Строк виконання етапів роботи	Примітка
1	Підбір і вивчення літератури по темі магістерської роботи	15.09.2025	виконано
2	Аналіз предметної області видобування знань в технологіях цифрової агрокультури	29.09.2025	виконано
3	Огляд методологій побудови карт знань	15.10.2025	виконано
4	Моделі та методології обробки знань в технологіях цифрової агрокультури	08.11.2025	виконано
5	Побудова онтологічних моделей карт видобутих знань в домені цифрової агрокультури	20.11.2025	виконано
6	Приклад імплементації моделі	30.11.2025	виконано
7	Затвердження пояснювальної записки роботи завідувачем кафедри	15.12.2025	виконано

Студент – магістр _____

(підпис)

Керівник роботи _____

(підпис)

АНОТАЦІЯ

Магістерська робота: 80 с., 31 рис., 31 джерело.

Тема: Онтологічні моделі карт видобутих знань в технологіях цифрової агрокультури

Мета роботи - розробка онтологічних моделей карт видобутих знань у сфері цифрової агрокультури, які забезпечать ефективне представлення, інтеграцію, зберігання та використання знань, отриманих у результаті інтелектуального аналізу даних.

Об'єкт дослідження - процеси управління знаннями у цифровій агрокультурі.

Предмет дослідження - онтологічні моделі карт знань, що забезпечують структурування знань, отриманих у результаті інтелектуального аналізу даних у сфері аграрного виробництва.

Результати дослідження

В роботі удосконалено методологію інтеграції знань із різномірних джерел шляхом формалізації процесів розпізнавання та семантичної інтерпретації даних а домені цифрової агрокультури

Висновок

Запропоновано концепцію онтологічної моделі карти знань, яка забезпечує уніфіковане представлення видобутих знань, їх інтерпретацію та подальшу інтеграцію у прикладні системи.

ЦИФРОВА АГРОКУЛЬТУРА, ЗНАННЯ, КАРТА ЗНАНЬ, ОНТОЛОГІЯ, ОНТОЛОГІЧНЕ МОДЕЛЮВАННЯ, УПРАВЛІННЯ ЗНАННЯМИ, ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, РЕПОЗИТОРІЙ ЗНАНЬ, АРХІТЕКТУРА ОНТОЛОГІЇ

ABSTRACT

Master Thesis: 80 pp., 31 fig., 31 sources.

Topic: Ontological models of extracted knowledge maps in digital agriculture technologies

The purpose of the work is to develop ontological models of extracted knowledge maps in the field of digital agriculture, which will ensure effective representation, integration, storage and use of knowledge obtained as a result of intelligent data analysis.

The object of the research is knowledge management processes in digital agriculture.

The subject of the research is ontological models of knowledge maps, which provide structuring of knowledge obtained as a result of intelligent data analysis in the field of agricultural production.

Research results

The work improves the methodology for integrating knowledge from diverse sources by formalizing the processes of recognition and semantic interpretation of data in the domain of digital agriculture

Conclusion

The concept of an ontological model of a knowledge map is proposed, which provides a unified representation of extracted knowledge, its interpretation and further integration into application systems.

DIGITAL AGRICULTURE, KNOWLEDGE, KNOWLEDGE MAP, ONTOLOGY, ONTOLOGICAL MODELING, KNOWLEDGE MANAGEMENT, INTELLECTUAL DATA ANALYSIS, KNOWLEDGE REPOSITORY, ONTOLOGY ARCHITECTURE

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ	10
ВСТУП.....	11
РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ВИДОБУВАННЯ ЗНАНЬ В ТЕХНОЛОГІЯХ ЦИФРОВОЇ АГРОКУЛЬТУРИ.....	14
1.1. Проблематика представлення та управління знаннями, отриманими з інтелектуального аналізу даних у секторі цифрової агрокультури	14
1.2. Концептуальні основи та класифікація знань у системах штучного інтелекту.....	18
1.2.1. Визначення та елементи знань	18
1.2.2. Ролі представлення знань.....	20
1.2.3. Типологія карт знань	21
1.3. Огляд методологій побудови карт знань	22
1.4. Технології зберігання знань	26
1.4.1. Реляційне зберігання графів знань.....	26
1.4.2. Нативне зберігання графів знань	27
1.4.3. Мова запитів та обробка.....	28
Висновки до розділу	28
РОЗДІЛ 2. МОДЕЛІ ТА МЕТОДОЛОГІЇ ОБРОБКИ ЗНАНЬ В ТЕХНОЛОГІЯХ ЦИФРОВОЇ АГРОКУЛЬТУРИ.....	30
2.1. Управління знаннями в галузі цифрової агрокультури	30
2.1.1. Експертні системи	30
2.1.2. Система управління знаннями	32
2.1.3. Репозиторій знань	35
2.2. Онтологічна модель карти знань для представлення видобутих знань	36
2.2.1. Генерація представлення знань.....	37
2.2.2. Роль онтології у представленні видобутих знань	38

2.3 Архітектура та компоненти онтологічно-орієнтованої моделі карти знань	38
2.3.1 Майнінг знань	39
2.3.2 Моделювання знань	40
2.4. Методологія вилучення знань	41
2.4.1. Розпізнавання сутностей	43
2.4.2. Оцінка знань	44
2.4.3. Публікація знань	45
2.4.4. Використання знань	46
2.5. Обробка знань для домену цифрової агрокультури на основі онтологічних моделей	47
2.5.1. Побудова онтології	48
2.5.2. Підготовка знань	49
2.5.3. Система управління знаннями	50
Висновки до розділу	51

РОЗДІЛ 3. ПОБУДОВА ОНТОЛОГІЧНИХ МОДЕЛЕЙ КАРТ ВИДОБУТИХ ЗНАНЬ В ДОМЕНІ ЦИФРОВОЇ АГРОКУЛЬТУРИ	53
3.1. Обґрунтування розробки онтології	53
3.1.1. Вимоги до онтології	53
3.1.2. Процес побудови онтології	54
3.1.3. Принципи проектування онтології	55
3.2. Процес розробки та структура онтології	57
3.2.1. Специфікація концептуального ядра	58
3.2.2. Технічні характеристики та інтеграція	59
3.2.3. Принципи дизайну онтології	60
3.3. Способи обробки і перетворення даних в онтології	62
3.4. Розпізнавання сутностей та підготовка знань	64
3.5. Архітектура системи розпізнавання сільськогосподарських сутностей	66

3.6. Приклад імплементачії моделі	69
Висновки до розділу	74
ВИСНОВКИ	76
ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ.....	78

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

CRF - Conditional Random Fields

ES - Expert System

EKG - Enterprise Knowledge Graph

KMap - Knowledge Map

KMS - Knowledge Management System

NEL - Named Entity Linking

NER - Named Entity Recognition

OAK - Ontology-bAsed Knowledge Map

SC - Semantic Class

SSN - Semantic Sensor Network

ВСТУП

Актуальність теми.

Сучасний етап розвитку сільського господарства характеризується стрімкою цифровізацією виробничих процесів, що зумовлює потребу в ефективному управлінні знаннями, отриманими з різномірних джерел даних. У цифровій агрокультурі щодня генерується величезний обсяг інформації, пов'язаної із моніторингом ґрунтового стану, біологічними характеристиками культур, прогнозуванням врожайності, управлінням зрошенням, застосуванням агрохімікатів тощо. Традиційні підходи до аналізу таких даних не здатні забезпечити повноцінну інтеграцію та ефективне використання накопичених знань. Саме тому виникає необхідність у впровадженні інтелектуальних систем, які базуються на онтологічному моделюванні та здатні створювати карти знань для представлення й подальшого використання результатів інтелектуального аналізу даних.

У даній роботі здійснено аналіз предметної області цифрової агрокультури, розглянуто сучасні підходи до представлення знань, обґрунтовано доцільність використання онтологічних моделей для побудови карт знань, а також розроблено концептуальну та архітектурну модель, орієнтовану на вирішення прикладних задач аграрного сектору.

Актуальність дослідження визначається зростаючою потребою аграрної галузі в систематизації, інтеграції та практичному використанні знань, отриманих із великих масивів даних. Цифрова агрокультура сьогодні базується на застосуванні інноваційних технологій: сенсорних мереж, супутникового моніторингу, безпілотних систем, платформ Big Data та штучного інтелекту. Проте наявність значної кількості різномірних джерел інформації ускладнює процес їх інтерпретації та інтеграції у єдиний інформаційний простір.

Онтологічне моделювання знань дозволяє вирішити зазначені проблеми, оскільки забезпечує формалізацію предметної області, уніфікацію

понять, побудову семантичних зв'язків між об'єктами та створення інтегрованих карт знань. Такі карти є ефективним інструментом не лише для зберігання та обробки інформації, але й для підтримки прийняття рішень, прогнозування та оптимізації агропроцесів.

Таким чином, актуальність роботи зумовлена необхідністю створення нових підходів до управління знаннями, що сприятимуть підвищенню ефективності сільськогосподарського виробництва, забезпеченню продовольчої безпеки та сталому розвитку аграрного сектору в умовах цифровізації.

Метою роботи є розробка онтологічних моделей карт видобутих знань у сфері цифрової агрокультури, які забезпечать ефективне представлення, інтеграцію, зберігання та використання знань, отриманих у результаті інтелектуального аналізу даних.

Об'єктом дослідження є процеси управління знаннями у цифровій агрокультурі.

Предметом дослідження є онтологічні моделі карт знань, що забезпечують структурування знань, отриманих у результаті інтелектуального аналізу даних у сфері аграрного виробництва.

Завдання дослідження:

1. Провести аналіз предметної області цифрової агрокультури та визначити основні проблеми управління знаннями.
2. Дослідити концептуальні основи та методології побудови карт знань.
3. Розробити онтологічну модель карти знань для представлення видобутих знань у цифровій агрокультурі.
4. Визначити архітектуру та ключові компоненти онтологічно-орієнтованої моделі карти знань.

Розробити прикладну імплементацію онтологічної моделі та перевірити її ефективність.

У роботі застосовано методи:

- аналізу та синтезу для дослідження підходів до управління знаннями;

- системного аналізу для визначення вимог до онтологічних моделей;
- методи семантичного моделювання для побудови онтологій;
- методи експертного оцінювання для верифікації побудованих моделей.

Наукова новизна отриманих результатів

Запропоновано онтологічну модель карти знань, орієнтовану на специфіку цифрової агрокультури та розроблено архітектуру системи розпізнавання сільськогосподарських сутностей на основі онтологій.

Практичне застосування результатів

Результати дослідження можуть бути використані у створенні систем підтримки прийняття рішень у сфері цифрового землеробства, розробці інтелектуальних аграрних інформаційних систем та у навчальному процесі для підготовки спеціалістів у галузі цифрових технологій агрокультури.

Структура магістерської роботи. Робота складається зі вступу, трьох розділів та висновків. Загальний обсяг роботи становить 80 сторінок, і містить 31 рисунок, список використаних джерел із 31 найменування.

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ВИДОБУВАННЯ ЗНАНЬ В ТЕХНОЛОГІЯХ ЦИФРОВОЇ АГРОКУЛЬТУРИ

1.1. Проблематика представлення та управління знаннями, отриманими з інтелектуального аналізу даних у секторі цифрової агрокультури

Ефективне управління знаннями (УЗ) є критичною передумовою для конвертації агрономічної інформації у рентабельні рішення в контексті цифрової агрокультури та землеробства. Сільськогосподарські знання генеруються або транслюються через емпіричний досвід фермерів, результати досліджень агрономів, або шляхом аналітичних процесів (інтелектуального аналізу даних), що здійснюються фахівцями з науки про дані. Конкретизуючи зазначене, знання, отримані саме з аналізу даних, вирізняються найвищим ступенем різноманітності, гнучкості та динамізму для підтримки цифрового землеробства.

Аналіз літературних джерел (огляд) дозволяє оцінити наявні підходи до відкриття знань в агропромисловому комплексі. Цей огляд також ідентифікував значні прогалини у представленні та управлінні гетерогенними знаннями в цифровому сільському господарстві. З одного боку, різноманітний та динамічний масив знань, отриманий в результаті інтелектуального аналізу даних, часто виявляється складним для безпосередньої експлуатації та використання різними категоріями кінцевих користувачів. Як приклад, модель, розроблена в процесі інтелектуального аналізу даних, може генерувати знання щодо оптимальних стратегій посіву озимої пшениці з метою максимізації врожайності. Однак такі знання є по суті артефактом прикладної моделі інтелектуального аналізу, який вимагає додаткової інтерпретації з боку експертів. Крім цього, подібні знання можуть бути інкорпоровані у наукові публікації або галузеві звіти. Існує тенденція до наявності численних звітів, присвячених ідентичній проблематиці інтелектуального аналізу даних, які, проте, відрізняються за джерелами,

форматами та аналітичними методологіями. Відсутність послідовності в обміні знаннями ускладнює процес пошуку релевантної інформації для фахівців з науки про дані або агрономів у великому обсязі різнорідних даних. Відтак, існує нагальна потреба у розробці стандартизованого методу представлення знань для підвищення ефективності їх зберігання, управління та пошуку.

З іншого боку, критично важливою вимогою до ефективного представлення знань є подолання проблеми неузгодженості визначень концепцій між різними джерелами знань. Навіть при посиланні на ідентичну концепцію, знання, отримані з різних досліджень, можуть нести різне семантичне навантаження. Наприклад, розглянемо два представлення знань, отримані з різних проєктів інтелектуального аналізу даних, що використовуються для прогнозування оптимальних сільськогосподарських умов для забезпечення високої врожайності озимої пшениці. Семантичне значення концепції «високий врожай» у цих двох представленнях може суттєво відрізнитися, оскільки її визначення залежить від специфічного діапазону або порогового значення, встановленого дослідниками. Як наслідок, подібні знання можуть виявитися несумісними для інтеграції в єдину систему або для зіставлення аналітичних результатів. Дана проблематика також може бути актуальною для прогностичних моделей, які використовують ідентичні вхідні атрибути, але базуються на різних значеннях цих атрибутів.

Таким чином, для представлення знань як набору правил, призначених для вирішення проблем, можуть бути використані системи, засновані на правилах (експертні системи). Проте експертні системи на основі правил демонструють низьку масштабованість, оскільки відсутні ефективні механізми для динамічного уточнення правил, а перевірка узгодженості всієї сукупності правил стає обтяжливою. Відтак, їм бракує гнучкості та адаптивності до мінливих умов. Наприклад, якщо система оперує агрономічними знаннями для відповіді на запит: «Які методи посіву озимої

пшениці для отримання високого врожаю у Великій Британії?», вона зіткнеться зі значними труднощами при обробці запиту: «Які методи посіву озимої пшениці для отримання високого врожаю в Ірландії?». Це обумовлено семантичною невідповідністю між географічними локаціями, навіть за умови ідентичності кліматичних чи ґрунтових умов. Для коректної відповіді на подібний запит такі системи вимагають оновлення своїх наборів правил, що спричиняє перевантаження механізму виведення та ускладнює процес верифікації узгодженості правил.

Альтернативним підходом до представлення знань у сфері штучного інтелекту є імплементація карти знань (knowledge map), яка забезпечує динамічне управління знаннями. Припустимо, фермери або агрономи висувають запитання щодо врожайності озимої пшениці в Україні, тоді як база знань містить інформацію виключно про врожайність у Великій Британії. Система знань потенційно може надати відповідь щодо врожайності в Україні, якщо вона здатна ідентифікувати подібність кліматичних або ґрунтових умов між Україною та Великою Британією. Це передбачає співіснування як подібних, так і відмінних параметрів між цільовим запитом та наявними знаннями. Наприклад, локації є географічно різними, але характеризуються подібними погодними або ґрунтовими характеристиками. Відтак, шляхом аналізу ґрунтових умов та географічної прив'язки, вилучених із системи, експерти можуть формувати рекомендації для фермерів, враховуючи відмінності у локації або часовому періоді. Ця здатність до контекстуального виведення підтверджує динамічну сутність карти знань.

Сучасна система цифрового сільського господарства, як ілюструє рисунок 1.1, є інтегративною комбінацією численних факторів, що формують єдиний комплекс для підвищення продуктивності та забезпечення сталого розвитку. Для досягнення окреслених цілей доцільно імплементувати уніфіковану базу знань цифрового сільського господарства з метою інтеграції

інформаційних систем, що функціонують на різних етапах, в єдину екосистему.

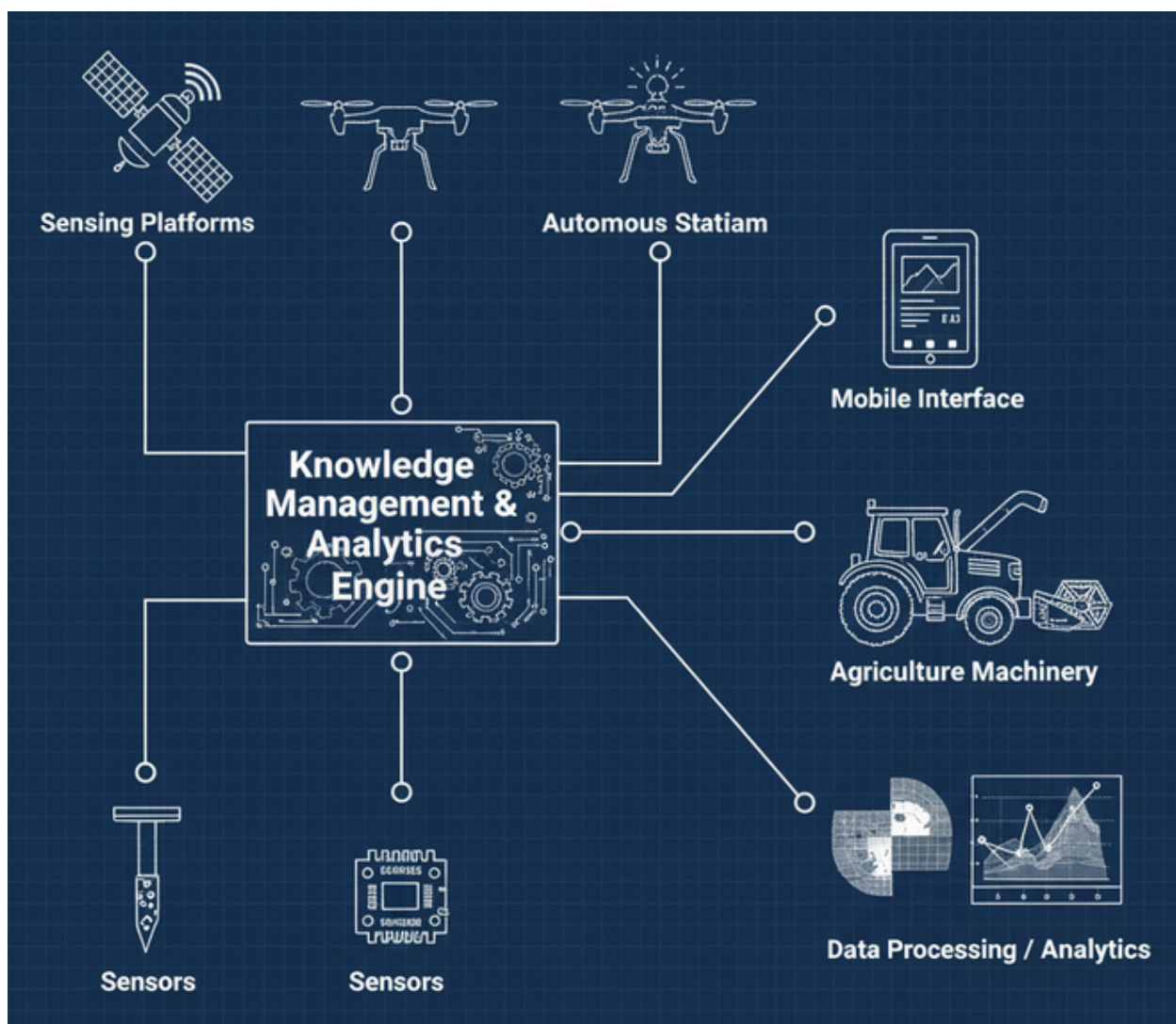


Рис. 1.1. Архітектура системи управління знаннями для сектору цифрової агрокультури

У контексті комп'ютерних наук, база знань може бути реалізована через онтологію, яка надає базові концепції та пов'язану інформацію у визначеній предметній області. Онтологія, що є концептом, ініційованим у рамках розробок Семантичного Вебу, визначається як формалізована семантична система, що містить термінологію, її визначення та специфікацію відношень між цими термінами.

Таким чином, метою даного проєкту є розробка моделі представлення знань, яка буде спроможною не лише репрезентувати виявлені знання у форматі, оптимізованому для пошуку, але й слугувати основою для побудови ефективних онтологічно-орієнтованих систем управління знаннями в секторі цифрової агрокультури.

1.2. Концептуальні основи та класифікація знань у системах штучного інтелекту

1.2.1. Визначення та елементи знань

Визначення поняття «знання» варіюється залежно від контексту його застосування. Часто знання трактується як інформація та розуміння, що індивідуально або колективно набуті суб'єктом. Натомість, дослідження у сфері представлення знань та міркувань пропонують більш формалізоване визначення, стверджуючи, що знання є відношенням між суб'єктом пізнання (знавцем) та пропозицією, що виражена простим декларативним реченням.

У контексті комп'ютерних наук знання, отримані з реальних прикладних систем, є визначальною умовою для розробки інтелектуальних агентів та систем штучного інтелекту (ШІ). Воно сприяє демонстрації раціональної поведінки в архітектурах ШІ. Представлення знань (ПЗ) є дисципліною, спрямованою на формалізацію знань про світ у форматі, який піддається інтерпретації комп'ютерними системами. Це дозволяє ШІ розв'язувати складні прикладні задачі, такі як діагностування патологій рослин на основі візуальних даних або прогнозування стану ґрунту, спираючись на параметри навколишнього середовища чи супутникові знімки.

Структура знань, що моделюються в системах ШІ, типово включає чотири основні елементи :

- Об'єкти (Objects) - пасивні сутності реального світу, що містять статичну інформацію (наприклад, рослина, пшениця, ячмінь, кукурудза).

- Факти (Facts) - декларативні твердження, що відображають істинний стан реального світу та його репрезентацій (наприклад, умови ґрунту, метеорологічні показники).

- Події (Events) - динамічні дії, що відбуваються у реальному світі (наприклад, посів насіння пшениці, збір урожаю, прогнозування погодних умов).

- Виконання (Performance) - опис поведінки, що інкапсулює процедурні знання про алгоритмічні дії та послідовності.

Залежно від методології організації різних елементів, знання можуть бути класифіковані на п'ять типових категорій. До них відносяться: декларативні знання, процедурні знання, структурні знання та евристичні знання.

Декларативні знання також відомі як концептуальні, пропозиційні або описові знання, вони описують концепції, факти та об'єкти. Прикладом є опис морфології озимої пшениці або фізико-хімічних характеристик ґрунтових умов.

Процурні знання відповідають за знання «як робити» щось, і включають правила, стратегії, послідовності дій та процедури. Архітектура GoogleNet є типовим представленням процедурних знань.

Структурні знання описують взаємозв'язки між різними концепціями, включаючи відносини типу «вид-частина», антонімію, синонімію та групування (наприклад, WordNet, ConceptNet).

Евристичні знання репрезентують експертні знання у певній предметній області. Вони включають емпіричні правила, що ґрунтуються на попередньому досвіді та обізнаності щодо прикладних підходів.

Зазначені типи знань екстрагуються, зберігаються та застосовуються в архітектурах ШІ. Вони можуть бути інкорпоровані в різні інформаційні системи, такі як бази знань, графи або онтології. Загалом, існують два основні підходи до набуття знань:

1) Ручна експертна інженерія знань (наприклад, створення таксономій, онтологій або експертних систем на основі правил);

2) Автоматичне відкриття знань за допомогою інструментів інтелектуального аналізу з структурованих та неструктурованих даних.

Дане дослідження сфокусоване на знаннях, отриманих автоматично в результаті завдань інтелектуального аналізу даних (Data Mining), включаючи класифікацію, кластеризацію, регресію та асоціативні правила. У наступних підрозділах буде детально розглянуто проблематику представлення знань та описано різні типи карт знань, які використовуються для їх моделювання.

1.2.2. Ролі представлення знань

Представлення знань виконує п'ять ключових функціональних ролей у системах ШІ (рисунок 1.2):

- Функціонування як замісника об'єкта.
- Встановлення набору онтологічних зобов'язань.
- Формування фрагментарної теорії для інтелектуального міркування.
- Забезпечення механізму для ефективних обчислень.
- Слугування засобом людської експресії.

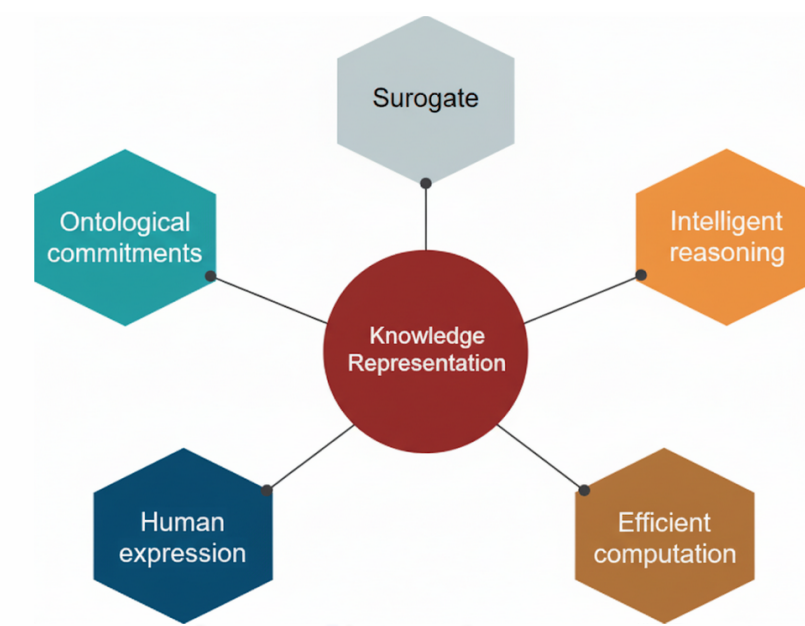


Рис. 1.2. П'ять ключових функціональних ролей у системах ШІ

Ці п'ять ролей забезпечують необхідний компроміс між людським сприйняттям та операційним розумінням агентів ШІ щодо знань реального світу, які вони репрезентують. Різні техніки для представлення знань застосовуються для інкорпорації типів знань (об'єкти, факти, події, виконання) та адаптації до цих п'яти ролей. До таких технік належать логічне та фреймове представлення, семантична мережа (включаючи онтології, карти знань та графи знань) і продукційні правила .

1.2.3. Типологія карт знань

Кarti знань підлягають класифікації за різними критеріями, проте найбільш поширеною є їхня типологія на три основні категорії : концептуальні, компетентнісні та процесові карти.

Компетентнісні карти (карти компетенцій): Використовуються для моделювання відносин між людським ресурсом та знаннями в межах організації. Така карта відображає рівень знань, навичок, здібностей та поведінки кожної особи. Наприклад, рисунок 1.3 ілюструє приклади карт компетентності для оцінки здібностей суб'єкта (Софія) та проміжного дизайнера продуктів.

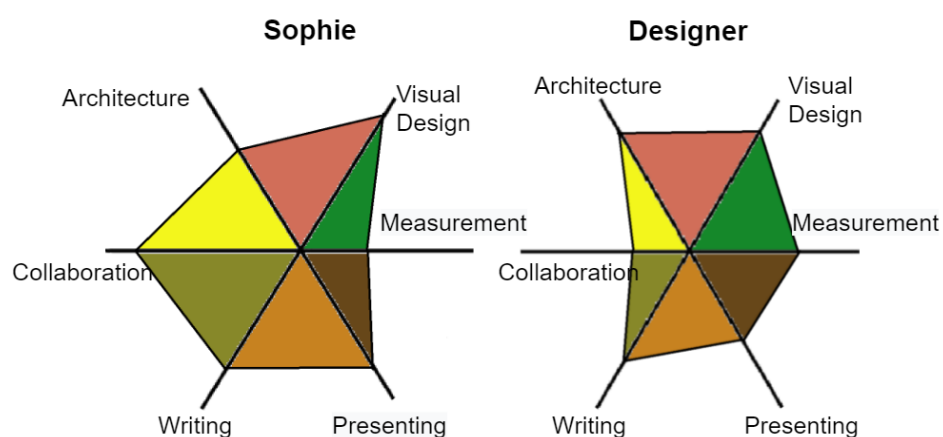


Рис. 1.3. Приклад карт компетенцій

Концептуальні карти описують взаємозв'язки між концепціями та джерелами знань (рис. 1.4).

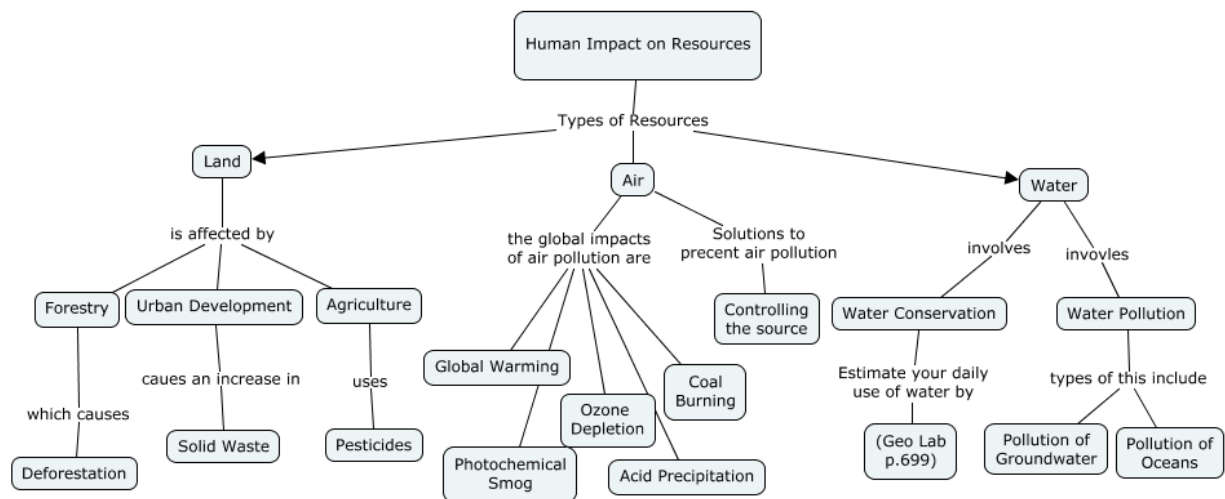


Рис. 1.4. Приклад концептуальних карт

Процесові карти слугують для опису послідовних етапів обробки завдань та процедур, таких як фази обробки даних, що використовуються в моделях прогнозування погоди.

1.3. Огляд методологій побудови карт знань

Наявні численні методологічні підходи до конструювання карт знань (КЗ). Більшість із них орієнтовані на створення корпоративних КЗ, тоді як менша частина призначена для розробки КЗ, специфічних для конкретної предметної області. У цьому розділі представлено огляд існуючих методологій побудови карт знань з подальшим обґрунтуванням вибору найбільш придатної методології для застосування в галузі цифрового сільського господарства.

Методологія одинадцяти кроків [4], є одинадцятиетапним підходом, призначеним для розробки корпоративних карт знань. Вона використовує стратегію етапів життєвого циклу розробки програмного забезпечення (SDLC) — від визначення вимог до тестування КЗ. Основна мета методу полягає у забезпеченні повного розміщення та категоризації всіх документів (як елементів знань) до впровадження системи у виробництво. Реалізація цього процесу вимагає використання спеціалізованого програмного

забезпечення Lotus Discovery Server (LDS) [5]. Обмеженням методології є залежність від специфічного програмного забезпечення, що створює бар'єри для участі співробітників, які не мають відповідного досвіду роботи з LDS.

Методологія шести кроків [9] розроблена для фіксації та представлення організаційних карт знань, які інтегрують профілі знань та бізнес-процеси. Методологічний процес охоплює наступні фази:

- 1) Розуміння знань в організації;
- 2) Аналіз карти процесу знань;
- 3) Витяг знань;
- 4) Профілювання знань;
- 5) Зв'язування карти знань;
- 6) Верифікація КЗ.

Ключовою особливістю є застосування онтології для забезпечення формального представлення знань на етапі розробки, що забезпечує високу концептуальну чіткість КЗ.

Методологія чотирьох фаз [7] складається з таких основних фаз: планування, збір, картографування та перевірка, призначених для візуалізації активів знань. Ця модель інтегрує чотири ключові підходи до картографування — соціальний, компетентнісний, процедурний та концептуальний — в єдину структуру карти. Чотири фази поділяються на одинадцять деталізованих кроків, необхідних для побудови та оцінки КЗ. Варто зазначити, що ця методологія переважно ґрунтується на концептуальній моделі, формально представленій як онтологія.

Методологія трьох етапів [8] відповідно містить три етапи побудови організаційних карт знань на основі онтологій. Початковий етап передбачає створення онтології, структурованої на основі аналізу медичних випадків. Другий етап зосереджений на розробці карти процесу із застосуванням стандартних нотацій (наприклад, блок-схем або мереж Петрі). На завершальному етапі відбувається витяг карти знань з фактичних прикладів процесів (представлених як медичні файли) шляхом використання словника

та відношень, визначених в онтології (етап 1). Методологія була успішно апробована в медичній організації та продемонструвала придатність для створення комбінованої концептуально-процесної карти знань (наведено на рис. 1.5).

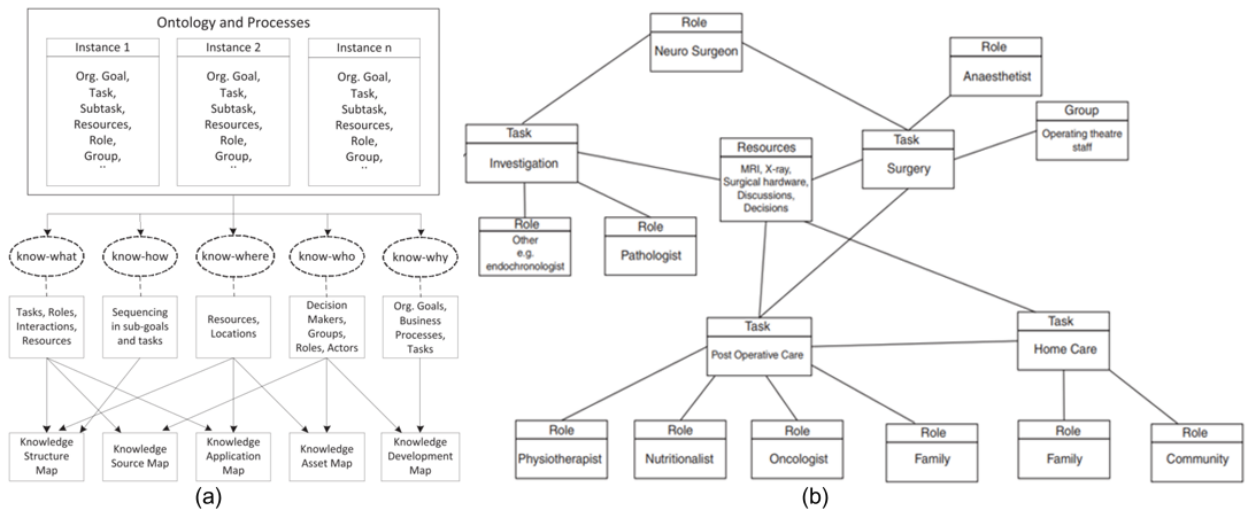


Рис. 1.5. Триетапна методологія побудови карт знань

Методологія семи кроків орієнтована на побудову корпоративних карт знань для матричних організаційних структур. Етапи включають:

- 1) Створення команди розробників;
- 2) Аналіз ресурсів знань;
- 3) Визначення меж бізнес-домену;
- 4) Визначення структури та відношень КЗ;
- 5) Вибір інструментарію розробки;
- 6) Визначення локалізації елементів знань;
- 7) Побудова початкової КЗ, оцінка та оновлення.

Незважаючи на демонстрацію застосовності у матричній організації, деталізація окремих кроків та обмеженість описів знанневих ресурсів є недоліками цієї методології.

В дослідженні [9] представили багаторівневу методологію для побудови корпоративної карти знань (рис. 1.6). Концепція включає три ключові рівні:

- 1) Індивідуальне позначення знань (тегування співробітниками власних елементів знань, наприклад, документів);
- 2) Вибір та анотування тем доменів (теми анотуються відповідно до проектів, у яких беруть участь співробітники);
- 3) Аналіз асоціацій між доменами (визначення зв'язку через розрахунок показників подібності).

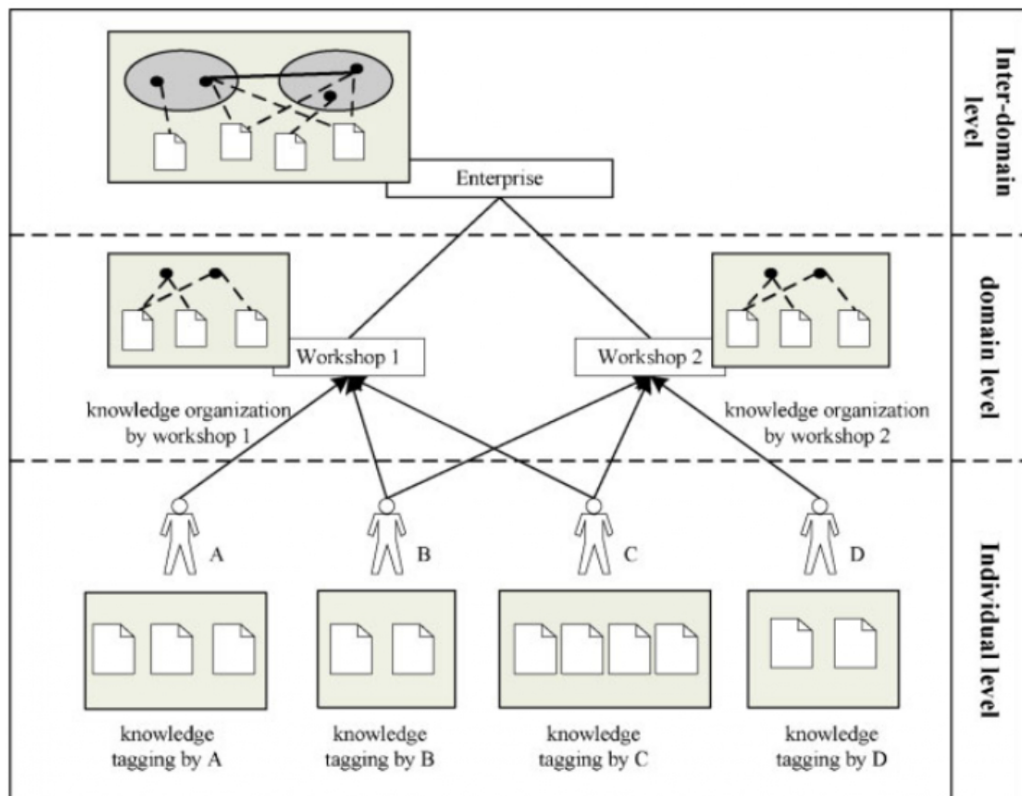


Рис. 1.6. Процедура трирівневої методології для побудови карт знань

У цій КЗ теми знань репрезентуються як вагові вектори, а зв'язок між темами відображається через коефіцієнт подібності. Методологія була розроблена для корпоративного застосування та успішно апробована на практиці. Результатом є набір елементів знань, їх міток, доменних тем та відповідних зв'язків, представлених ваговими коефіцієнтами подібності.

1.4. Технології зберігання знань

У контексті цього дослідження, незважаючи на різноманітність існуючих типів представлення знань, знання моделюються як граф знань (Knowledge Graph, KG), що складається з концепцій та відношень. Відповідно, технології зберігання знань будуть розглядатися з позиції технік збереження графів знань. Зберігання знань визначає спосіб представлення та збереження знань у комп'ютерних програмах. Різні підходи до зберігання впливають на методи ефективного індексування, доступу та виконання запитів до даних. У науковій літературі виділяють два основних підходи до зберігання графів знань: реляційний та нативний.

1.4.1. Реляційне зберігання графів знань

Реляційні бази даних все ще широко використовуються для зберігання графів знань, використовуючи такі схеми:

- Таблиця трійок (Triple Table) - схема прямого зберігання даних RDF. Таблиця triple_table зберігає всі трійки (суб'єкт, предикат, об'єкт) в одній таблиці (рис. 1.7).

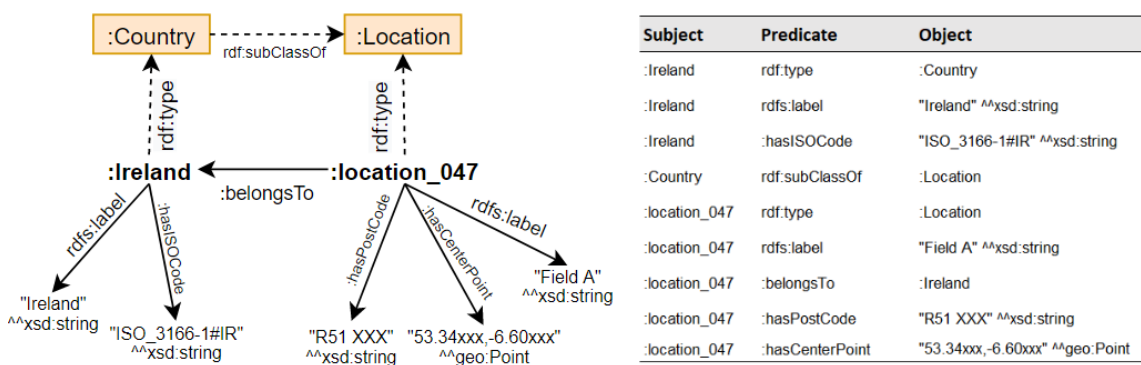


Рис. 1.7. Зберігання графів знань з використанням таблиці трійок

- Горизонтальна таблиця (Horizontal Table) - кожен рядок зберігає всі предикати та об'єкти для одного суб'єкта, що є еквівалентом списку суміжності KG. Усі рядки розміщуються в одній таблиці (рис. 1.8).

Subject	rdf:type	rdfs:label	:hasPostCode	:hasCenterPoint	:hasISOCODE
:Ireland	:Country	"Ireland" ^{^^xsd:string}			"ISO 3166-1#IR" ^{^^xsd:string}
:location_047	:Location	"Field A" ^{^^xsd:string}	"R51 XXX" ^{^^xsd:string}	"53.34xxx,-6.60xxx" ^{^^geo:Point}	

Рис. 1.8. Підхід використання горизонтальної таблиці

- Таблиця властивостей (Property Table) - граф декомпозується на різні таблиці властивостей. Суб'єкти, що мають спільні властивості, кластеризуються в одній таблиці (рис. 1.9). Таблиця властивостей, як правило, містить один стовпець для суб'єкта та n додаткових стовпців для зберігання відповідних властивостей даного суб'єкта.

Country

Subject	rdfs:label	:hasISOCODE
:Ireland	"Ireland" ^{^^xsd:string}	"ISO_3166-1#IR" ^{^^xsd:string}

Location

Subject	rdf:subClassOf	:belongsTo	:hasPostCode	:hasCenterPoint
:location_047	:Location	:Ireland	"R51 XXX" ^{^^xsd:string}	"53.34xxx,-6.60xxx" ^{^^geo:Point}

Рис. 1.9. Підхід використання таблиці властивостей

1.4.2. Нативне зберігання графів знань

Нативне зберігання графів знань (KG) стосується базової структури бази даних, яка безпосередньо містить графові дані (вузли та ребра). Цей підхід має два технічні рішення залежно від вимог до масштабованості:

- RDF Triplestores - призначені для локального або серверного використання (наприклад, Jena, Neo4j).
- Розподілене сховище KG - Призначені для великомасштабних систем (наприклад, Microsoft Trinity, Virtuoso, GraphDB).

RDF Triplestores (RDF-сховища) – це графові бази даних, оптимізовані для зберігання та пошуку трійок — кортежів даних сутностей, що включають суб'єкт, предикат та об'єкт. Переваги RDF Triplestores включають:

- Відсутність необхідності у попередньо визначеній схемі бази даних чи абстрактних об'єктах (таких як таблиці) перед збереженням даних.

- Трійки RDF легко експортуються/імпортуються у форматі N-Triple, що спрощує обробку великих обсягів кортежів.
- Протокол SPARQL підтримує ефективне виконання запитів до великомасштабних RDF-трійок.
- Оскільки технології RDF та RDF Triplestores базуються на унікальних URI, це сприяє легкому обміну та зв'язуванню баз даних.
- Хоча попереднє визначення схеми не є обов'язковим, RDF Triplestores може виводити нові факти або потенційні відношення при інтеграції з існуючою онтологією.

1.4.3. Мова запитів та обробка

Для доступу, пошуку та оновлення вмісту кожна база даних вимагає мови запитів. Хоча для RDF були запропоновані різні мови (наприклад, RQL, SeRQL), протокол SPARQL став стандартом для графів знань, що базуються на RDF-сховищах. SPARQL розроблений для дослідження різноманітних джерел даних, незалежно від того, чи зберігаються дані нативно як RDF. Стандартизація SPARQL та його основа на RDF спрощує дослідження, прийняття та розширення існуючих схем.

У дослідженні [11] автори показали, що графи знань є потужною концепцією для запити великих масивів даних, але їхня доступність для кінцевих користувачів часто ускладнена через необхідність володіння спеціалізованими мовами запитів, такими як SPARQL. Автори також представили систему для перекладу запитань, сформульованих природною мовою, у запити SPARQL. Це вказує на потенціал адаптації технологій RDF та SPARQL як основного механізму зберігання RDF у системі.

Висновки до розділу

Цей розділ надав необхідне теоретичне підґрунтя для магістерського дослідження, включаючи визначення ключових концепцій. У рамках

методології було проведено огляд найпоширеніших підходів для побудови карт знань, як загальноприйнятих, так і корпоративно-орієнтованих. Крім того, описано сучасні технології зберігання знань, які будуть використані для створення репозиторію знань у запропонованій моделі.

РОЗДІЛ 2. МОДЕЛІ ТА МЕТОДОЛОГІЇ ОБРОБКИ ЗНАНЬ В ТЕХНОЛОГІЯХ ЦИФРОВОЇ АГРОКУЛЬТУРИ

2.1. Управління знаннями в галузі цифрової агрокультури

2.1.1. Експертні системи

Експертна система (ЕС) являє собою комп'ютерну програму, призначену для імітації процесу прийняття рішень людським експертом шляхом застосування накопичених знань. У структурі ЕС досвід і знання фахівців формалізуються у вигляді продукційних правил типу "ЯКЩО-ТОДІ" та фактів. Ці правила в подальшому використовуються системою для вирішення проблемних завдань і генерації відповідей на запити (рис. 2.1) [9].

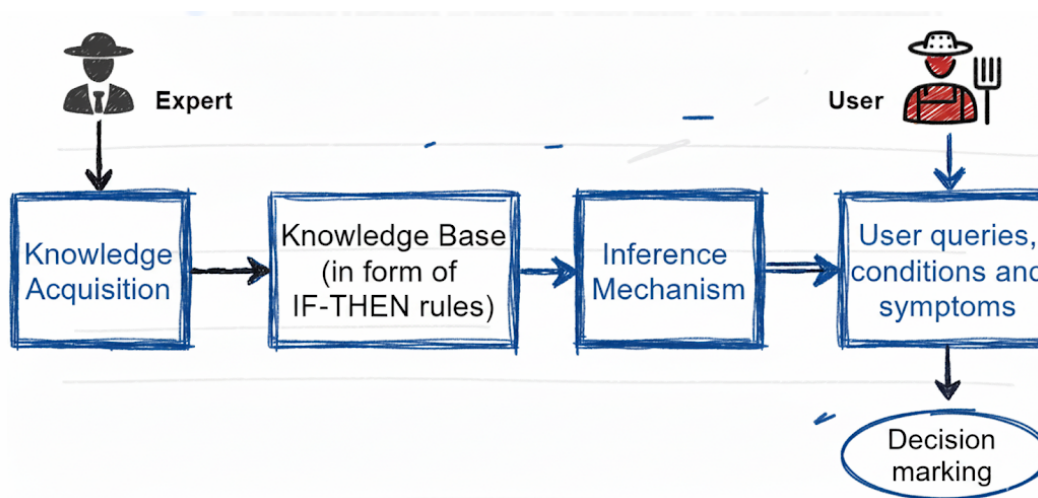


Рис. 2.1. Типова архітектура експертної системи

Отже, ЕС переважно оперують знаннями, структурованими як набір правил. Здобуті від експертів знання вимагають представлення саме у цьому форматі. Нарешті, механізми виведення здійснюють логічні міркування на основі правил, використовуючи техніки прямого ланцюжка (або комбінацію прямого та зворотного ланцюжка) для формування остаточних рішень.

AgriEnt є веб-орієнтованою платформою на основі знань для управління комахами-шкідниками польових культур. Система підтримки прийняття рішень AgriEnt на основі знань має чотири рівні: рівень даних, семантичний рівень, рівень вебсервісів та рівень представлення (рис. 2.2).

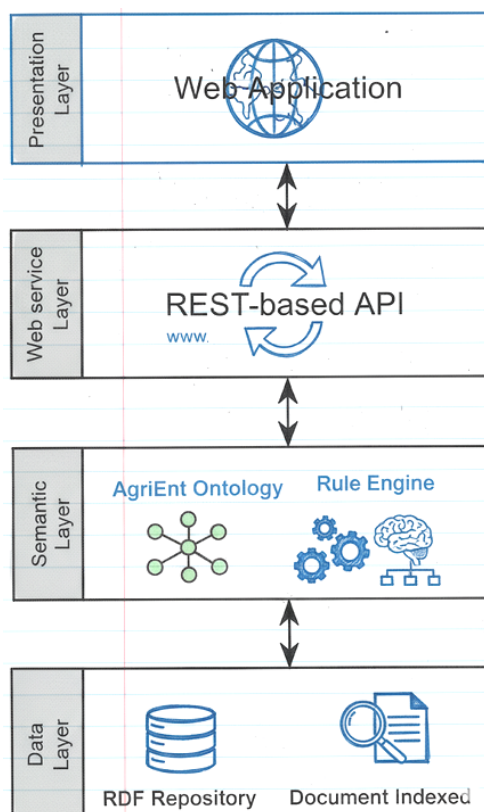


Рис. 2.2. Архітектура ЕС AgriEnt

AgriEnt містить онтологію та механізм виведення на основі правил для забезпечення діагностики комах-шкідників, які впливають на культуру, та надання рекомендацій щодо відповідного лікування. Онтологія AgriEnt має кілька основних класів, типових для діагностики хвороб сільськогосподарських культур (таких як культура, комах-шкідник, симптом, хвороба та лікування) та типових типів їхніх взаємозв'язків (hasSymptom, isCausedBy, hasTreatment). Зрештою, платформа AgriEnt продемонструвала найвищий показник точності (87.5%) при діагностиці комах-шкідників, що вражають кукурудзу, та найнижчий показник (78.5%) для шкідників рису, на основі 149 тестових випадків [12].

Хоча ЕС широко використовуються і відіграють певну роль у цифровому сільському господарстві, особливо в діагностиці хвороб рослин, вони також мають низку властивих обмежень, які важко подолати. Було визначено такі недоліки ЕС у цифровій агрокультури:

- База знань або набір правил потребують постійного розширення та оновлення для адаптації до нових умов та середовищ.

- Система може демонструвати невизначеність у діагностиці, якщо вхідні дані є низької якості, наприклад, зображення з низькою роздільною здатністю.

- Збільшення бази знань спричиняє складність в управлінні та належній послідовності її елементів.

- Для обробки складних запитів необхідно включати та інтегрувати більше суміжних модулів знань, наприклад, модулі дефіциту поживних речовин або прогнозування погоди.

- Експертним системам бракує гнучкості та здатності адаптуватися до мінливих умов середовища або збільшення кількості вхідних умов.

- Обслуговування та розширення бази правил може бути складним для відносно великих систем.

2.1.2. Система управління знаннями

Як було зазначено в першому розділі, граф знань (ГЗ) інкорпорує сукупність вузлів (концепцій), атрибутів (властивостей) та відносин. Як правило, ГЗ характеризуються значною різноманітністю концепцій і складною структурою відносин. З огляду на це, для їх зберігання часто застосовуються технології пов'язаних даних (linked data) або графових баз даних. Графові бази даних є особливо ефективними для обробки елементів даних зі складними взаємозв'язками. Відповідно, технології ГЗ та графових баз даних є перспективними підходами для управління знаннями.

Останнім часом проведено низку досліджень, сфокусованих на застосуванні ГЗ або графових баз даних для управління знаннями. По-перше,

було запропоновано використання графових баз даних (наприклад, Node4j) для процесів управління знаннями в організаційному контексті. Дослідження представило діаграму потоку даних системи управління знаннями, що включає чотири ключові етапи:

- 1) реєстрація та оновлення експертизи,
- 2) пошук експертизи,
- 3) узгодження експертизи з існуючими документами,
- 4) узгодження експертизи з працівниками знань.

У роботі представлена концептуальна реляційна модель даних з чотирма сутностями, які забезпечують цей процес (включаючи експертизу, працівника знань, документ та експертизу працівника). Таким чином, графи, збережені в графовій базі даних, містять екземпляри цих чотирьох типів сутностей та їхні відносини, такі як має, автор, зв'язок_з (як показано на рис. 2.3).

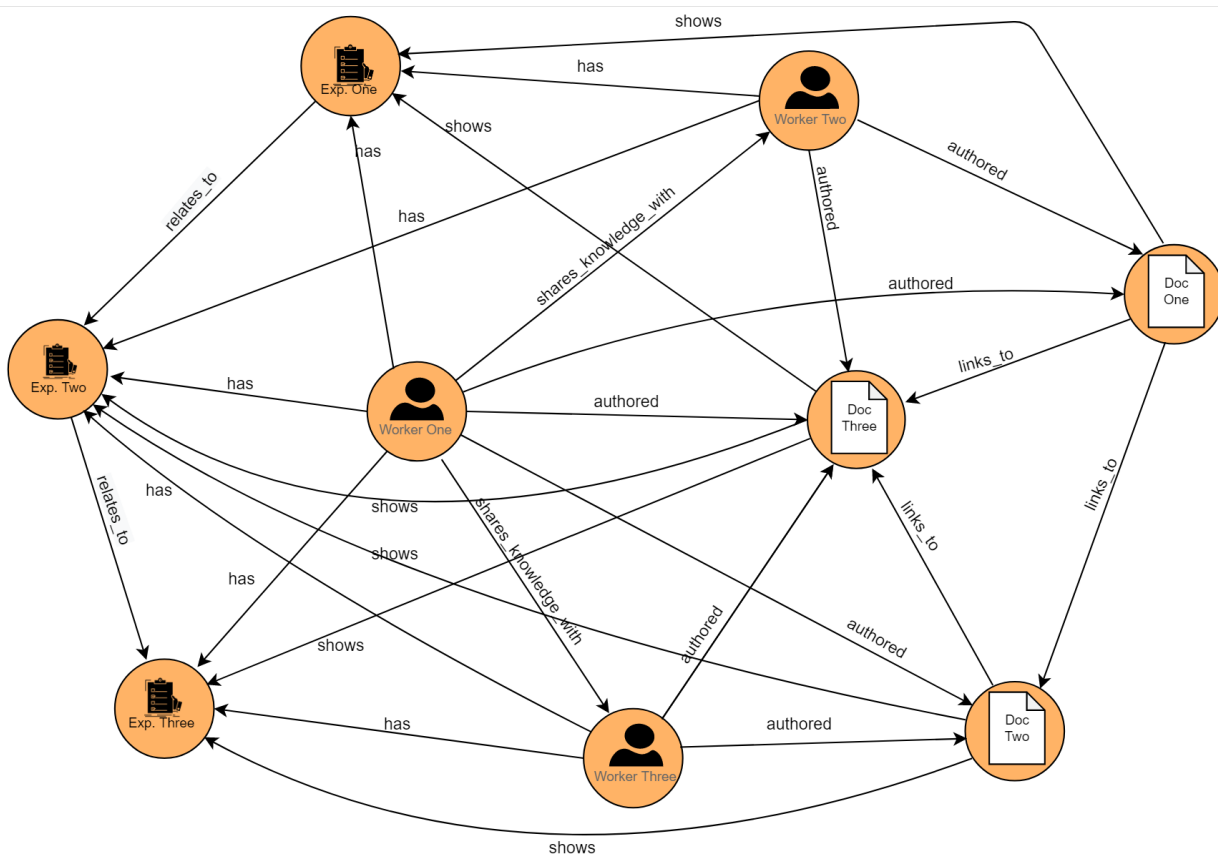


Рис. 2.3. Приклад графу знань в системі управління знаннями

Крім того, в [14] запропонували корпоративний граф знань (КГЗ) як семантичний підхід до управління знаннями. Автори визначили КГЗ як формальну модель для інкорпорування пов'язаних корпоративних даних (LED). Архітектура системи управління корпоративними знаннями є шестирівневою, включаючи:

- 1) сховище даних для тексту, медіа, неструктурованих даних, RDF;
- 2) семантичне озеро даних;
- 3) КГЗ;
- 4) шар онтологічної узгодженості;
- 5) GUI, JSON/XML, SPARQL, REST API;
- б) корпоративні додатки.

В цій структурі КГЗ функціонує як система інтеграції корпоративних даних, що містить онтології та екземплярні дані. Основна роль КГЗ полягає у зв'язуванні екземплярних даних на основі попередньо визначених формальних відносин та обмежень, встановлених в онтологіях.

В [15] запропонували концептуальну структуру для управління знаннями в сільськогосподарському секторі, реалізовану в три етапи. На першому етапі автори представили концептуальну структуру системи управління знаннями. На другому етапі була розроблена онтологія добрив, а третій етап передбачав створення інтерфейсу для доступу реальних користувачів до цієї онтології.

Основним результатом цього дослідження є побудова онтології добрив, яка включає 90 концепцій, 25 об'єктних властивостей та 36 властивостей даних. Інтерфейс для доступу до знань з онтології знаходиться на стадії розробки та базується на SPARQL-запитах та RDF для зберігання онтології. Загалом, хоча автори позиціонують це як концептуальну структуру для управління знаннями, вона скоріше є структурою для спільного використання знань з онтології, а не повноцінною системою управління знаннями.

В цілому, застосування технологій графів знань або графових баз даних для систем управління знаннями залишається на початковій стадії. Проте, цей підхід демонструє переваги у представленні складних взаємозв'язків.

2.1.3. Репозиторій знань

Репозиторій знань — це онлайн-платформа, призначена для збору, організації, зберігання та забезпечення доступу до знань. Корпоративні репозиторії знань здебільшого мають приватний доступ і використовуються для управління внутрішньою та власницькою інформацією. Натомість, доменні репозиторії знань часто є публічно доступними і слугують для управління загальнодоступною інформацією, специфічною для певної предметної області. У сільському господарстві існує кілька репозиторіїв знань з відкритим доступом.

The screenshot displays the AGROVOC Multilingual Thesaurus interface. On the left, there is a hierarchical tree view under 'plant products > cereals > barley > winter barley'. The main content area shows the following details for the preferred term 'winter barley':

- PREFERRED TERM:** winter barley
- DEFINITION:** Winter barley varieties require a period of exposure to cool temperatures (vernalization; loci listed earlier) to initiate flowering. They are usually photoperiod-sensitive, which prevents them from flowering in the winter, even if the cold requirement has been fulfilled. (en)
- BROADER CONCEPT:** barley (en)
- IN OTHER LANGUAGES:** A table listing translations in Arabic, Chinese, French, Georgian, German, Norwegian Bokmål, Romanian, Spanish, and Turkish.
- URI:** http://aims.fao.org/aos/agrovoc/c_5b8bcf68
- Download this concept:** RDF/XML TURTLE JSON-LD. Created 11/11/20, last modified 8/2/21
- EXACTLY MATCHING CONCEPTS:** A table listing related URIs from lod.nal.usda.gov and vocabulary.worldbank.org.

Рис. 2.4. Тезаурус AGROVOC

Зокрема, тезаурус AGROVOC є багатомовною словниковою системою, але він також опублікований як репозиторій знань з відкритим доступом для

довідкових цілей. Репозиторій AGROVOC підтримує перелік або ієрархічну структуру концепцій у цій базі знань. Він також надає атрибути та відносини для кожної концепції. Оскільки ця база знань є багатомовною, репозиторій підтримує як багатомовний інтерфейс, так і вміст концепцій. Користувачі мають можливість здійснювати пошук сільськогосподарських концепцій відповідною мовою. Цей репозиторій можна використовувати для отримання базової інформації про культури, такі як озима пшениця, озимий ячмінь тощо.

2.2. Онтологічна модель карти знань для представлення видобутих знань

Для ілюстрації концепції онтологічно-орієнтованої моделі карти знань (ontology-based knowledge map model) для генерації представлень знань із видобутої інформації, рисунок 2.5 демонструє представлення знань для моделі кластеризації.

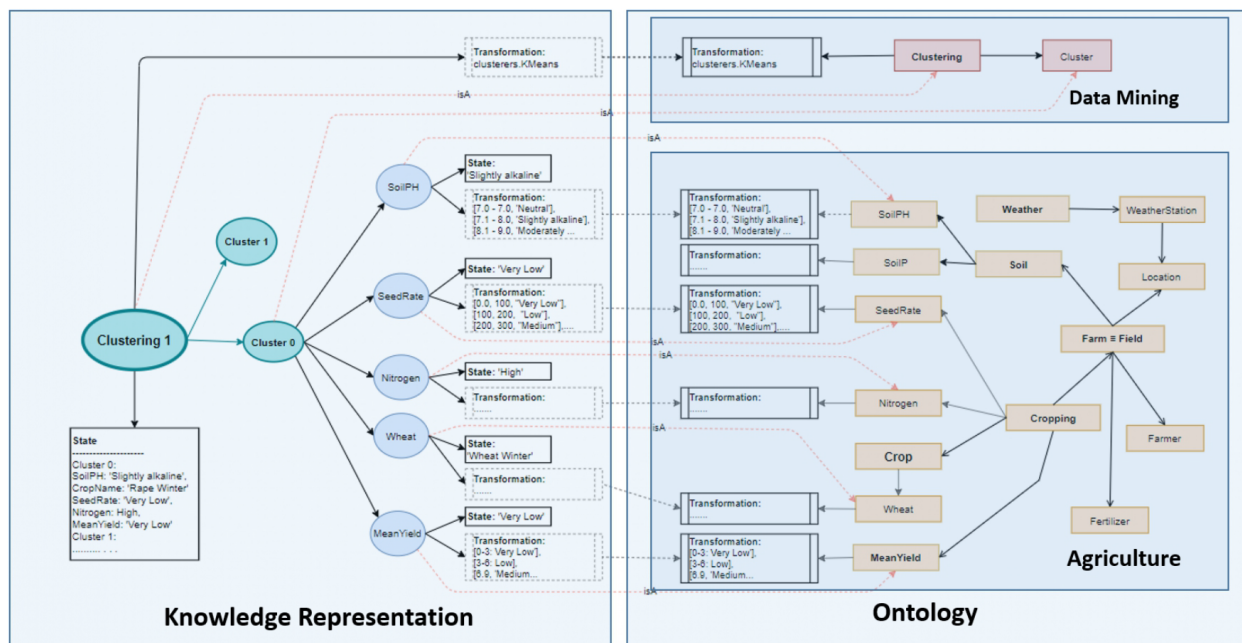


Рис. 2.5. Візуальне представлення онтологічно-орієнтованої моделі карти знань

Ця модель кластеризації має п'ять вхідних умов (input conditions): рН ґрунту, норма висіву, вміст азоту, назва сорту пшениці та середня врожайність. Для адекватного представлення будь-якого знання необхідні фонові знання (background knowledge). В ідеалі, онтологія забезпечує ці фонові знання, що охоплюють перелік концепцій предметної області та їхніх трансформацій. Права частина рисунку 2.5 відображає дві частини онтології: добування даних (data mining) та сільське господарство. Вона використовується як еталон знань для представлення будь-яких елементів знань.

2.2.1. Генерація представлення знань

Представлення знань (ліва частина рисунку 2.5) генерує список інстансів (instances) – SoilPH, SeedRate, Nitrogen, Wheat та MeanYield – на основі вхідних умов та виходу моделі знань (з посиланням на концепції в онтології). Далі ці інстанси пов'язуються з трансформаціями концепцій в онтології, щоб відобразити способи обробки даних цих концепцій.

З іншого боку, також створюються обчислювальні інстанси (computing instances) на основі завдання добування даних і пов'язуються з обчислювальними концепціями в обчислювальній частині онтології. Це також включає посилання на алгоритми, використані у знанні.

Якщо знання є процесним знанням, його представлення включає інстанси та трансформації для відображення концепцій у знанні та способів обробки даних цих концепцій.

Якщо знання є фактичним знанням (fact knowledge), представлення знань також включає значення (values) цих інстансів.

У наведеному прикладі, список інстансів містить інстанси рН ґрунту, норми висіву, вмісту азоту, назви пшениці та середньої врожайності з вхідних умов. Ці інстанси пов'язані з відповідними трансформаціями (Soil pH, Seed rate, Nitrogen, Wheat name та Mean yield) на основі кроків обробки елемента знань. Ці трансформації є попередньо визначеними в онтології.

Інстансам також присвоюються значення або стани (states), якщо вони є елементами знань. У результаті генерується представлення знань (рис. 2.5).

2.2.2. Роль онтології у представленні видобутих знань

Цей приклад демонструє важливу роль онтології у представленні елементів видобутих знань у цифровому сільському господарстві. Концепції є попередньо визначеними в онтології. Представлення знань лише створює інстанси цих концепцій для посилання на них у представленні.

Трансформації також попередньо визначені в онтології. Представлення знань створює відносини (relationships), щоб зв'язати їх із трансформаціями в онтології. Представлення знань також створює стани на основі реальних значень (real values) в елементі знань (за наявності).

Таким чином, онтологія надає попередньо визначені концепції та трансформації, які необхідні для формування представлень знань. Оскільки вони попередньо визначені та спільно використовуються для всієї моделі, ці концепції та трансформації, представлені у різних відображеннях знань, є семантично узгодженими (semantically consistent) у своїх визначеннях. Концепції та трансформації семантично пояснюються за допомогою попередньо визначених описів та відносин, що містяться в онтології.

Навпаки, за відсутності онтології для сприяння представленню елементів знань (як у наведеному прикладі кластеризації), елементи знань можуть використовувати різні визначення для представлення однієї й тієї ж концепції. Крім того, кожен елемент знань потребував би власного набору визначень для інтерпретації його семантики.

2.3 Архітектура та компоненти онтологічно-орієнтованої моделі карти знань

Як ілюструє рисунок 2.6, архітектура онтологічно-орієнтованої моделі карти знань (ОАК) інтегрує наступні ключові функціональні компоненти:

- 1) Майнінг знань,
- 2) Моделювання знань,
- 3) Витяг знань,
- 4) Оцінка знань,
- 5) Публікація знань,
- 6) Використання знань.

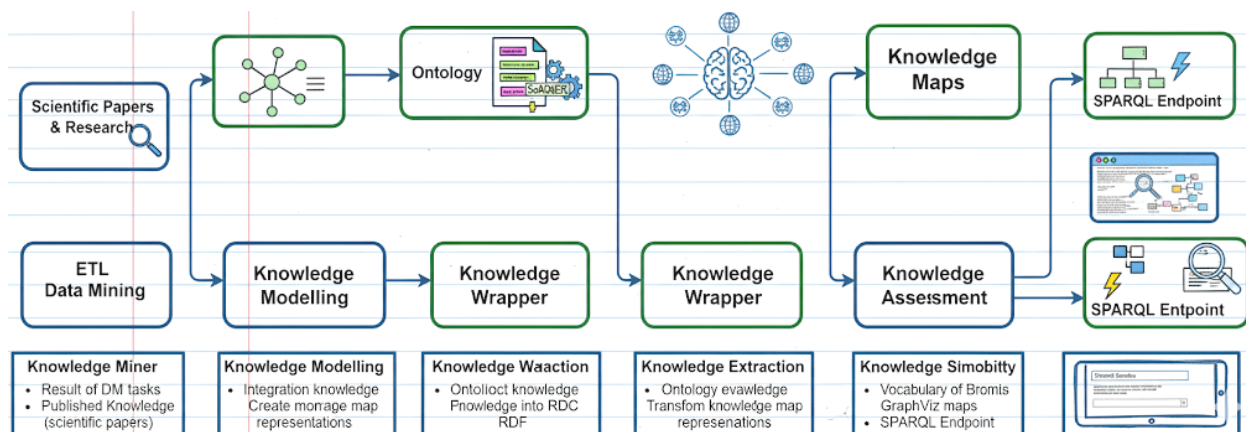


Рис. 2.6. Архітектура онтологічно-орієнтованої моделі карти знань

2.3.1 Майнінг знань

Майнінг знань є первинним компонентом, відповідальним за екстракцію знань із джерел аналізу даних. Цей компонент може бути реалізований як модуль добування даних (data mining) або використовувати наукові статті як вхідні ресурси.

Модуль добування даних охоплює інструменти та методи, що застосовуються для багатовимірного аналізу наборів даних, виявлення прихованих закономірностей та формалізації виявлених взаємозв'язків. У контексті цього дослідження ці методи включають класифікацію, кластеризацію, регресію та асоціативні правила.

Варто зазначити, що знання, отримані з модулів добування даних, як правило, зберігаються в структурованій формі, тоді як знання з наукових статей виражені у неструктурованій формі (текстовий контент). Ця

відмінність у типі знань визначає вибір відповідних технік обробки на наступному етапі (витяг знань).

2.3.2 Моделювання знань

Моделювання знань передбачає розробку структурної схеми для представлення знань, ґрунтуючись на чотирьох типах завдань добування даних. На цьому етапі також відбувається ідентифікація типу знання: фактичні знання (Fact Knowledge) або процесуальні знання (Process Knowledge).

Процес моделювання знань залежить від вхідних ресурсів

Результати завдань добування даних - якщо ресурсом є результати процесу добування даних (зберігаються в певному типі), моделювання знань вимагає схеми даних цільового набору даних для екстракції та представлення процесу знань. Ці знання переважно можуть бути представлені як карти Фактичних знань.

Якщо знання опубліковано як наукова стаття, необхідно ідентифікувати завдання добування даних (класифікація, кластеризація, регресія та асоціативне правило) для вибору відповідного типу та визначення знань. Ці знання зазвичай представляються як карти процесуальних знань.

Оскільки головна увага цього дослідження зосереджена на видобутих знаннях, даний модуль має на меті ідентифікувати тип вхідного представлення знань, яке відповідає одному з чотирьох типів:

- класифікація,
- кластеризація,
- регресія,
- асоціативне правило.

У даному контексті, елемент представлення знань охоплює наступні сутності: модель/завдання, алгоритми або системи обробки; умови, цілі, трансформації, концепції середовища, місця, набори даних, метрики.

2.4. Методологія вилучення знань

Модуль обгортки знань (Knowledge Wrapper) є центральним компонентом для трансформації знань, отриманих з модуля видобутку знань (Knowledge Mining), у формат, придатний для інтеграції в систему управління знаннями та подальшого зберігання. Цей модуль здійснює агрегацію результатів видобутку, ідентифікуючи типи завдань добування даних, застосовані алгоритми, перелік релевантних аграрних концепцій, кореляційні функції трансформації та відповідні стани. На основі цих даних генеруються представлення карт знань, які в подальшому конвертуються у формат RDF (Resource Description Framework) та імпортуються до сховища знань.

Процедура відображення отриманих знань у онтологічно-орієнтовані представлення карт знань формалізована в модулі обгортки знань. Цей модуль відповідає за перетворення видобутих знань у розширене представлення знань $k_{Extend} = (\{i\}, \{r\}, \{t\}, \{s\})$, згідно з визначенням, перед їхньою конвертацією в RDF-трійки та імпортуванням до RDF-трійкового сховища. Цей процес включає шість послідовних кроків (рисунок 2.7):

Крок 1. Вибір шаблону добування даних, який базується на конкретному алгоритмі добування даних. Це включає генерацію екземплярів добування даних (наприклад, для завдань класифікації, кластеризації, регресії та асоціативних правил) та встановлення посилань на алгоритми добування даних як об'єкти трансформації для цих екземплярів.

Крок 2. Вилучення доменних концепцій з результатів видобутку та ідентифікація трансформацій для кожної концепції. Прикладами таких концепцій є вхідні характеристики, як-от SoilPH, SeedRate, Nitrogen, Wheat, MeanYield або SpectralImages.

Крок 3. Локалізація вилучених концепцій в онтологічній структурі та генерація відповідних аграрних екземплярів для кожної локалізованої концепції. Ці екземпляри зв'язуються з екземплярами добування даних (з

Кроку 1) на основі структури завдань добування даних (класифікація, кластеризація, регресія та асоціативне правило).

Крок 4. Локалізація трансформацій даних або алгоритмів у онтологічній частині карти знань з подальшим їх зв'язуванням з аграрними екземплярами.

Крок 5. Ідентифікація станів, асоційованих з кожною концепцією в результатах видобутку, генерація відповідних станів та їх зв'язування з екземплярами.

Крок 6. Генерація представлення знань у форматі RDF-трийок та імпортування їх до RDF-трийкового сховища.

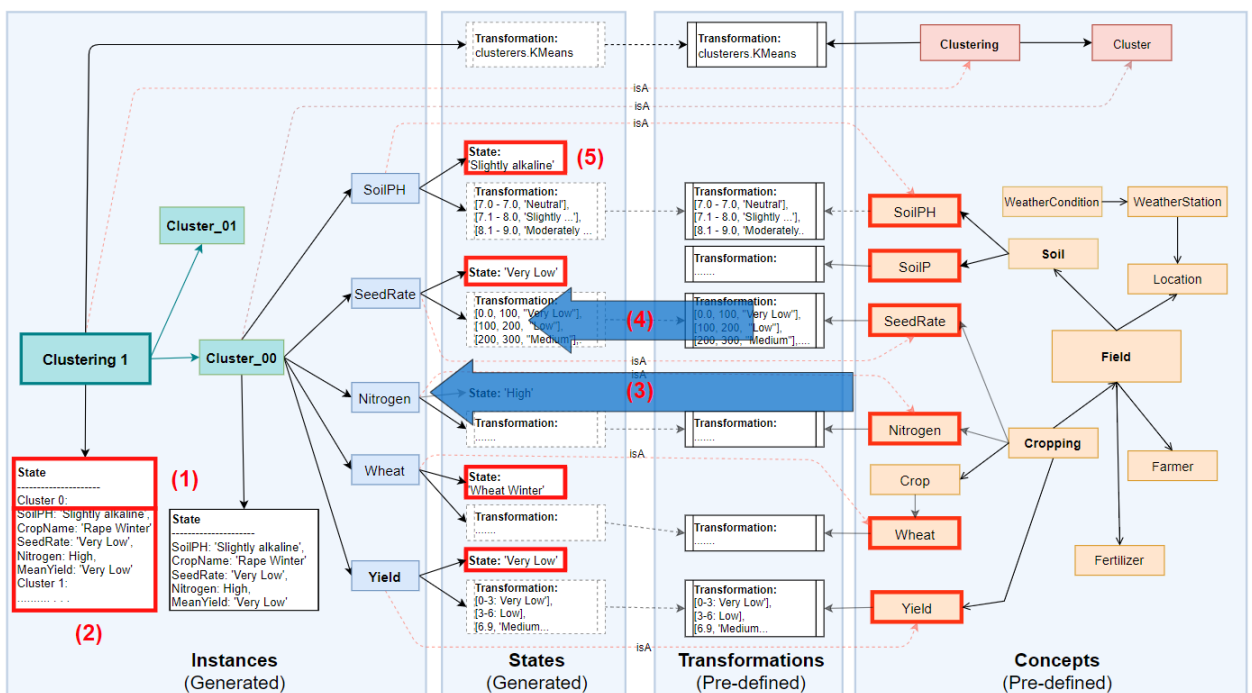


Рис. 2.7. Генерація представлення знань для завдання кластеризації

Після завершення цих шести кроків створюється представлення знань у вигляді набору RDF-трийок. Важливо зазначити, що процес локалізації концепцій або трансформацій в онтології може бути автоматизованим, напівавтоматизованим або ручним, залежно від ефективності попередніх етапів (вилучення концепцій і трансформацій з результатів видобутку) та наявності допоміжних інструментів. Різні джерела знань вимагатимуть

адапованих методів вилучення. Наприклад, для результатів модулів добування даних концепції можуть бути зіставлені за правилами, тоді як трансформації вводяться вручну. У випадку наукових статей концепції та трансформації розпізнаються з неструктурованих даних за допомогою інструментів обробки природної мови (NLP), таких як модулі розпізнавання іменованих сутностей. Цей процес є більш складним і характеризується більшим різноманіттям результатів. Наприклад, сутності "пшениця", "озима пшениця", "яра пшениця" або "гібридна пшениця" всі посилаються на концепцію "Wheat", тоді як алгоритм Long short-term memory (LSTM) може мати варіанти, такі як "Long short-term memory", "LSTM" або "штучна рекурентна нейронна мережа".

Набір екземплярів {i} формується на кроках 1 та 3, тоді як набір трансформацій {t} створюється на кроках 1 та 4. Набір станів {s} генерується на кроці 5, проте не всі представлення знань включають набори станів. Наприклад, у моделі для прогнозування врожайності вхідні значення виникають лише під час виконання моделі, тому набір станів для такого представлення знань відсутній. Нарешті, набір відношень {r} базується на відношеннях isA, hasTransformation, hasState, hasCondition та predicts.

2.4.1. Розпізнавання сутностей

Тоді як результати завдань добування даних можуть бути безпосередньо трансформовані в карту знань, знання, що містяться в наукових публікаціях, вимагають вилучення з текстових джерел. У цьому контексті модуль розпізнавання наукових сутностей (Scientific Entity Recognition) застосовується для ідентифікації та вилучення сутностей з наукових статей. Для представлення знань, отриманих з наукових публікацій, виділяється щонайменше шість основних типів сутностей:

- Завдання (Task) - одне з п'яти завдань добування даних: кластеризація, класифікація, регресія, асоціативне правило та вибір даних.

- Метод (Method) - посилається на алгоритми добування даних та підходи, що застосовуються для вирішення завдань добування даних в елементах знань. У моделі ОАК сутності, що виконують роль методів, є алгоритмами добування даних, представленими як підклас обчислювального алгоритму.

- Матеріал (Material) - відноситься до доменних концепцій у завданнях добування даних. У моделі ОАК сутності, що виконують роль матеріалів, є екземплярами типових доменних концепцій.

- Ціль (Target) - також відноситься до доменних концепцій у завданнях добування даних. У моделі ОАК сутності, що виконують роль цілей, є екземплярами типових доменних концепцій.

- Набір даних (Dataset) - стосується наборів даних, що використовуються для навчання та оцінки елементів знань.

- Метрика (Metric) - стосується метрик оцінки. У моделі ОАК сутності метрик представлені як підклас обчислювального алгоритму.

Крім того, для вилучення та представлення в моделі можуть бути розглянуті додаткові типи сутностей, такі як місце (location), особа (person), організація (organization), час (time), подія (event), продукт (product) тощо. Наразі лише місце відіграє роль контексту в представленнях знань, тоді як інші не мають прямої функціональної ролі. Однак ці теги сутностей можуть бути використані для побудови більш розширених карт знань у майбутньому.

2.4.2. Оцінка знань

Модуль оцінки знань (Knowledge Assessment) призначений для рецензування карт знань перед їх інтеграцією в модуль сховища знань. В основному, цей модуль обчислює оцінки карт знань на основі визначень процедурних або фактичних знань.

Для валідації представлень знань у цьому дослідженні інформація в кожному представленні карти знань поділяється на три категорії: базова інформація, основна інформація та підрядна інформація. Вагові коефіцієнти

для цих трьох категорій становлять 20%, 40% та 40% відповідно, як деталізовано нижче:

- Базова інформація (20%): Загальні відомості про знання, такі як автори, назва дослідження, часові рамки дослідження та джерело знань.
- Основна інформація (40%): Ключова інформація про знання, включаючи алгоритми, умови, цілі, а також трансформації для умов та цілей.
- Підрядна інформація (40%): Деталі, що стосуються наборів даних, оцінки (метрики та значення), місцеположення та контексту дослідження.

У цьому дослідженні кожне представлення знань, що містить базову інформацію, оцінюється приблизно на 50-60% для базових знань. Однак оцінка може досягати 100%, якщо представлення містить повну інформацію, що відповідає розширеним знанням. Наразі відсотки для кожної категорії базуються на відносній необхідності кожної частини для елементів знань і є суб'єктивними. Ця схема оцінки може бути скоригована для адаптації до різних джерел знань і потребує подальших досліджень.

2.4.3. Публікація знань

Система управління знаннями вимагає сервера баз даних, який підтримує RDF-трійкове сховище та протокол SPARQL для обробки запитів. Це дослідження пропонує використання нативного зберігання знань за допомогою технології Triplestore. Існують різні технології для нативного зберігання графів знань, такі як Apache Jena, Virtuoso RDF Triple Store та база даних Blazegraph. Всі ці технології графових баз даних базуються на RDF-трійковому сховищі та протоколі SPARQL.

Модуль публікації знань (Knowledge Publishing) відомий також як модуль RDF-репозиторію, модуль сховища знань або, іншими словами, репозиторій OAKMaps. Цей модуль отримує фонові знання конкретного домену з попередньо визначеної онтології, а також вилучені представлення знань з модуля обгортки знань, і потім зберігає їх у RDF-трійковому сховищі як набір RDF-трійок.

Онтологічно-орієнтована модель карти знань включає два шари знань. Перший шар складається з фонових знань, побудованих на основі попередньо визначеної онтології. Ця онтологія містить дві частини: доменні знання та обчислювальні знання. Цей шар визначає більшість доменних концепцій у картах знань та загальні відношення між ними. Цей шар також використовується для інтерпретації знань у другому шарі. Другий шар включає представлення знань, отриманих з результатів добування даних, які були виведені з наборів даних за допомогою алгоритмів добування даних (інтегрованих у модуль видобутку знань). Цей шар знань вилучається з процесу добування даних та імпортується модулем обгортки знань. Отже, перший шар знань є статичним, тоді як другий шар знань є більш динамічним.

2.4.4. Використання знань

У моделі ОАК модуль використання знань (Knowledge Exploitation) призначений для сприяння користувачам у доступі до карт знань, збережених у сховищі знань, та їх запиті різними способами. Існує кілька підходів для реалізації цього модуля, включаючи SPARQL Endpoint, браузер знань та графічні інтерфейси користувача на основі графів.

По-перше, система потребує SPARQL Endpoint для доступу та виконання SPARQL-запитів до сервера RDF-репозиторію. Цей модуль може залежати від технологій, використаних для сервера RDF-репозиторію. Наприклад, сервер з RDF Apache Jena використовуватиме Fuseki для SPARQL Endpoint, тоді як сервер Virtuoso використовуватиме Virtuoso SPARQL Query Editor. Цей модуль підтримує дослідників даних та агрономів, дозволяючи їм отримувати доступ до системи для перегляду знань, вилучення сутностей або відповіді на запитання на основі протоколу SPARQL та мови запитів RDF.

По-друге, браузер знань є зручним модулем для експлуатації знань, що забезпечує найзручніший спосіб доступу до знань. Він не вимагає від

користувачів глибоких знань про технології RDF або SPARQL і зазвичай реалізується як веб-додаток. Однак він потребує аналізатора запитів для перетворення запитів природною мовою на SPARQL. Для цієї мети доступні різні інструменти, такі як PAROT або NL2Query. Крім того, карти знань з модуля репозиторію знань можуть бути доступні за допомогою RDF-бібліотек, таких як фреймворк JENA для Java, RDFLib або SPARQLWrapper для Python.

Нарешті, програма з графічним інтерфейсом користувача на основі графів є корисним засобом експлуатації знань, таким як GraphViz або LodLive. Модуль з графічним інтерфейсом користувача на основі графів може бути розроблений як самостійна програма або як плагін, інтегрований у браузер знань для візуального представлення карт знань у вигляді графів.

У цій роботі модуль використання знань складається з двох рівнів: SPARQL Endpoint та браузера знань. SPARQL Endpoint реалізований за допомогою JENA-Fuseki, тоді як браузер знань є веб-додатком, розробленим на Python з використанням SPARQLWrapper.

2.5. Обробка знань для домену цифрової агрокультури на основі онтологічних моделей

Незважаючи на потенціал запропонованої моделі для обробки широкого спектру типів знань, дане дослідження обмежує фокус представленням знань, отриманих з результатів дата-майнінгу в цифровому сільському господарстві, оскільки це становить основну предметну область роботи. У цьому розділі деталізується застосування моделі ОАК (Ontology-based Agricultural Knowledge) для обробки видобутих знань у домені цифрового сільського господарства.

Результати огляду літератури щодо методологій створення карт знань демонструють, що побудова карт знань або управління ними в системі

управління знаннями (Knowledge Management System, KMS) традиційно включає чотири ключові етапи:

- Визначення меж та домену карти знань - для цього дослідження доменом є цифрове сільське господарство.

- Визначення ресурсів знань або матеріалів - джерелом є знання, отримані з результатів дата-майнінгу або відповідних наукових статей.

- Побудова представлень знань як карт знань - це досягається через витяг сутностей та побудову представлень знань на основі запропонованої моделі ОАК.

- Оцінка, використання та оновлення карти знань - включає оцінку представлення знань, зберігання та отримання.

Ці чотири етапи корелюють із структурними елементами моделі ОАК. Отже, при їх інтеграції з шістьма етапами моделі ОАК, процес побудови карт знань для домену цифрової агрокультури може бути декомпозиційований на три основні завдання (рис. 2.9):

1. Побудова онтології.
2. Процес підготовки знань.
3. Побудова системи управління знаннями на основі моделі.

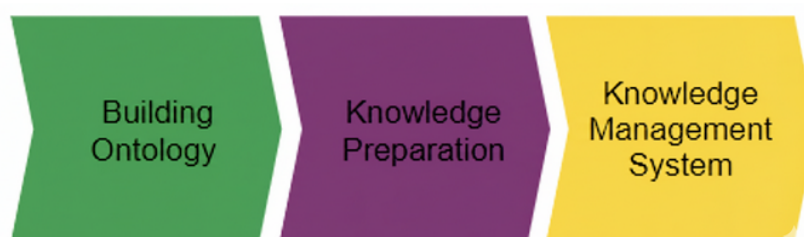


Рис. 2.9. Процес декомпозиції задач для побудови карти знань

2.5.1. Побудова онтології

Онтологія є фундаментальним компонентом моделі, що виконує подвійну функцію у побудові карт знань:

- Визначення базових концепцій у знаннях.
- Формування структури кожної карти знань.

Для виконання цих функцій онтологія поділяється на дві частини:

1. Доменна онтологія

Ця частина забезпечує базові концепції для конкретного домену, яким у цьому дослідженні є цифрове сільське господарство. Оскільки онтологія слугує для представлення знань, отриманих з даних, ці концепції охоплюють сутності, які можуть виникати в заданому наборі даних. Цей сегмент відповідає доменним концепціям (CD) та трансформаціям даних (TD). У контексті цифрового сільського господарства базові концепції включають сутності, пов'язані з культурами, добривами, хворобами, поживними речовинами або умовами вирощування (наприклад, ґрунт, погода).

2. Обчислювальна онтологія

Цей компонент підтримує представлення результатів завдань дата-майнінгу (наприклад, прогностичних регресій, класифікаційних моделей, кластеризації або відкриття шаблонів). Він відповідає обчислювальним концепціям (CDM) та обчислювальним алгоритмам (TDM). Сюди також входять концепції, що використовуються в процесах дата-майнінгу, такі як набір даних, алгоритми дата-майнінгу та метрики оцінки.

При застосуванні запропонованої моделі для обробки видобутих знань у цифровому сільському господарстві це завдання конкретизується як побудова онтології сільськогосподарських обчислень.

2.5.2. Підготовка знань

Цей етап охоплює вибір ресурсу знань та його подальшу обробку з метою витягування карт знань та їх трансформації у репозиторій знань.

Зважаючи на природу видобутих знань у цифровому сільському господарстві, дане дослідження обирає наукові статті як основний ресурс знань для застосування до моделі ОАК. Цей вибір призводить до трьох послідовних підзавдань:

- Витяг сутностей

Це завдання розпізнавання іменованих сутностей (Named Entity Recognition, NER), що є критичним елементом для пошуку інформації в обробці природної мови (Natural Language Processing, NLP). У контексті сільськогосподарських наукових статей це завдання називається Науковим сільськогосподарським розпізнаванням іменованих сутностей (SciAGNER). Модуль SciAGNER використовується для ідентифікації сільськогосподарських сутностей та сутностей дата-майнінгу (таких як завдання, алгоритми, компоненти, набори даних, метрики оцінки та результати).

Витяг відносин

Використовується для визначення відносин між витягнутими сутностями з подальшою побудовою карт знань для моделі ОАК. До цих відносин належать: `hasLocation`, `hasAlgorithm`, `hasCondition`, `predicts`, `hasDataset`, `hasEvaluation`, `hasTransformation` та `hasState`. Ці відносини витягуються та генеруються напівавтоматично за допомогою модуля Обголонки знань (Knowledge Wrapper).

Для витягування сутностей у процесі підготовки знань для цифрового сільського господарства, це дослідження пропонує семантично-орієнтований підхід до завдання NER.

2.5.3. Система управління знаннями

Проектування системи управління знаннями базується на багат шаровій архітектурі (рис. 2.10), що розділяє потоки обробки знань.

Шар Отримання знань (Knowledge Acquisition Layer) - відповідає за споживання знань, використовуючи широкий спектр плагінів-обголонок для перетворення видобутих знань у формат, придатний для наступного шару.

Шар Карт знань (Knowledge Map Layer) - відповідає за зберігання та індексування даних, використовуючи модель RDF (Resource Description Framework). Цей шар надає інтерфейси Доступу до даних (Data Access) для обробки запитів SPARQL.

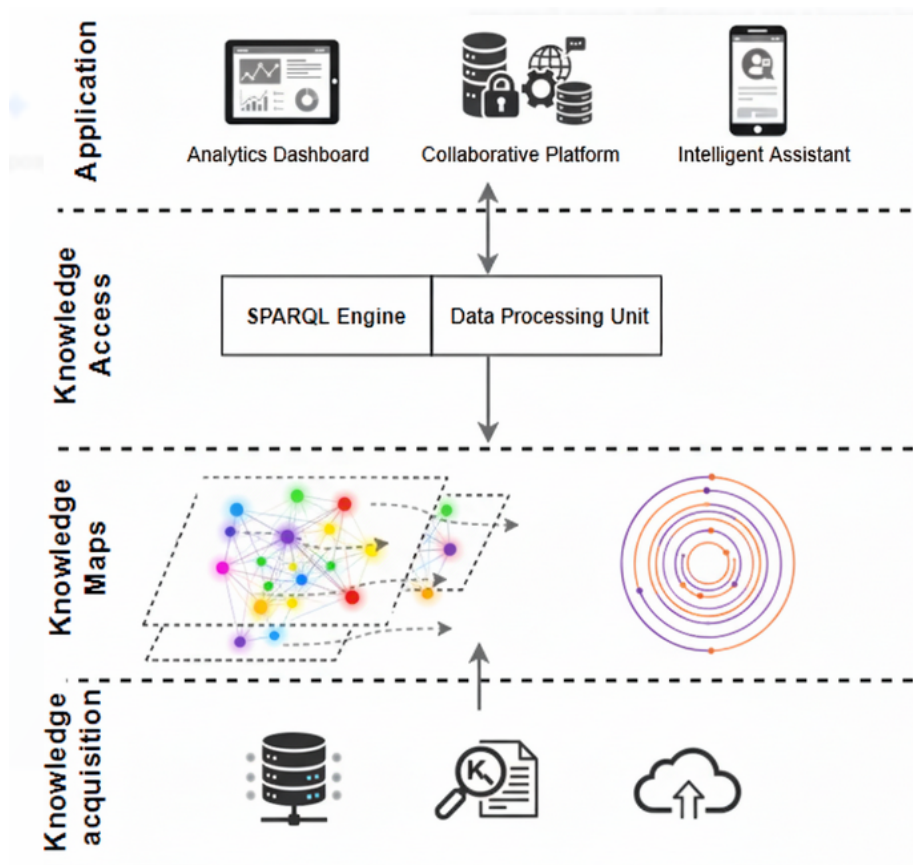


Рис. 2.10. Багатошарова архітектура моделі

Шар додатків (Application Layer) - тут Двигун SPARQL дозволяє розробникам програм запитувати дані через SPARQL Endpoint (для одноразових запитів, використовуючи розширення мови запитів SPARQL 1.1) або SPARQL-додаток.

У даній роботі онтологія сільськогосподарських обчислень (O) та представлення знань (K) є представленнями доменних карт знань, доступних для обробки комп'ютерами. Це технічне питання розглядається через технології зберігання знань та обробки запитів знань.

Висновки до розділу

У цьому розділі було представлено архітектуру онтологічно-орієнтованої моделі карти знань (ОАК). Ця модель структурована навколо шести ключових модулів (зокрема, майнінг знань, моделювання знань, витяг

знань, оцінка знань, публікація знань та використання знань), призначених для представлення, зберігання та отримання знань, отриманих із завдань дата-майнінгу.

Для успішного застосування моделі ОАК та її імплементації у формі програмного забезпечення в конкретній предметній галузі необхідне виконання мінімум трьох основних етапів:

- Побудова онтології.
- Підготовка знань.
- Створення системи управління знаннями.

Застосування запропонованої моделі до обробки знань у домені цифрової агрокультури також вимагає виконання цих трьох ключових завдань: побудова онтології для домену, підготовка знань, створення системи управління знаннями.

РОЗДІЛ 3. ПОБУДОВА ОНТОЛОГІЧНИХ МОДЕЛЕЙ КАРТ ВИДОБУТИХ ЗНАНЬ В ДОМЕНІ ЦИФРОВОЇ АГРОКУЛЬТУРИ

3.1. Обґрунтування розробки онтології

Попередній аналіз виявив існування низки онтологій у домені цифрової агрокультури (наприклад, AGROVOC, що надає сільськогосподарські словники; AgOnt, орієнтована на IoT; Citrus Ontology; Plant Ontology (PO), що описує анатомію рослин). Однак, кожна з них має специфічну мету і не призначена для комплексної обробки знань у цифровому сільському господарстві. Отже, виникає необхідність у створенні нової онтології, яка інтегрує загальні сільськогосподарські концепції та їх взаємозв'язки з обчислювальними аспектами. Крім того, для ефективної обробки та представлення результатів досліджень дата-майнінгу критично важливою є обчислювальна частина онтології. Існуючі онтології, такі як OntoDM, пропонують структуру для завдань дата-майнінгу, алгоритмів та компонентів попередньої обробки даних, включаючи 856 класів і 480 екземплярів. Хоча OntoDM є міцним фундаментом через широкий спектр охоплених концепцій, йому бракує специфічних екземплярів та концепцій, необхідних для точного представлення моделей знань, які зберігаються в рамках моделі ОАК.

Згідно з архітектурою, описаною в попередньому розділі, запропонована онтологія повинна складатися з доменної онтології (для домену цифрової агрокультури) та обчислювальної онтології. Таким чином, у цьому розділі представлено процес побудови онтології сільськогосподарських обчислень, призначеної для використання у запропонованій моделі ОАК.

3.1.1. Вимоги до онтології

В інформатиці онтологія визначається як формальне представлення доменних знань через набір концепцій, об'єктів та відносин між ними.

Створення онтологій здійснюється за допомогою спеціалізованих інструментів (наприклад, Protégé, SmartTools Ontology Editor, TopBraid Composer), причому Protégé є найбільш поширеним у наукових колах завдяки його статусу безкоштовного візуального редактора з відкритим вихідним кодом. Результати зберігаються у форматах OWL, RDF або XML, що забезпечує комп'ютерну доступність та оброблюваність доменних знань.

Для коректної побудови та концептуалізації онтології необхідно чітко визначити вимоги, що включають:

1. Межі онтології.

Виходячи з обмежень та опису побудови карт знань для цифрового сільського господарства, мета полягає у створенні онтології, здатної представляти елементи видобутих знань у домені цифрового сільського господарства.

2. Сценарії онтології.

Онтологія матиме дві частини: сільськогосподарську та обчислювальну. Через широкий спектр завдань дата-майнінгу, ця версія зосередиться виключно на представленні елементів знань, отриманих з чотирьох поширених завдань дата-майнінгу: класифікація, кластеризація, регресія та асоціативні правила.

3. Компетентність онтології. Для забезпечення компетентності формується набір питань компетентності. Онтологія повинна бути здатною відповідати на ці питання за допомогою своїх аксіом.

3.1.2. Процес побудови онтології

Побудова онтології базується на загальноприйнятих методологіях розробки онтологій і включає наступні послідовні етапи:

1. Визначення меж та області застосування.

Визначаються компоненти онтології - сільськогосподарські концепції та концепції дата-майнінгу, необхідні для представлення знань у цифровому

сільському господарстві. Область застосування охоплює цифрове сільське господарство та вищезазначені чотири завдання дата-майнінгу.

2. Збір та аналіз термінології.

Здійснюється збір та аналіз термінів з сільськогосподарських та дата-майнінгових джерел, включаючи статті, словники та існуючі онтології (AGROVOC, OntoDM, Plant Ontology, Agronomy Ontology).

3. Концептуалізація.

Визначаються основні концепції та відносини. Створюється ієрархія класів, а також властивості та обмеження для кожного класу. Основні концепції поділяються на:

- Сільськогосподарські концепції: Культури, ґрунт, погода, хвороби, добрива тощо.
- Концепції дата-майнінгу: Алгоритми, завдання, набори даних, оцінка тощо.
- Формалізація: Концептуальна модель перетворюється на формальну онтологію з використанням мови OWL (Web Ontology Language), де чітко визначаються класи, властивості та обмеження.

4. Реалізація.

Онтологія реалізується за допомогою інструментів (наприклад, Protégé) та зберігається у форматі OWL або RDF, забезпечуючи її готовність до інтеграції в системи управління знаннями.

3.1.3. Принципи проектування онтології

Більшість онтологій зазвичай описують класи (концепції), екземпляри та відношення, а також можуть включати обмеження, правила, аксіоми та функціональні терміни. Як формальне представлення Карт знань (KMaps), онтологія запропонована з наступними основними компонентами:

- Концепції. Охоплюють концепції в сільському господарстві та концепції, необхідні для представлення чотирьох основних завдань дата-майнінгу. Приклади сільськогосподарських концепцій включають поле,

фермер, культура, організація, місцезнаходження, продукт тощо, тоді як концепції дата-майнінгу включають кластеризацію, класифікацію, регресію та асоціативні правила.

- Трансформації. Екземпляри або об'єкти, що представляють атомний або базовий рівень онтології.

- Відношення. Способи, якими концепції (а згодом і екземпляри) можуть бути пов'язані між собою.

Побудова сільськогосподарської онтології загалом складається з двох етапів:

1) створення основної сільськогосподарської онтології на базі існуючих баз знань у сільському господарстві або дата-майнінгу,

2) збагачення сільськогосподарської онтології шляхом видобутку сільськогосподарських сутностей з наявних ресурсів.

З огляду на вимоги до бази знань для додатків цифрового сільського господарства та моделі, що базується на знаннях для дата-майнінгу в сільському господарстві, онтологія має важливе значення. Ця онтологія побудована з горизонтальною описовою структурою, що включає три шари:

1. Шар специфікації.

Є семантичним каркасом онтології, який використовується для представлення концепцій у сільському господарстві та дата-майнінгу.

2. Шар імплементації.

Включає необхідні екземпляри, які використовуються для пошуку та порівняння різних імплементацій сільськогосподарських концепцій та алгоритмів дата-майнінгу.

3. Шар додатків.

Включає необхідні трансформації, які використовуються для представлення процесів обробки сільськогосподарських атрибутів та виконання алгоритмів дата-майнінгу.

Побудова онтології здійснюється пошарово. Шар специфікації є основною онтологією і детально описується в цьому розділі. Шар

імплементатії розширюється лексикою з інших ресурсів,. Нарешті, шар додатків базується на створенні трансформацій для онтології.

У цьому дослідженні доменна частина онтології містить чотири піддомени (рис. 3.1):

1. Сільське господарство
2. IoT (Інтернет речей)
3. Географічний
4. Бізнес під-домен

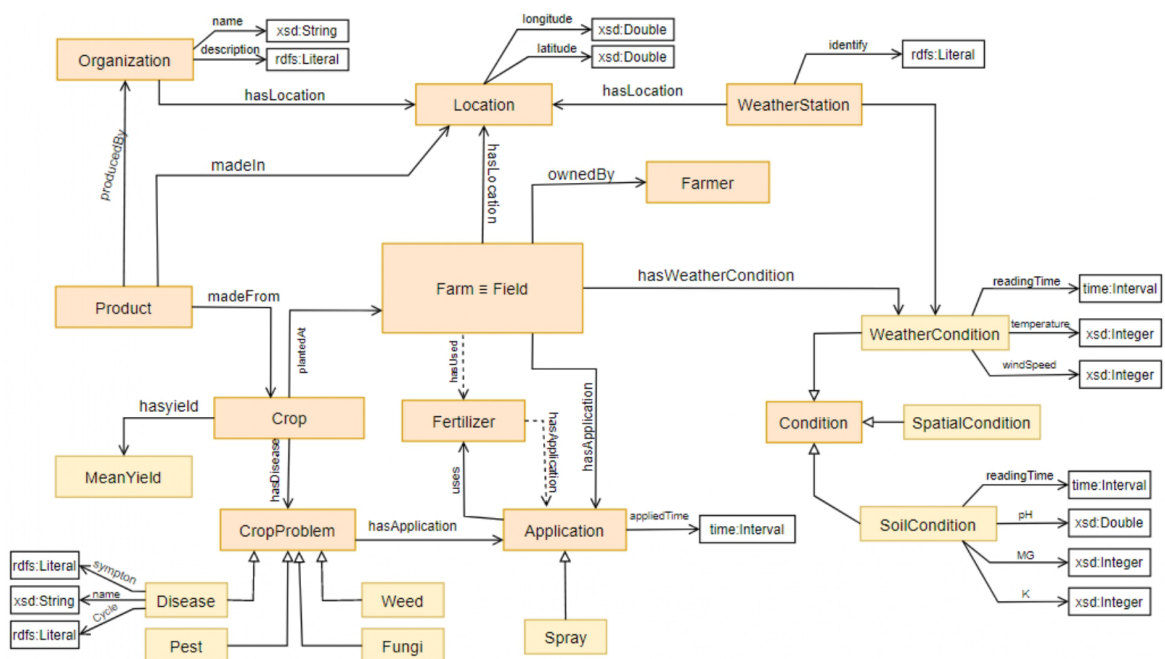


Рис. 3.1. Основні концепції побудови онтології для цифрової агрокультури

Ці чотири піддомени дозволяють онтології описувати базові концепції в сільськогосподарському домені та пов'язаних тематичних піддоменах.

3.2. Процес розробки та структура онтології

Процес побудови онтології включає три основні етапи:

1. Побудова доменно-специфічної ієрархії знань.

2. Визначення слотів категорій та представлення аксіом.

3. Набуття знань (Knowledge acquisition) шляхом заповнення значень слотів екземплярів із використанням існуючих баз знань.

Дотримуючись цієї послідовності, онтологія була побудована таким чином: Етапи (1) і (2) були сфокусовані на створенні основних концепцій та відношень у кожному піддомені як для сільськогосподарської, так і для обчислювальної частин онтології. Ці концепції успадковуються від загального класу Entity та його двох підкласів: VirtualEntity та PhysicalEntity. Етап (3) зосереджувався на збагаченні онтології екземплярами (переважно сільськогосподарськими) та трансформаціями для обробки даних (які використовуються переважно в дата-майнінгу). Цей прогрес забезпечив відповідність онтології її тришаровій структурі (шари специфікації, імплементації та додатків), про яку йшлося раніше.

3.2.1. Специфікація концептуального ядра

Після побудови ієрархії знань, онтологія надає не лише загальний огляд домену сільського господарства, але й детально описує ключові сільськогосподарські концепції, такі як Crop (Культура), Disease (Хвороба), Farm (Ферма), Product (Продукт) тощо. Вона також встановлює відношення між сільськогосподарськими концепціями та пов'язаними з ними концепціями, такими як погодні умови, стан ґрунту, добрива, характеристики ферми. Наприклад, рис. 3.1 ілюструє відношення між Crop та Farm, між Farm та Location чи WeatherCondition, або між Crop та Disease. Ці відношення є основою для подальших міркувань (reasoning), якщо це необхідно.

Крім сільськогосподарських, до онтології додано концепції з домену дата-майнінгу (рис. 3.2). Клас DataMining охоплює чотири поширені завдання (Classification, Clustering, Regression, Association Rule) та один клас для розширення моделі (Data Selection). Крім того, онтологія містить класи Algorithm, Dataset, Evaluation, Evaluation-Metric для представлення різних типів екземплярів у моделі ОАК.

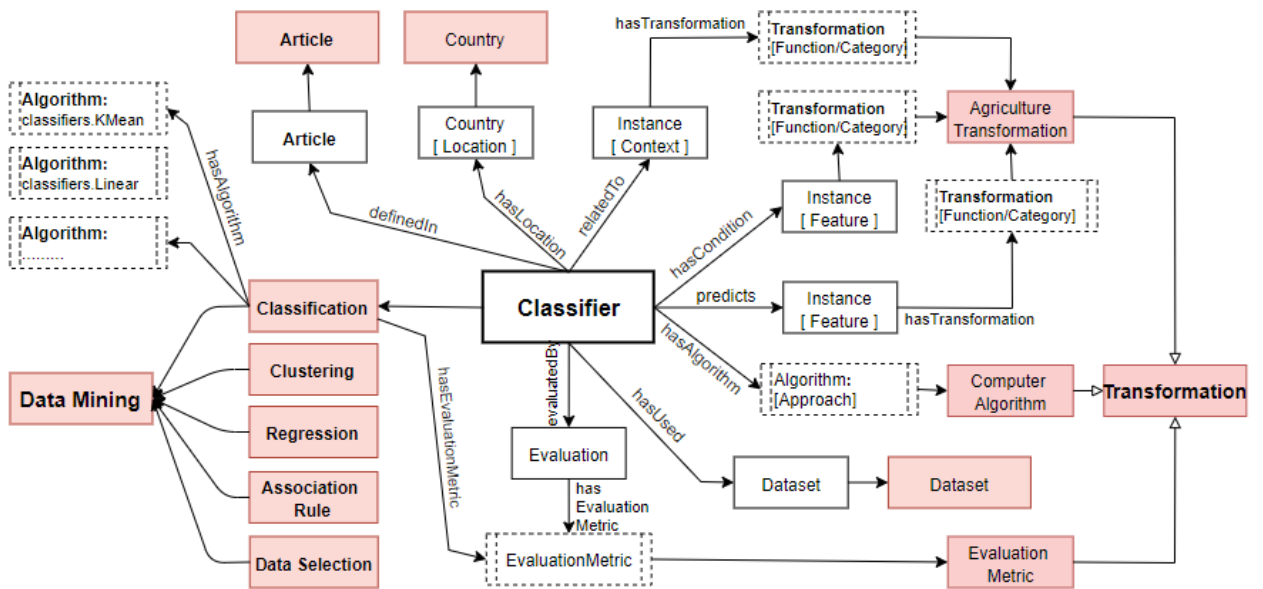


Рис. 3.2. Основні концепції інтелектуального аналізу даних в онтології

Ці концепції та їхні відношення формують каркас знань, необхідний для трансформації видобутих знань із завдань дата-майнінгу у представлення знань та їх імпорту до карт знань (KMaps). До цих відношень належать *hasAlgorithm*, *hasCondition*, *predicts*, *hasDataset*, *hasLocation*, *evaluationBy* тощо (рис. 3.2).

3.2.2. Технічні характеристики та інтеграція

Основне ядро онтології було побудовано за допомогою інструменту Protégé (рис. 3.3). Онтологія включає класи та аксіоми, пов'язані із сільським господарством та обчисленнями. Онтологія надає загальний огляд сільськогосподарського домену з найбільш загальними концепціями та їхніми зв'язками, включаючи погодні умови, стан ґрунту, добрива та опис ферм. Онтологія має чотири основні тематичні піддомени сільськогосподарської частини, чия архітектура спрямована на представлення концепцій, які можуть зустрічатися в сільськогосподарських наборах даних.

Інтеграція сільськогосподарських концепцій з концепціями дата-майнінгу (наприклад, класифікація, кластеризація, регресія) дозволяє представляти видобуті знання. Наприклад, концепції регресія, рН ґрунту,

назва культури та відповідні функції трансформації використовуються для представлення моделі знань для прогнозування рН ґрунту на основі найближчих полів.

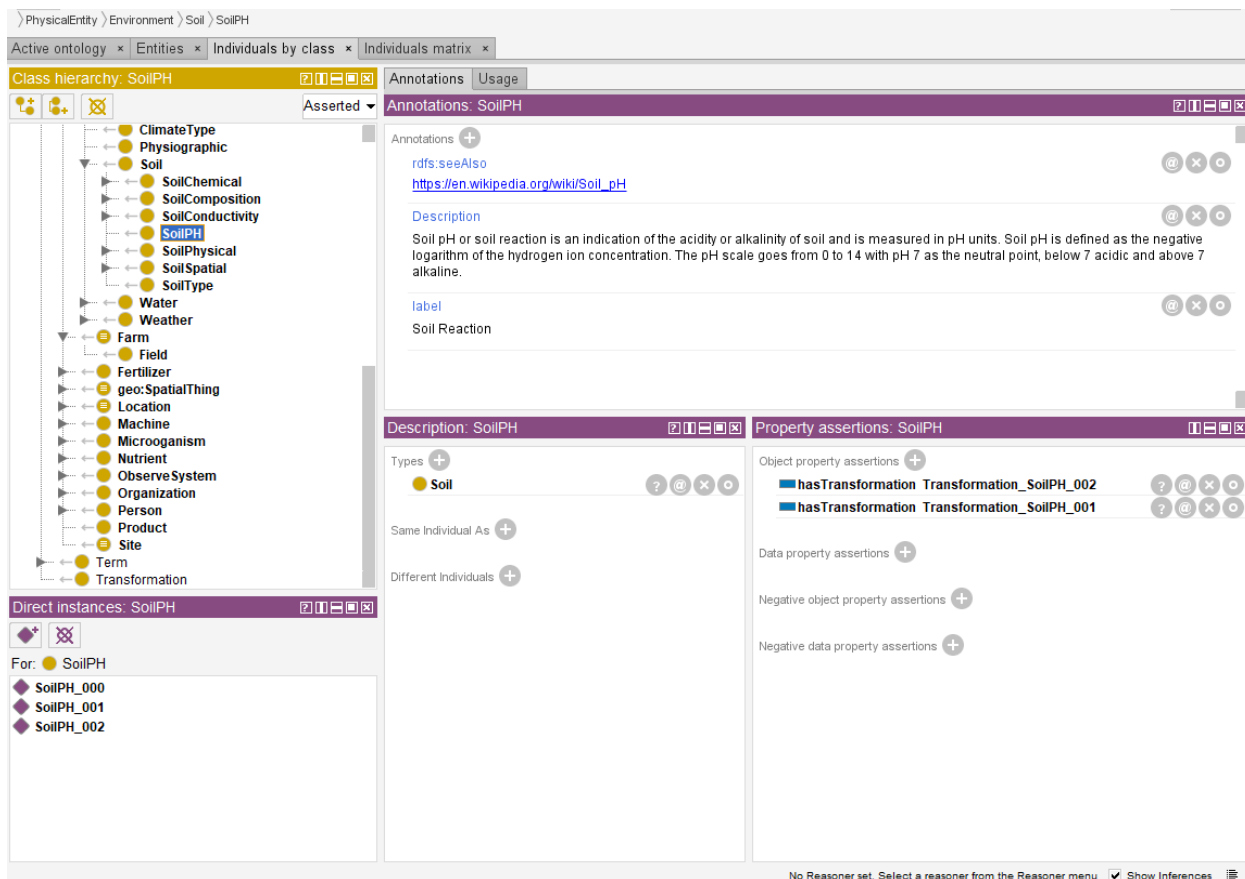


Рис. 3.3. Вигляд онтології реалізований засобами Protege

Отже, дана онтологія є ядерною онтологією для побудови карт знань у сільському господарстві. Її ієрархічна структура сприяє інтеграції доступних ресурсів для створення більш широких та точних карт знань.

3.2.3. Принципи дизайну онтології

Для забезпечення прозорості та пояснюваності онтології, а отже, і моделі ОАК, кожен концепт та відношення в онтології формально визначаються із залученням стандартних атрибутів метаданих.

URI (Universal Resource Identifier) - Унікальний ідентифікатор ресурсу.

rdfs:label - Назва концептів або екземплярів (титул).

`rdfs:comment` - Опис концептів, екземплярів, відношень або трансформацій для цілей пояснення.

`dc:identifier` - Формула, вираз або функція, необхідна для обчислення та трансформації даних, якщо застосовується.

`rdfs:isDefinedBy` - Посилання на джерела або творців концептів.

`rdfs:seeAlso` - Зовнішні посилання для отримання додаткової інформації щодо кожного концепту.

Ці атрибути є базовою інформацією для кожної сутності в онтології та критично важливі для досягнення зрозумілого дизайну.

Для збагачення шару імплементації та забезпечення широкої доменної охоплюваності онтологія інтегрує термінологію з відомих зовнішніх джерел. AGROVOC є найбільш відомим сільськогосподарським лексиконом, що містить понад 36,000 концептів на 33 мовах. AGROVOC доступний у дистрибутивах Core та LOD (Linked Open Data) у форматі RDF/XML. Ця база знань була обрана як найбільш придатний еталон для узгодження та оцінки нової онтології.

Крім AGROVOC, інші існуючі онтології та популярні довідкові ресурси, такі як Plantix та PlantVillage, були використані як цінні матеріали для вилучення термінології. Географічні класи в пропонуваній онтології охоплюють два основні адміністративні рівні, а також загальні просторові сутності: Country (Країна), Subcountry (Субкраїна), `geo:SpatialThing` та `geo:Point`.

Відношення включають: довготу/широту, населення, властивості площі, відношення частина-ціле між географічними екземплярами, а також `hasLocation`, `isLocationOf` та `isProducedAt` між географічними та іншими екземплярами (рослини, продукти). Інші географічні екземпляри мають базові відношення (довгота, широта, адреса, поштовий індекс), а також відношення `hasCountry` та `hasSubCountry` для вказівки їхньої приналежності до адміністративних одиниць.

3.3. Способи обробки і перетворення даних в онтології

Трансформації відіграють ключову роль як в онтології, так і в моделі ОАК. Вони забезпечують представлення способів попередньої обробки та перетворення даних у процесі дата-майнінгу.

Трансформації в онтології виконують три основні функції:

- Пояснення процесу. Деталізують методи обробки та перетворення даних на етапах дата-майнінгу.
- Представлення результатів. Формалізують видобуті знання, отримані в результаті процесу дата-майнінгу.
- Підтримка. Надають спеціалістам з даних додаткові опції для трансформації атрибутів даних.

Як подано рис. 3.4, трансформації включають трансформації даних та обчислювальні алгоритми.

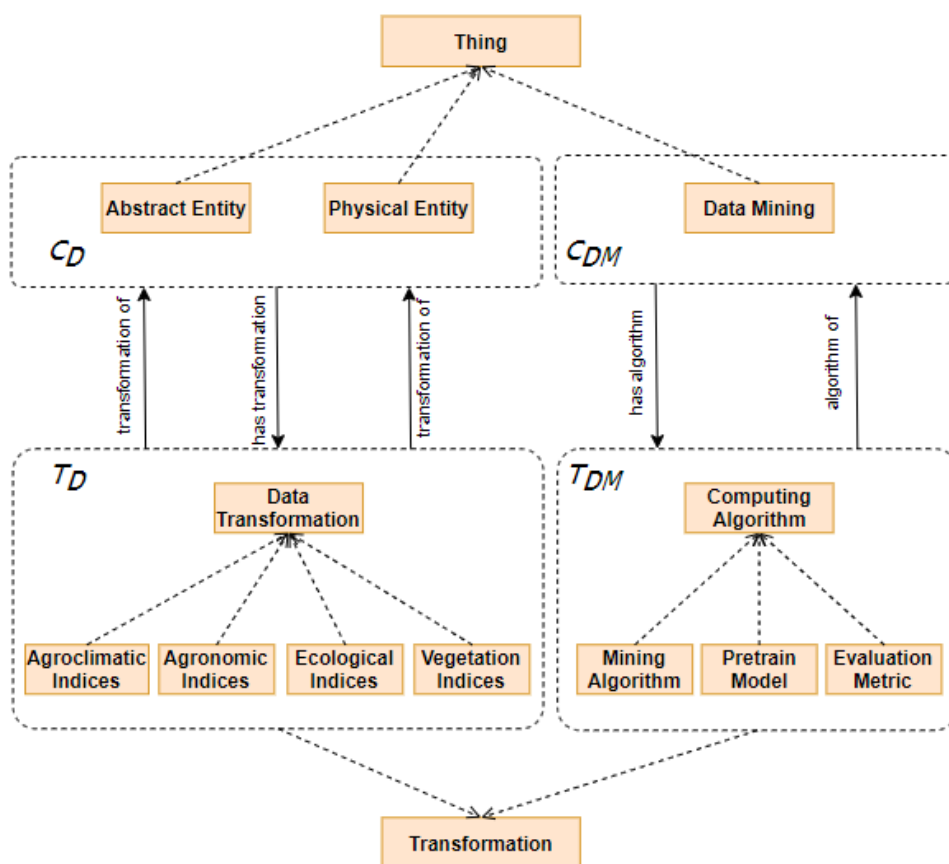


Рис. 3.4. Представлення архітектури онтології

Кожна трансформація або індекс у класі Transformation та його підкласах має наступні атрибути:

- rdfs:label: Назва трансформації або індексу.
- rdfs:comment: Опис трансформації або індексу.
- rdfs:acronym: Аббревіатура назви трансформації/індексу.
- dc:identifier: Формула або функція для обчислення та перетворення даних.
- rdfs:isDefinedBy: Джерела або автори трансформації.
- rdfs:reference: Документ для посилання на оригінальну інформацію.

У сільському господарстві та суміжних доменах (метеорологічному, агрокліматичному, екологічному) численні індекси використовуються для спостереження та прогнозування стану навколишнього середовища. Індекси слугують ефективним інструментом управління та комунікації, оскільки вони згортають множинні показники в єдине значення, хоча їхня простота може приховувати складність екосистеми. Вони використовуються для характеристики системних атрибутів, таких як фізичні драйвери екосистемних процесів.

В онтології індекси класифікуються як трансформації і включають: агрокліматичні індекси, вегетаційні індекси та агрономічні індекси (рис. 3.5).

Індекси	Опис/Приклади
Агрономічні Індекси	Індекси ґрунту, бур'янів, поживних речовин, води
Агрокліматичні Індекси	GDD, AFD, CDD, FX
Вегетаційні Індекси	LAI, NDVI, GDVI, RVI, SAVI, SAVO
Екологічні Індекси	Індекси води: WZI, WFI
Попередньо навчені Моделі	GoogLeNet, AlexNet
Алгоритми Майнінгу	Linear, DT, ANN, LSTM, LASSO
Метрики Оцінки	Accuracy, CCR, Jscore, F1, R2, RSME
Трансформації Даних	RGB або BW для кольору, Типи SoilPH

Рис. 3.5. Перелік індексів та їх описів

Важливою складовою трансформацій є обчислювальні алгоритми. Ці алгоритми використовуються для виконання завдань дата-майнінгу, включаючи класифікацію, кластеризацію, регресію та асоціативні правила.

Збір алгоритмів здійснювався вручну з різних джерел, таких як програмні бібліотеки (Scikit-learn, NLTK, Huggingface), та наукові статті. Початкові списки алгоритмів були отримані з програмних бібліотек. Потім для кожного алгоритму була зібрана необхідна інформація з різних ресурсів, та програмні бібліотеки, згідно зі структурою базової інформації. До онтології також додано інформацію про загальновідомі алгоритми, метрики оцінки для завдань дата-майнінгу та поширені попередньо навчені моделі, зібрані аналогічним процесом.

3.4. Розпізнавання сутностей та підготовка знань

Своєчасне та економічно ефективно оновлення даних має вирішальне значення для розвитку точного сільського господарства. У цьому контексті дані традиційно охоплюють цифрові формати (дистанційне зондування, карти ґрунту/врожайності, метеорологічні дані). Однак, значний обсяг даних представлений у вигляді текстових джерел (сільськогосподарські новини, бізнес-новини, політичні документи, мас-медіа). Ці текстові джерела можуть розглядатися як соціальні сенсори, які також використовуються для актуалізації інформації. Ефективний аналіз текстових даних не лише забезпечує актуальність інформації, але й підтримує експертів у прийнятті рішень в управлінні точним сільським господарством. У таких системах обробки новин тегування сутностей є критично важливим для вилучення географічних осередків, новинних подій або кластерів історій.

У сфері обробки природної мови (NLP) розпізнавання іменованих сутностей (NER) спрямоване на класифікацію слів у документі за попередньо визначеними цільовими класами сутностей. NER вважається фундаментальною задачею для багатьох завдань NLP та пошуку інформації

(IR). Окрім застосування у IR, NER підтримує процес побудови онтологій. Як показано на рисунку 3.5, NER є базовим етапом у модулях вилучення екземплярів та фактів у процесі онтологічного вилучення інформації, а також у створенні карт знань чи репозиторіїв знань.

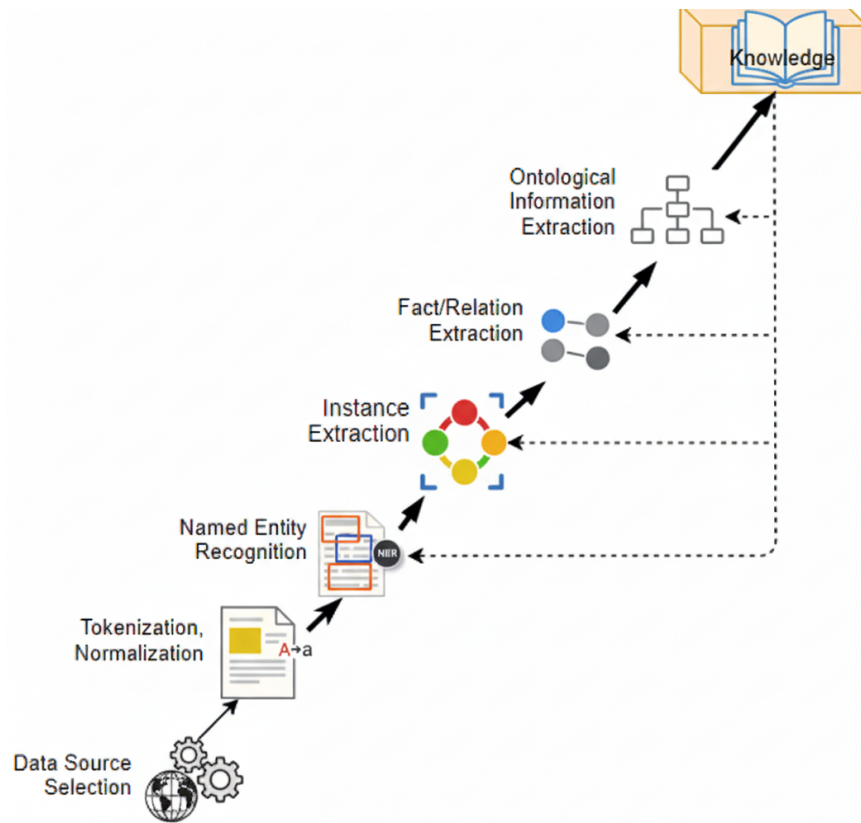


Рис. 3.5. Процес розпізнавання сутностей в інформації

Неструктуровані дані зазвичай зберігаються як сирий текст, що ускладнює доступ та аналіз цінної інформації. Отже, NER є необхідним інструментом для ідентифікації значущих сутностей, роблячи їх придатними для аналітичних додатків. Хоча спочатку розпізнавалися загальні іменовані сутності (місцезнаходження, особи, організації), кількість тегів у наборах NER зростає, щоб надати більш специфічні значення.

Загалом, іменовані сутності є власними назвами і можуть бути виявлені за орфографічними ознаками, ознаками на рівні слова та географічними довідниками. Техніки аналізу даних використовуються для класифікації та усунення неоднозначності. Однак для завдання Розпізнавання

сільськогосподарських сутностей (AGER) ці ознаки мало корисні, оскільки загальні сільськогосподарські сутності не є власними назвами і не мають типових орфографічних ознак, що створює проблему для систем розпізнавання.

Для розробки компонента AGER пропонується процес NER для сільськогосподарського домену, що включає:

- 1) розробку нового набору тегів для сільськогосподарських сутностей;
- 2) використання семантично-орієнтованого підходу AGER для створення анотованого корпусу AGER;
- 3) використання глибоких нейронних мереж (DNN) для побудови модуля AGER, навченого на цьому корпусі.

Семантично-орієнтований підхід AGER поєднує словниковий підхід та семантичні класи, дозволяючи розпізнавати сутності без навчання на анотованих корпусах.

Вилучення іменованих сутностей у сільськогосподарському домені має кілька застосувань, таких як вилучення конкретної інформації з мас-медіа та відстеження появи сільськогосподарських сутностей у часі та просторі. Прикладом є системи виявлення та моніторингу спалахів хвороб на основі мас-медіа, де NER є обов'язковим модулем для ідентифікації сутностей (місцезнаходження, організації, події, хвороби).

3.5. Архітектура системи розпізнавання сільськогосподарських сутностей

Це дослідження сфокусоване на тегуванні іменникових фраз, які класифікуються як одна із сільськогосподарських категорій, наприклад, [великий кролик]ANI (тварина), [озима пшениця]CRP (культура) або [снігова буря]DST (катастрофа).

При розпізнаванні сутностей у вхідних реченнях виникає проблема неоднозначності. Розглянемо приклад: "Площа пшениці швидко

збільшується у 2018 році." (The wheat area is increasing quickly in 2018.). Іменникову фразу [The wheat area]NP можна інтерпретувати трьома способами:

Випадок 1: [The wheat area]CRP — уся іменникова фраза розпізнається як сутність CROP (Культура).

Випадок 2: The [wheat]CRP area — лише "wheat" розпізнається як сутність CROP.

Випадок 3: The wheat area — жодна сутність не розпізнається.

У Випадку 1 фраза "wheat area" (площа пшениці) більше стосується інформації про площу, ніж про культуру. У Випадку 2 "wheat" виступає модифікатором слова "area". У Випадку 3 жодна сутність не розпізнається, оскільки "area" не належить до визначених тегів. У рамках цього підходу ми обрали Випадок 3 для подібних ситуацій, оскільки кінцевою метою є розпізнавання тегованих сутностей, які мають цілісне значення об'єктів, класифікованих за однією із сільськогосподарських міток.

Традиційні підходи до NER покладаються на орфографічні ознаки, ознаки на рівні слова та географічні довідники. Однак ці ознаки ефективні для власних іменованих сутностей. Оскільки сільськогосподарські сутності є загальними сутностями у специфічному домені, орфографічні ознаки стають неефективними, що підкреслює необхідність застосування семантичних підходів для доменного розпізнавання сутностей, такого як AGER

Запропонований підхід до AGER має дворівневу архітектуру (рисунок 3.6):

Етап 1: Семантично-орієнтований підхід та створення корпусу

На першому етапі розроблено семантично-орієнтований підхід AGER для автоматичного анотування сільськогосподарських сутностей та створення початкового корпусу AGER. Цей підхід використовує три семантичні ресурси для ідентифікації сутностей та пошуку відповідників у сирому корпусі. Отриманий корпус підлягає ручному рецензуванню та корекції анотаторами для створення фінального корпусу AGER.

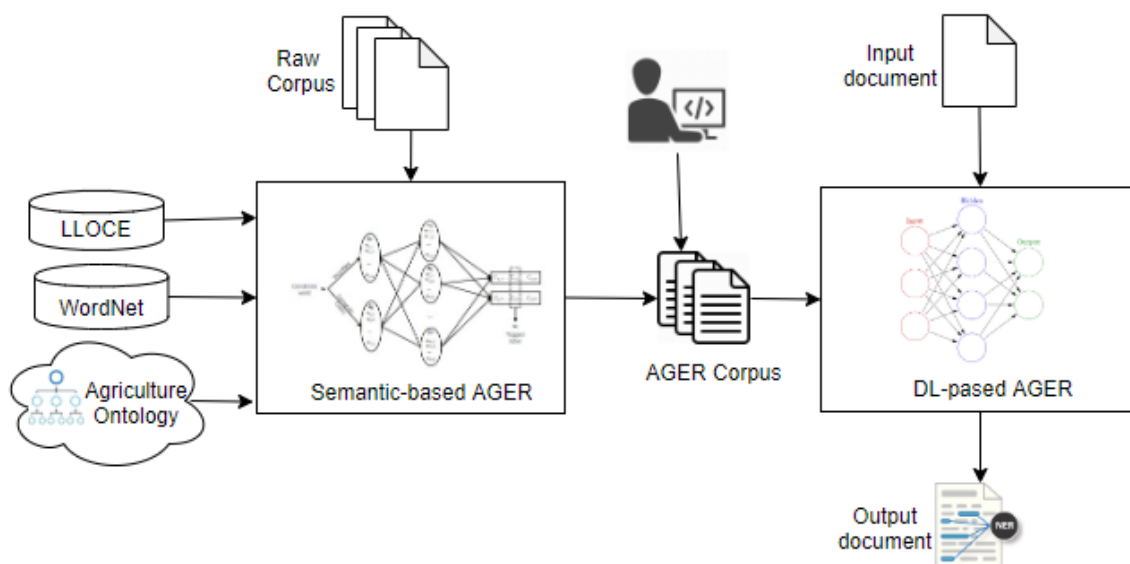


Рис. 3.6. Архітектура системи

Ресурси: Географічний довідник (gazetteer) створюється на основі сільськогосподарської онтології для формування унітарних наборів сільськогосподарських тегів. Два інші моделі класифікації ґрунтуються на WordNet та Семантичних Класах. Географічний довідник слугує простим видом бази знань для довідкового пошуку.

Етап 2: Модель глибокого навчання

На другому етапі застосовуються архітектури BiLSTM-CNN та BiLSTM-CRF для класифікації вхідних токенів за однією з 12 міток сільськогосподарських сутностей або міткою NONE.

Для забезпечення семантичної основи AGER використовуються наступні ресурси.

WordNet є великою лексичною базою даних англійської мови. Іменники, дієслова, прикметники та прислівники згруповані в набори когнітивних синонімів (синсети). Англійська версія містить 117,000 синсетів. WordNet використовується для вимірювання семантичних відстаней між словами-кандидатами та словами в географічному довіднику кожного класу.

Як другий семантичний ресурс використовуються словники LDOCE (Longman Dictionary of Contemporary English) та LLOCE (Longman Lexicon of Contemporary English).

LDOCE організований як словник із ієрархічними семантичними класами (32 основні коди). LLOCE, менший словник, організований семантично: 16,000 слів у 2,441 наборі. Його ієрархічна структура має три рівні. Наприклад, Клас А ("Життя та живі істоти") включає Групу А50-А61 ("Тварини/Ссавці"), а Набір А53 містить "Кішка та подібні тварини" (Рис. 3.7).

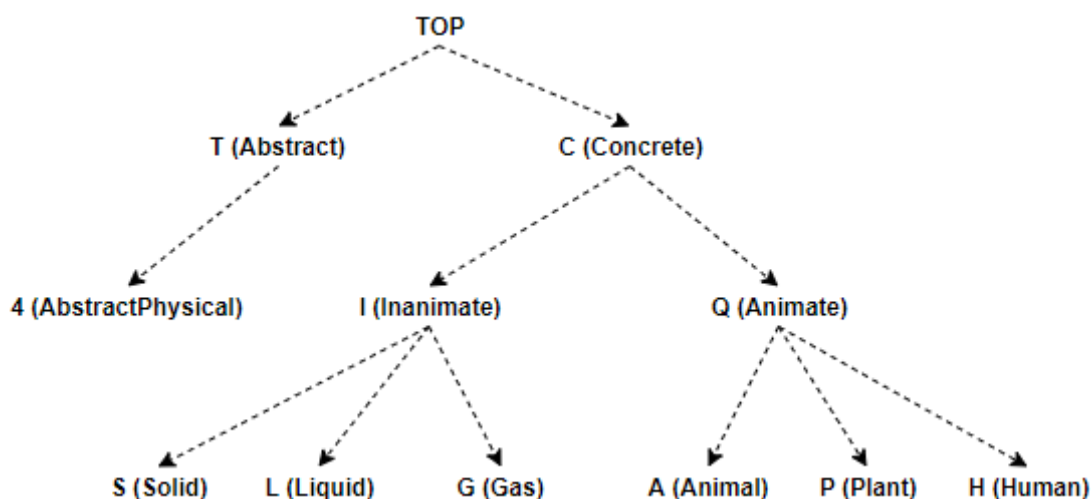


Рис. 3.7. Ієрархія класів

Ці ієрархічні набори слів корисні для пошуку синонімів або споріднених значень слів. Підхід використовує LLOCE для пошуку синонімів та порівняння їх із вмістом географічного довідника кожного класу.

3.6. Приклад імплементації моделі

У цьому розділі деталізуються етапи обробки та впровадження запропонованої моделі застосування, починаючи від ресурсів знань і закінчуючи їхньою експлуатацією в цифровому сільському господарстві. Зокрема, процес включає попередню обробку ресурсів знань, їх моделювання за допомогою моделі ОАК та трансформацію в репозиторій знань. Останній

підрозділ описує розгортання модуля Knowledge Browser (Браузер знань) для пошуку знань.

У контексті проекту ми використовуємо набори даних для проведення експериментів та вилучення знань із сирих даних. Видобуті знання можуть бути виявлені безпосередньо з цих наборів даних. Набори даних включають декілька піднаборів, що стосуються поля, ґрунту, врожайності, добрив та погоди, кожен з яких містить таку інформацію:

- Поле: місцезнаходження (довгота, широта), висота над рівнем моря, межа, площа та карта поля.

- Сільськогосподарські культури: робоча площа, назва культури, назва сорту, норма висіву, дата посіву, дата збору врожаю, цільова врожайність, середня врожайність.

- Погода: температура повітря, напрямок/швидкість/пориви вітру, хмарність, видимість, сонячна радіація, евапотранспірація, опади, тривалість сонячного сяйва та снігові накопичення.

- Зразки ґрунту: рН, Р, К, МG (хімічні характеристики), пісок, мул, глина, текстура ґрунту, вологість ґрунту, органічна речовина та температура ґрунту.

- Захист: фунгіциди, гербіциди, інсектициди, молюскоциди, регулятори росту та десиканти.

Усі точки даних є результатом відбору проб ґрунту, моніторингу врожайності та сільськогосподарської діяльності протягом п'яти років (2015–2019). Ці набори даних пройшли попередню обробку, очищення та верифікацію для досягнення високої якості, необхідної для завдань дата-майнінгу.

Для отримання видобутих знань для моделі, в рамках роботи було проведено експеримент із прогнозування рН ґрунту полів на основі наборів даних.

Експеримент ґрунтується на значеннях рН сусідніх полів із застосуванням поширених алгоритмів регресії: лінійна регресія, регресія

опорних векторів (SVR), регресія дерева рішень, LASSO (Least Absolute Shrinkage and Selection Operator), випадкові ліси та градієнтний бустинг дерева регресії.

Для кожного поля у наборі даних визначаються найближчі поля, що знаходяться в межах заданого радіусу (r_i) від 200 до 2,000 метрів. Для прогнозування атрибута рН об'єкта даних у (поля у) в експерименті оцінюються середні, максимальні та мінімальні значення рН на основі значень рН сусідніх полів у радіусі r_i :

$$pH_{\min}(y, r_i) = \min(pH(x_j));$$

$$pH_{\max}(y, r_i) = \max(pH(x_j));$$

$$pH_{\text{avg}}(y, r_i) = \frac{\sum_{j=1}^k pH(x_j)}{k};$$

де k — кількість сусідів у радіусі r_i (наприклад, 200 м, 300 м, 400 м та 500 м), x_j — сусіднє поле у цій області ($j=\{1..k\}$).

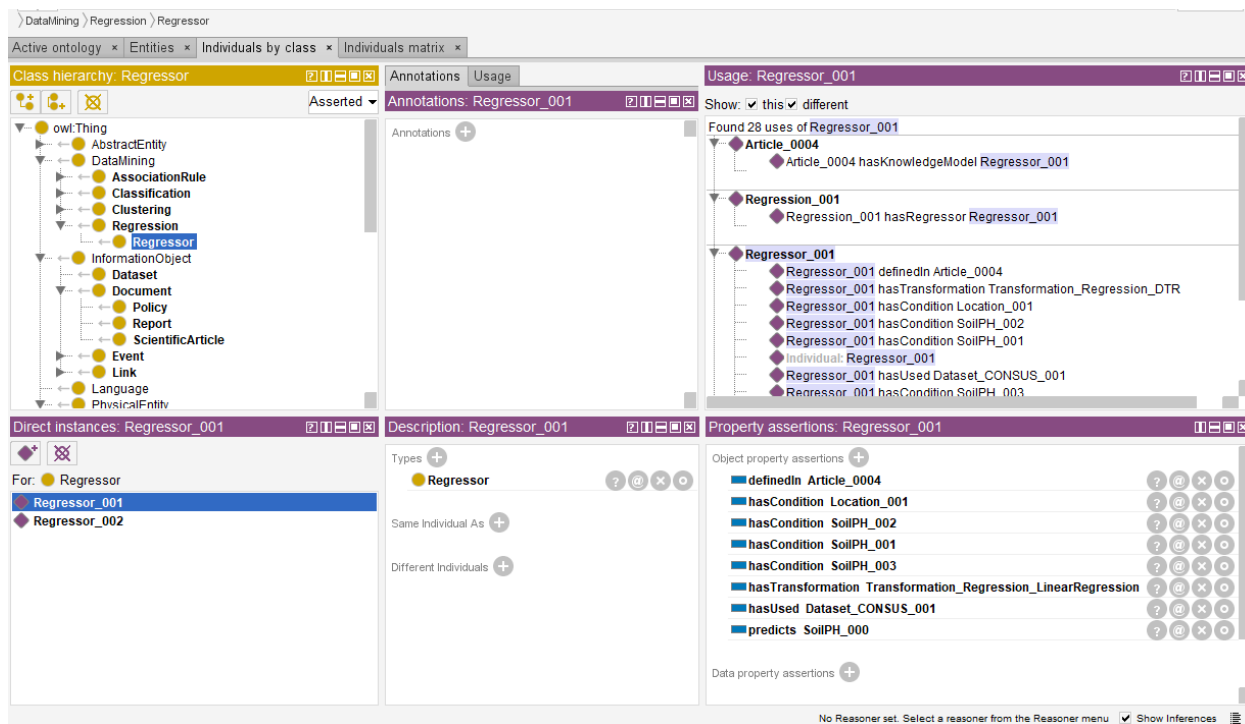


Рис. 3.8. Приклад моделі прогнозування ґрунту засобами Protege

Видобуте знання може бути представлено як репрезентація знань. Вхідними даними моделі прогнозування є інформація рН, а вихідним результатом також є значення рН, але вони використовують різні трансформації. Зокрема, вхідні ознаки обчислюються як pH_{max} , pH_{min} та pH_{avg} , тоді як вихідна ознака прогнозується як вихідне значення рН поля. (Ці умови визначені як екземпляри Soil_001, Soil_002, Soil_003 та Soil_000, як показано на рисунку 3.8).

Приклад видобутого знання Regressor_001 (рис. 3.8) визначений у статті Article_004. На панелі праворуч (рис. 3.8) відображаються умови, набір даних, алгоритми (трансформація) та прогнозований атрибут (SoilPH) для елемента знань Regressor_001.

Semantic Search Portal є однією з ключових прикладних систем експлуатації знань, призначеною для пошуку видобутих елементів знань користувачами на основі вхідних запитів або ключових слів. Вхідний запит може бути простим реченням, наприклад, "прогнозувати врожайність культури" (рис. 3.9).

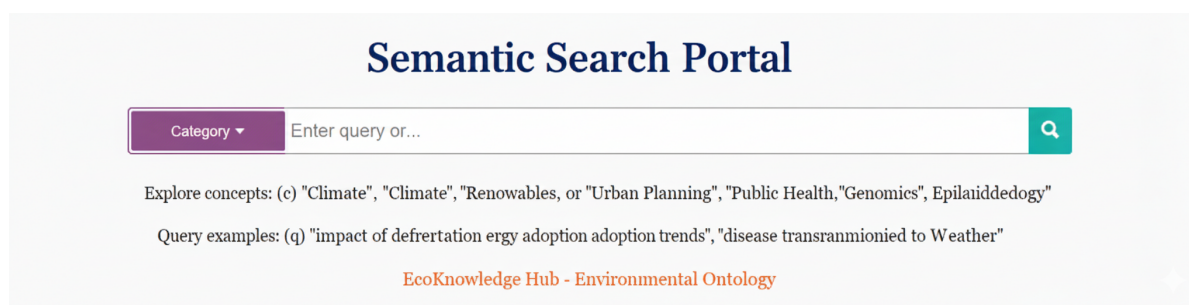


Рис. 3.9. Вікно пошуку системи

Процес обробки запиту включає послідовні етапи:

- Виявлення концептів зі вхідних пошукових запитів.
- Сегментація концептів на складові частини моделі знань.
- Генерація запитів SPARQL.
- Виконання запитів SPARQL за допомогою SPARQL Engine.

- Відображення результатів (веб-орієнтоване) з вихідних даних SPARQL Engine.

Для забезпечення обробки запитів, орієнтованих на людину, ядро системи використовує евристичний підхід для аналізу вхідних запитів, поділу їх на функціональні ролі та генерації відповідних запитів SPARQL. Евристика базується на географічному довіднику (gazetteer) для ідентифікації елементів представлень знань.

Ядро також застосовує базові техніки NLP для нормалізації токенів із вхідних запитів.

У моделі ОАК, Semantic Search Portal функціонує як система пошуку знань, що базується на RDF-сховищі. Пошуковий механізм генерує скрипт SPARQL для запиту до системи. Наприклад, вхідний запит "predict Yield in the United Kingdom" (прогнозувати Врожайність у Великій Британії) шукає моделі прогнозування, які мають Yield (Врожайність) як цільовий атрибут (Target) і стосуються United Kingdom (Велика Британія) як контексту (Location).

SPARQL-запит виконується в модулі системи через SPARQL API, а вихідні RDF-результати візуалізуються у вигляді веб-сторінки для кінцевих користувачів (рис. 3.10).

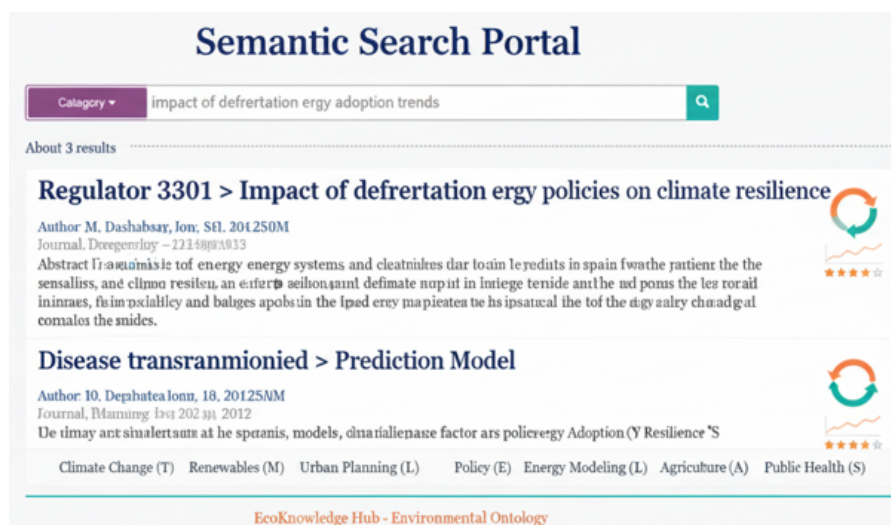


Рис. 3.10. Візуалізація результатів

Наприклад, для запиту, що розглядається, було знайдено 4 елементи знань. Кожен елемент включає вхідні умови, вихідні цілі, алгоритми обробки даних, а також місцезнаходження та середовище експериментів.

При розширенні запиту до "predict Yield in the United Kingdom based on Temperature" (прогнозувати врожайність у великій британії на основі температури), що включає додаткові умови, було знайдено лише один валідний результат (рис. 3.11).

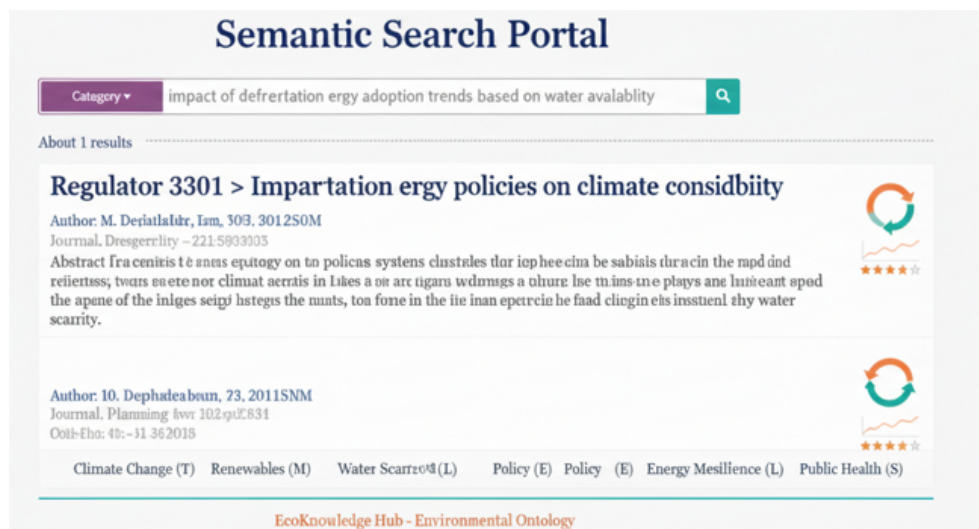


Рис. 3.11. Результат пошуку для комплексного запиту

У іншому випадку, коли запит змінюється на "predict Temperature in the United Kingdom based on Yield" (прогнозувати температуру у великій британії на основі врожайності), не знайдено жодного відповідного результату. Хоча набір термінів однаковий, зміна їхньої функціональної ролі призводить до зміни результату пошуку, що підтверджує чутливість системи до семантики запиту.

Висновки до розділу

Отже, в цьому розділі описано три основні етапи реалізації системи управління знаннями на основі моделі ОАК (Ontology-bAsed Knowledge Map). Ця система була імплементована впроваджена як Semantic Search

Portal, що виконує функції експлуатації знань. Етапи включали: вилучення знань, їх моделювання у формі представлення знань, зберігання в репозиторії знань на основі технології RDF-сховища, і, нарешті, Semantic Search Portal для пошуку знань за вхідними запитами. Ця система може використовуватись для представлення, зберігання та пошуку видобутих знань, отриманих із результатів дата-майнінгу та наукових робіт у домені цифрового сільського господарства та агрокультури.

ВИСНОВКИ

В магістерській роботі досліджено онтологічні моделі карт видобутих знань в технологіях цифрової агрокультури. У першому розділі було здійснено системний аналіз теоретичних та прикладних аспектів представлення й управління знаннями у сфері цифрової агрокультури. Встановлено, що сучасний розвиток агротехнологій базується на обробці великих обсягів даних, отриманих із різноманітних джерел — сенсорних систем, супутникового моніторингу, геоінформаційних платформ, експертних систем та машинного навчання. Окреслено ключову проблему – складність інтеграції та уніфікації знань, видобутих у процесі інтелектуального аналізу даних, що зумовлює необхідність застосування онтологічних підходів.

Розглянуто концептуальні основи класифікації знань у системах штучного інтелекту, визначено їхню роль у побудові карт знань та описано основні типи таких карт. Встановлено, що ефективність побудови та застосування карт знань безпосередньо залежить від правильного вибору методології, що враховує як формальні, так і семантичні характеристики даних. Окрему увагу приділено аналізу сучасних технологій зберігання знань, де порівняно можливості реляційних і нативних систем зберігання графів знань, а також розглянуто аспекти використання мов запитів для їхньої обробки.

У другому розділі проведено глибокий аналіз сучасних моделей та методологій обробки знань у контексті цифрової агрокультури. Розглянуто особливості функціонування експертних систем, систем управління знаннями та репозиторіїв, які виступають основними інструментами для накопичення, структуризації та поширення аграрних знань.

Запропоновано концепцію онтологічної моделі карти знань, яка забезпечує уніфіковане представлення видобутих знань, їх інтерпретацію та подальшу інтеграцію у прикладні системи. Обґрунтовано, що ключову роль у

даному процесі відіграє онтологія, оскільки вона дозволяє формалізувати доменну область та встановити семантичні зв'язки між об'єктами агрокомплексу.

Визначено архітектуру онтологічно-орієнтованої моделі карти знань, де виділено основні компоненти: майнінг знань, моделювання, вилучення, оцінку, публікацію та використання знань. Запропонована методологія передбачає поетапний процес — від розпізнавання сутностей і формування концептів до створення систем управління знаннями, орієнтованих на специфіку цифрової агрокультури.

У третьому розділі безпосередньо реалізовано процес побудови онтологічних моделей карт знань для предметної області цифрової агрокультури. Визначено вимоги до створення онтології, включаючи концептуальну повноту, формальну коректність, масштабованість та можливість інтеграції з іншими інформаційними системами. Обґрунтовано основні принципи проектування та розроблено структуру онтології, яка відображає концептуальне ядро галузі.

Детально розглянуто процеси специфікації, технічних характеристик та інтеграційних аспектів побудови онтології. Показано, що успішна імплементація можлива лише за умови застосування формалізованих методів обробки і перетворення даних, які дозволяють інтегрувати інформацію з різнорідних джерел. Розроблено архітектуру системи розпізнавання сільськогосподарських сутностей, що забезпечує ідентифікацію та семантичну інтерпретацію ключових понять аграрної галузі.

Окрему увагу приділено прикладу імплементації запропонованої моделі, що підтвердив її практичну ефективність для вирішення завдань управління знаннями в агросекторі. Таким чином, результати третього розділу доводять доцільність і ефективність застосування онтологічного підходу для побудови карт знань, що сприяють розвитку інноваційних рішень у сфері цифрової агрокультури та забезпечують новий рівень інтеграції й управління аграрними даними.

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Zheng, Y., et al. (2012). Construction of the ontology-based agricultural knowledge management system. *Computers and Electronics in Agriculture*.
2. Deepa, R., et al. (2022). An effective automated ontology construction based on the Jaccard relative extractor and Naïve Bayes classifier. *ETRI Journal*.
3. Saravanan, K. S., et al. (2024). Innovative agricultural ontology construction using NLP approaches. *Computers and Electronics in Agriculture*.
4. Bhattacharya, S., Pandey, M. (2024). Enhancing Agricultural Knowledge Management using an efficient & novel ontology-based approach leveraging BERT-GPT and Graph Recurrent Q Learning Network. *Journal of Emerging Sciences (JES)*.
5. “Building Knowledge Graphs Towards a Global Food System” (2025). arXiv preprint.
6. Uchihira, N., Yoshida, M. (2020). Agricultural Knowledge Management Using Smart Voice Messaging Systems: Combination of Physical and Human Sensors. arXiv preprint.
7. Le-Khac, N.-A., Aouad, L. M., Kechadi, M.-T. (2007). Knowledge map: toward a new approach supporting knowledge management in distributed data mining. In *Proceedings of ICAS 2007*. (IEEE)
8. Lecocq, R., Valcartier, D. (2006). Knowledge mapping: a conceptual model. DRDC Valcartier Technical Report (Canada).
9. Kim, S., Suh, E., Hwang, H. (2003). Building the knowledge map: an industrial case study. *Journal of Knowledge Management*, 7(2), 34–45.
10. Liu, J., Goering, C. E., Tian, L. (2001). A neural network for setting target corn yields. *Transactions of ASAE*, 44(3).
11. Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L., Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and Electronics in Agriculture*, 121, 57–65.

12. Karimi, Y., Prasher, S. O., Patel, R. M., & Kim, S. H. (2006). Application of support vector machine technology for weed and nitrogen stress detection in corn. *Computers and Electronics in Agriculture*, 51(1–2), 99–109.
13. Eppler, M. J. (2006). A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. *Information Visualization*, 5(3), 202–210.
14. Pei, X., Wang, C. (2009). A study on the construction of knowledge map in matrix organizations. In 2009 International Conference on Management and Service Science.
15. Abdullah, A., Murali, E., S., Balusamy, B., Rajashree, S. (2023). Design of Multiple Ontology Based Agro Knowledge Mining Model. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(7), 47–56.
16. FAO. (2024). AGROVOC: Multilingual Controlled Vocabulary. Food and Agriculture Organization of the United Nations.
17. Ingram, J. (2019). Searching for meaning: Co-constructing ontologies with farmers. *New Zealand Journal of Agricultural Research*.
18. Spivak, D. I., & Kent, R. E. (2012). Ologs: A Categorical Framework for Knowledge Representation. *PLOS ONE*.
19. Bishop, T. F. A., McBratney, A. B. (2001). A comparison of prediction methods for creation of field-extent soil property maps. *Geoderma*, 103(1–2).
20. Cheng, Y., Kuang, J., Cheng, D., Zheng, J., Gao, M., Zhou, A. (2019). AgriKG: an agricultural knowledge graph and its applications. In DASFAA 2019. *Lecture Notes in Computer Science*, vol. 11448.
21. Chenglin, Q., Qing, S., Pengzhou, Z., Hui, Y. (2018). Cn-MAKG: China meteorology and agriculture knowledge graph construction based on semi-structured data. In IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS).

22. Maltas, A., Charles, R., Jeangros, B., Sinaj, S. (2013). Effect of organic fertilizers and reduced tillage on soil properties, crop nitrogen response and crop yield: results of a 12-year experiment. *Soil & Tillage Research*, 126, 11–18.
23. Loreto, E., et al. (2015). Integrated sensor networks and data fusion in precision agriculture. *Journal of Sensors*.
24. Zhang, Y., et al. (2021). Semantic integration in agriculture: challenges and opportunities. *Computers in Agriculture and Food Systems*.
25. Guarino, N. (1998). Formal Ontology in Conceptual Analysis and Knowledge Representation. *International Journal of Human-Computer Studies*, 43(5–6), 625–640.
26. Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220.
27. Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *The Knowledge Engineering Review*, 11(2), 93–136.
28. Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology*. Stanford Knowledge Systems Laboratory Technical Report.
29. Bernstein, A., Kaufmann, E. (2006). GINO – A Guided Input Natural Language Ontology Editor. In *Proceedings of ISWC / ASWC*.
30. Shadbolt, N., Hall, W., Berners-Lee, T. (2006). The Semantic Web Revisited. *IEEE Intelligent Systems*, 21(3), 96–101.
31. Bizer, C., Heath, T., Berners-Lee, T. (2009). Linked Data – The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22.