Факультет комп’ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)
Кафедра комп’ютерних наук
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

магістр
(назва освітнього ступеня)

на тему: Семантичне управління озерами сховищ даних "розумних міст"

Виконав: студент VI курсу, групи СНм-61
спеціальності 122 Комп’ютерні науки
(шифр і назва спеціальності)

(підпис) Лебідко Д.М.
(прізвище та ініціали)

Керівник (підпис) Марценко С.В.
(прізвище та ініціали)

Нормоконтроль (підпис) Дуда О.М.
(прізвище та ініціали)

Завідувач кафедри (підпис) Боднарчук І.О.
(прізвище та ініціали)

Рецензент (підпис) Оробчук О.Р.
(прізвище та ініціали)

Тернопіль
2023

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних наук
(повна назва кафедри)

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ Боднарчук І.О.
(підпис) (прізвище та ініціали)
«_25_» _грудня_ 2023 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

на здобуття освітнього ступеня Магістр
(назва освітнього ступеня)

за спеціальністю 122 Комп'ютерні науки
(шифр і назва спеціальності)

Студенту Лебідко Дмитро Миколайович
(прізвище, ім'я, по батькові)

1. Тема роботи Семантичне управління озерами сховищ даних "розумних міст"

Керівник роботи Марценко Сергій Володимирович, к.т.н., доцент кафедри КН
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «_24_» _листопада_ 2023 року № _4/7-1099_____

2. Термін подання студентом завершеної роботи 26 грудня 2023р.

3. Вихідні дані до роботи Наукові публікації про семантичне управління даними, сховища даних, озера сховищ даних та "розумні міста"

4. Зміст роботи (перелік питань, які потрібно розробити)
Вступ. 1 Управління семантичними даними в озерах даних «розумних міст». 2 Семантичне моделювання для інтеграції даних «розумних міст». 3. Доступ до даних «розумних міст» на основі онтології. 4 Охорона праці та безпека в надзвичайних ситуаціях. Висновки. Додатки

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)
1 Титульна сторінка. 2 Тема, Мета, Об'єкт, Предмет дослідження. 3 Завдання дослідження. 4 Актуальність дослідження. 5 Озера даних «розумних містах». 6 Таксономія концепцій і методів. 7 Порівняння озер даних із семантичним шаром. 8. Порівняння платформ семантичного моделювання. 9 Висновки. 10 Завершальний слайд.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Охорона праці	Сенчишин В.С., доцент		
Безпека в надзвичайних ситуаціях	Клепчик В.М., ст. викладач		

7. Дата видачі завдання 24 листопада 2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	25.11.2023	*Виконано*
2.	Підбір наукових джерел про семантичне управління	26.11.2023-28.11.2023	*Виконано*
	даними, сховища даних, озера сховищ даних та		
	"розумні міста"		
3.	Опрацювання наукових публікацій та збір даних по	29.11.2023-1.12.2023	*Виконано*
	темі роботи		
4.	Виконання дослідження згідно мети кваліфікаційної	2.12.2023-4.12.2023	*Виконано*
	роботи		
5.	Оформлення розділу «Управління семантичними	5.12.2023-7.12.2023	*Виконано*
	даними в озерах даних «розумних міст»»		
6.	Оформлення розділу «Семантичне моделювання для	8.12.2023-10.12.2023	*Виконано*
	інтеграції даних «розумних міст»»		
7.	Оформлення розділу «Доступ до даних «розумних	11.12.2023-13.12.2023	*Виконано*
	міст» на основі онтології»		
8.	Виконання завдання до підрозділу «Охорона праці»	14.12.2023-15.12.2023	*Виконано*
9.	Виконання завдання до підрозділу «Безпека в	16.12.2023-17.12.2023	*Виконано*
	надзвичайних ситуаціях»		
10.	Оформлення кваліфікаційної роботи	18.12.2023-19.12.2023	*Виконано*
11.	Нормоконтроль	19.12.2023-20.12.2023	*Виконано*
12.	Перевірка на плагіат	21.12.2023	*Виконано*
13.	Попередній захист кваліфікаційної роботи	22.12.2023	*Виконано*
14.	Захист кваліфікаційної роботи	26.12.2023	

Студент _____ Лебідко Д.М.
(підпис) (прізвище та ініціали)

Керівник роботи _____ Марценко С.В.
(підпис) (прізвище та ініціали)

АНОТАЦІЯ

Семантичне управління озерами сховищ даних "розумних міст" // Кваліфікаційна робота освітнього рівня «Магістр» // Лебідко Дмитро Миколайович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем і програмної інженерії, кафедра комп'ютерних наук, група СНм-61 // Тернопіль, 2023 // С. 69, рис. – 6, табл. – 2, кресл. – 10, додат. – 1, бібліогр. – 64.

Ключові слова: дані, озера даних, розумне місто, семантичний опис сховища даних, онтологія, управління.

Кваліфікаційна робота присв'ячена дослідженню семантичного управління озерами сховищ даних «розумних міст». В першому розділі описано міські джерела семантичних даних. Проаналізовано стан та досліджень в царині озер даних «розумних міст». Наведено ключові терміни та визначення управління семантичними даними в озерах даних. Розроблена таксономія управління семантичними даними. Розглянуті моделі метаданих для озер даних, які враховують семантику. Описані існуючі системи семантичних озер даних «розумних міст». Виконано порівняння та обговорення семантичних озер даних «розумних міст». В другому розділі досліджено системи семантичного моделювання. Розглянуто автоматизоване присвоєння семантичних міток. Висвітлена автоматизована генерація семантичної моделі. Виконано порівняння та обговорення семантичних моделей озер даних «розумних міст». В третьому розділі кваліфікаційної роботи описано OBDA над реляційними даними «розумних міст». Розглянуто OBDA над нереляційними даними «розумних міст». Досліджено виробництво семантичних озер даних «розумних міст». Описано використання семантичних озер даних «розумних міст» в розумному місті. В четвертому розділі кваліфікаційної роботи описано медичні профілактичні заходи щодо збереження здоров'я та працездатності користувачів комп'ютерів та відеодисплейних терміналів. Розглянуто особливості заходів електробезпеки на підприємствах.

ANNOTATION

Semantic management of "Smart cities" data lakes // The educational level "Master" qualification work // Lebidko Dmytro Mykolaiovych // Ternopil Ivan Pulyuy National Technical University, Faculty of Computer Information Systems and Software Engineering, Department of Computer Science, SNm-61 group // Ternopil, 2023 // P. 69, fig. – 6, tables – 2, posters – 10, annexes – 1, ref. – 69.

Key words: data, data lakes, smart city, semantic description of data storage, ontology, management.

The qualification work is devoted to the study of the semantic management of "smart cities" data storage lakes. The first chapter describes urban sources of semantic data. The status and researches in the realm of the lakes of these "smart cities" have been analyzed. Key terms and definitions of semantic data management in data lakes are provided. A taxonomy of semantic data management has been developed. Metadata models for data lakes that take semantics into account are reviewed. Existing systems of semantic data lakes of "smart cities" are described. A comparison and discussion of the semantic lakes of the data of "smart cities" was carried out. In the second section, semantic modeling systems are studied. Automated assignment of semantic labels is considered. The automated generation of the semantic model is highlighted. A comparison and discussion of semantic models of data lakes of "smart cities" was carried out. The third section of the qualification work describes OBDA on the relational data of "smart cities". Considered OBDA on non-relational data of "smart cities". The production of semantic data lakes of "smart cities" was studied. The use of semantic data lakes of "smart cities" in a smart city is described. The fourth section of the qualification work describes preventive medical measures to preserve the health and working capacity of users of computers and video display terminals. The features of electrical safety measures at enterprises are considered.

ПЕРЕЛІК СКОРОЧЕНЬ І ТЕРМІНІВ

БД – База даних.

СУБД – Система управління базами даних.

DQ (англ. Data Quality) – Якість даних.

ELT (англ. Extract-Load-Transform) – Видобування-Завантаження-Перетворення.

Goods (англ. Google Dataset Search) – Пошук набору даних Google.

GKG (англ. Google Knowledge Graph) – Граф знань Google.

KG (англ. Knowledge Graph)– Граф знань.

OBDA (англ. Ontology Based Data Access) – Доступ до даних на основі онтології.

OWL (англ. Web Ontology Language) – Мова веб-онтології.

RDF (англ. Resource Description Framework) – Структура опису ресурсу.

STD (англ. Semantic Type Detection) – Виявлення семантичного типу.

STI (англ. Semantic Table Interpretation) – Інтерпретація семантичною таблицею.

SDM (англ. management of semantic data) – Управління семантичними даними.

SRQL (англ. Source Retrieval Query Language) – Мова запитів пошуку джерел.

ЗМІСТ

ВСТУП 7
1 УПРАВЛІННЯ СЕМАНТИЧНИМИ ДАНИМИ В ОЗЕРАХ ДАНИХ «РОЗУМНИХ МІСТ» 9
1.1 Міські джерела семантичних даних 9
1.2 Стан та досліджень в царині озер даних «розумних міст» 13
1.3 Ключові терміни та визначення управління семантичними даними в озерах даних 14
1.4 Таксономія управління семантичними даними 16
1.5 Моделі метаданих для озер даних, які враховують семантику 19
1.6 Існуючі системи семантичних озер даних «розумних міст» 20
1.7 Порівняння та обговорення семантичних озер даних «розумних міст» 23
1.8 Висновок до першого розділу 25
2 СЕМАНТИЧНЕ МОДЕЛЮВАННЯ ДЛЯ ІНТЕГРАЦІЇ ДАНИХ «РОЗУМНИХ МІСТ» 26
2.1 Системи семантичного моделювання «розумних міст» 26
2.2 Автоматизоване присвоєння семантичних міток «розумного міста» 29
2.3 Автоматизована генерація семантичної моделі «розумного міста» .. 30
2.4 Порівняння та обговорення семантичних моделей озер даних «розумних міст» 34
2.5 Висновок до другого розділу 36
3 ДОСТУП ДО ДАНИХ «РОЗУМНИХ МІСТ» НА ОСНОВІ ОНТОЛОГІЇ ... 37
3.1 OBDA над реляційними даними «розумних міст» 40
3.2 OBDA над нереляційними даними даними «розумних міст» 43
3.3 Виробництво семантичних озер даних «розумних міст» 48
3.4 Використання семантичних озер даних «розумних міст» в розумному місті 50
3.5 Висновок до третього розділу 55
4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ 56

4.1 Медичні профілактичні заходи щодо збереження здоров’я та працездатності користувачів комп’ютерів та відеодисплейних терміналів 56
4.2 Особливості заходів електробезпеки на підприємствах 57
4.3 Висновок до четвертого розділу 59
ВИСНОВКИ 60
ПЕРЕЛІК ДЖЕРЕЛ 62
ДОДАТКИ

ВСТУП

Актуальність теми. Останніми роками з’явилися озера даних, щоб керувати великими обсягами неоднорідних даних для сучасної аналітики даних. Одним із способів запобігти перетворенню озер даних на непрацездатні болота даних є семантичне керування даними. Деякі підходи пропонують зв’язати метадані з графами знань на основі принципів пов’язаних даних, щоб надати більше значення та семантики даним в озері. Такий семантичний рівень може бути використаний не тільки для управління даними, але також для вирішення проблеми інтеграції даних з різнорідних джерел, щоб зробити доступ до даних більш виразним і сумісним.

Через зростаючу залежність від даних їх безперебійне надання та споживання стає все більш важливим для різних зацікавлених сторін. Озера даних були запропоновані для вирішення проблем в управлінні неструктурованими та структурованими джерелами даних. Багато організацій створили озера даних для збору даних із різнорідних джерел, які потім можуть бути використані в проектах науки про дані [1]. Тому семантичне управління озерами сховищ даних «розумних міст» є актуальним напрямком досліджень.

Мета і задачі дослідження. Метою даної кваліфікаційної роботи освітнього рівня «Магістр» є підвищення рівня повноти подання інформації щодо методів і інструментів для семантичного управління озерами сховищ даних «розумних міст». Для досягнення поставленої мети потрібно виконати ряд завдань, зокрема:

– Проаналізувати стан досліджень в озер даних «розумних міст».

– Дослідити існуючі на даний час моделі метаданих для озер даних, які враховують семантику.

– Проаналізувати методи системи семантичного моделювання.

Об’єкт дослідження семантичне управління озерами сховищ даних "розумних міст".

Предмет дослідження. методи і інструменти для семантичного управління озерами сховищ даних "розумних міст".

Наукова новизна одержаних результатів кваліфікаційної роботи полягає у тому, що отримали подальше застосування моделі метаданих для озер даних, які враховують семантику «розумних міст».

Практичне значення одержаних результатів. Виконано порівняння OBDA над реляційними та нереляційними даними «розумних міст»

Апробація результатів магістерської роботи. Основні результати проведених досліджень обговорювались на XI науково-технічної конференції «Інформаційні моделі, системи та технології» Тернопільського національного технічного університету імені Івана Пулюя (м. Тернопіль, 2023 р.).

Публікації. Основні результати кваліфікаційної роботи опубліковано у двох працях конференції (Див. додатки А).

Структура й обсяг кваліфікаційної роботи. Кваліфікаційна робота складається зі вступу, чотирьох розділів, висновків, списку літератури з 64 найменувань та одного додатку. Загальний обсяг кваліфікаційної роботи складає 69 сторінок, з них 47 сторінок основного тексту, який містить 6 рисунків та 2 таблиці.

1 УПРАВЛІННЯ СЕМАНТИЧНИМИ ДАНИМИ В ОЗЕРАХ ДАНИХ «РОЗУМНИХ МІСТ»

1.1 Міські джерела семантичних даних

Опис джерел даних із значущими метаданими є важливим для зручності використання даних «розумного міста». Особливо для користувачів з обмеженими знаннями предметної області або коли вони не знайомі з набором даних, важливі метадані є незамінними. Відображення необроблених даних із джерел даних у семантично насичених моделях підвищує зручність використання та інтерпретацію даних. Використання графів знань (KG), обладнаних якісно-сформованою виразною онтологією для забезпечення чітко визначеного значення [2], виявилося корисним різними способами, наприклад, для інтеграції даних «розумних міст» за допомогою доступу до даних на основі онтології (OBDA) [3].

Ці переваги були визнані в дослідженнях і практиці озер даних в останні роки. Хоча перші реалізації були спрямовані на ефективну обробку великих даних за допомогою розподілених масштабованих систем, наприклад Hadoop, також була визнана потреба в належному управлінні метаданими «розумних міст» та якістю даних [4]. Було запропоновано декілька підходів для управління семантичними даними (SDM) в озерах даних. Озера семантичних даних потребують інтеграції технологій Big Data і Semantic Web. У семантичній мережі RDF і OWL є основними мовами для подання даних як набору пов’язаних елементів даних. Ці мови також можна використовувати для подання метаданих «розумних міст» у формі онтологій або KG. Незважаючи на те, що технології пов’язаних даних і семантичного вебу стали більш зрілими за останні роки, обсяг даних, що розглядаються в застосунках семантичного вебу, набагато менший, ніж у застосунках великих даних «розумних міст». Таким чином, масштабованість до великих гетерогенних наборів даних є основною проблемою для застосування технологій семантичного вебу в озерах даних «розумних міст».

Семантичні дані розширюють звичайні метадані, які можна видобути, наприклад, схему, типи даних, щоб передавати контекстну інформацію, яка не властива окремому джерелу даних «розумного міста». Згідно з [5], є два важливі процеси, які забезпечують якість і прийняття SDM:

1) створення та підтримка концептуалізацій;

2) опис джерел даних «розумного міста» у формі семантичних моделей, пов'язуючи вихідні схеми з концептуалізацією.

Семантично добре анотований джерело даних «розумного міста» можна виявити за допомогою концептуальних подань даних і зрозуміти за допомогою наданої контекстної інформації, що зберігається в моделі. Створення таких семантичних моделей «розумного міста» передбачає декодування поточного джерела даних, перевірку придатних KG і зв'язування атрибутів даних із поняттями в KG. Тому KG можна розглядати як загальну модель даних, яка забезпечує концептуальний опис ресурсів даних організації. Семантичну модель можна розглядати як проекцію сутностей і зв'язків KG на джерела даних.

Більшість концепцій семантичної моделі «розумного міста» взято з KG, але для опису конкретних наборів даних можна визначити нові класи та зв'язки.

Семантична модель є додатковим шаром між рівнем даних і рівнем знань. Основна ідея семантичної моделі подана на рисунку 1.1. Необроблені набори даних в озері даних подані на нижньому рівні даних; вони можуть мати різні формати, наприклад табличні дані чи ієрархічні дані JSON, але мати частково перекриваючий вміст. Як правило, озера даних можуть автоматично отримувати інформацію про схему з джерел. Семантичну модель можна подати у вигляді графа з вузлами та ребрами, отриманими з KG, збагаченого вузлами та ребрами для подання відображення та розширеної семантики джерел даних. У прикладі на рис. 1.1 ми повторно використовуємо концепції та властивості з онтології WikiData (префікс wd) як частину нашого KG. За допомогою явного кодування семантичного типу та зв'язку між атрибутами джерела в графі семантична модель точно описує передбачуване значення джерела даних. Платформа SDM повинна підтримувати процес, наприклад, пропонуючи відповідні концепції та відповідні відображення під час створення семантичної моделі.

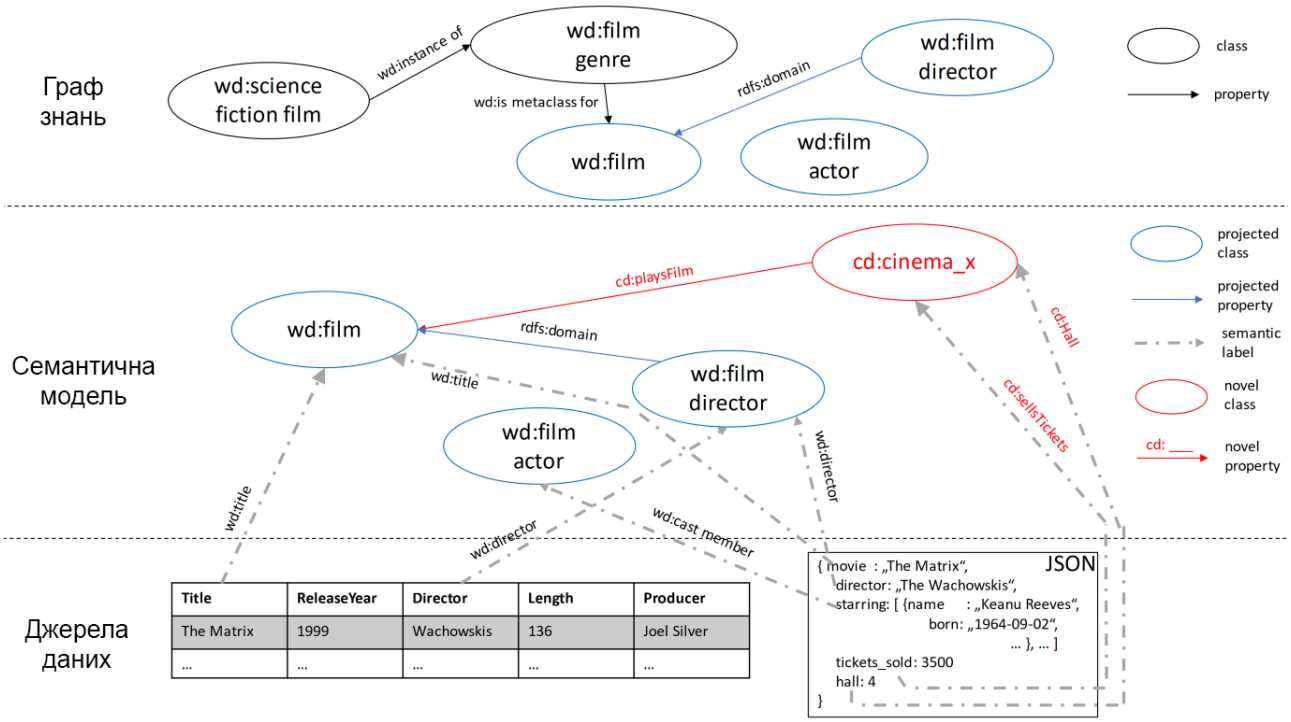

Рисунок 1.1 – Управління семантичними даними для різнорідних джерел даних в озері даних «розумного міста» [6]

Крім того, замість того, щоб використовувати лише статичний KG, семантична модель може також вводити нові класи та властивості. Ця потреба в еволюції виникає, коли користувачі надають джерела даних «розумного міста», які включають концепції та зв'язки, які не охоплюються KG. У цьому прикладі документ JSON містить дані про певну послугу, префікс cd – ознака домена, який потрібно змоделювати явно. Ці нові знання мають бути структурованими, тим самим постійно розвиваючи рівень знань [7].

Семантичні мітки – це прямі анотації елементів джерела даних «розумного міста», наприклад, атрибутів схеми з елементами семантичної моделі. Семантика визначається відповідним предикатом і об'єктом. Семантична мітка зазвичай подається у вигляді трійки: атрибут схеми, предикат, об'єкт. На рис. 1.1 подана семантична мітка атрибута таблиці «Назва» побудована через об'єкт «wd:film» і предикат «wd:назва», що моделює зв'язок між ними. Озеро семантичних даних «розумного міста» має сприяти вилученню семантичної моделі з необхідних наборів даних, тобто більш концептуального опису наборів даних, включаючи концепції та їхні зв'язки. Потім семантичні моделі пов'язуються з наборами

даних «розумного міста», щоб знати, який набір даних відноситься до якого поняття [8].

На рисунку 1.2 запропоновано розширену версію архітектури озера даних, де функції, пов'язані з метаданими, збагачені семантикою.

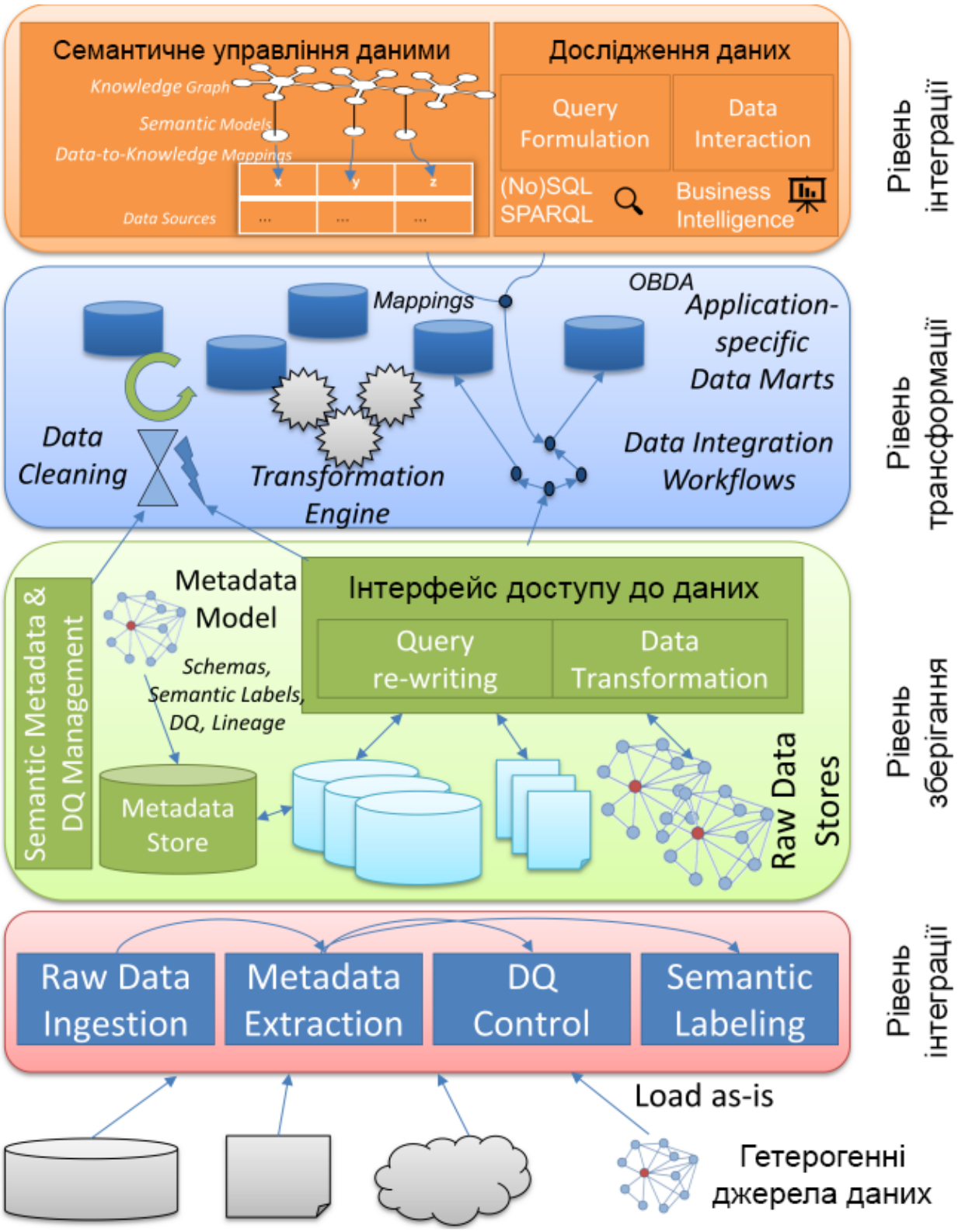

Рисунок 1.2 – Архітектура озера семантичних даних, розширена з [6]

Наприклад, компонент семантичного маркування на рівні прийому додає семантичні мітки до витягнутих елементів метаданих «розумного міста». Управління якістю даних (DQ) також можна покращити, наприклад, шляхом перевірки наявності семантичної інформації для джерела даних «розумних міст». Семантична інформація, мітки, моделі, KG тощо, управляється на рівні зберігання в розширеному сховищі семантичних метаданих «розумного міста», який забезпечує, наприклад, відображення, необхідні для OBDA. Зокрема, семантичні метадані повинні сприяти використанню та інтерпретації даних «розумного міста». Таким чином, рівень взаємодії має декілька додаткових компонентів, наприклад, для перегляду KG і семантичних моделей, для визначення запитів за допомогою мови семантичних запитів, наприклад SPARQL1, і редакторів для вдосконалення семантичних відображень і моделей «розумного міста».

1.2 Стан та досліджень в царині озер даних «розумних міст»

Проведемо огляд та аналіз останніх розробок у SDM, пов’язаних, зокрема, із семантичними озерами даних «розумного міста». На відміну від ряду досліджень у цій галузі, зосередимося на підходах, які можна масштабувати до великих даних і сприяти семантичній інтеграції даних. Автори [9] зосередились на загальних архітектурах озер даних та управлінні метаданими «розумного міста», а також обговорює переваги та недоліки озер даних та альтернатив їх дизайну. Подібним чином подано комплексний огляд дослідницьких питань для проектування та побудови загальних озер даних. Вони класифікують існуючі підходи та системи на основі їх надає функції для озер даних і забезпечує ретельне порівняння існуючих рішень і обговорення відкритих дослідницьких завдань.

Автори [10] досліджують існуючі підходи до семантичного моделювання, обговорюють їхні сильні та слабкі сторони для використання в реальному світі та подають майбутні виклики та необхідні напрямки досліджень, на яких

спільнота повинна зосередитися, щоб зробити управління семантичними даними прийнятним у «розумних міст».

В [11] подано теоретичне обговорення формальних деталей реалізації в OBDA. Було описано екосистему інструментів і конкретні випадки використання в широкому діапазоні комерційних застосувань «розумних міст».

1.3 Ключові терміни та визначення управління семантичними даними в озерах даних

Семантичні інструменти часто застосовують стандарти W3C, наприклад, RDF, OWL, але деякі промислові рішення «розумних міст» не зосереджуються на сумісності зі стандартом, а моделюють KG та семантичні метадані іншим способом. Подамо ключові визначення щодо озер даних.

Визначення 1. «Озеро даних – це гнучка, масштабована система зберігання та керування даними, яка отримує та зберігає необроблені дані з різнорідних джерел у їх оригінальному форматі та забезпечує обробку запитів і аналітику даних у режимі «на льоту».» [12]

Зокрема, тут обробка запитів у режимі «on-the-fly» відноситься до принципу ELT. У сховищах даних використовується класична ідея ETL: дані перетворюються в єдину схему перед завантаженням. Лише тоді запити можна буде виконувати до цього інтегрованого набору даних. В озерах даних «розумного міста» необроблені дані зберігаються на гетерогенному рівні, обробка запитів повинна мати можливість працювати з гетерогенними системами зберігання. Наприклад, як це робить Apache Spark зі своєю мовою запитів SparkSQL. У контексті великих даних така поведінка також називається схемою при читанні. Ще один важливий аспект: розглядаємо лише ті підходи, які не вимагають перетворення даних в уніфіковану схему, оскільки такі підходи просто не можна масштабувати до великих даних.

Визначення 2. Озера семантичних даних – це особлива форма традиційних озер даних, які розширюють можливості за допомогою семантичного рівня, який збагачує та семантично пов’язує збережені дані. Семантичний рівень забезпечує

набори даних в озері «розумного міста» зв’язками від метаданих набору даних до концептуальних та логічних моделей, які інкапсулюють знання, потенційно зовнішні щодо вмісту даних, наприклад знання предметної області.

Визначення 3. Семантичні мітки демонструють зв’язок між джерелами даних і концепціями з метою опису семантики атрибута джерела даних «розумного міста», з яким пов’язана мітка.

Семантичні мітки є основною концепцією впровадження семантичного керування даними на практиці. Вони доповнюють метадані, які можна видобути, наприклад, типи даних, розміри, формати тощо, щоб передати контекстну інформацію, яка може не бути притаманною даному джерелу даних «розумного міста». Найпоширенішими семантичними мітками є ті, які з’єднують елементи схеми джерела даних у зв’язку один-до-одного, наприклад, стовпець таблиці або вузол в ієрархічному та вкладеному наборі даних, наприклад, файл JSON, їхнім аналогам із KG «розумного міста». Однак їх також можна використовувати, наприклад, на рівні окремих значень даних «розумного міста», для певного рядка таблиці, для позначення зв’язків між окремими записами схеми, цілим набором даних або будь-яким іншим відповідним поняттям. У літературі для семантичних міток використовуються різні терміни, наприклад, семантичне збагачення, посилання, теги, типи, профілювання або анотації [13]. Ідея семантичних міток відрізняється від підходів, які моделюють увесь каталог метаданих та даних «розумного міста» як KG, наприклад, тим фактом, що їх основною метою є надання контексту, отже, додаткової інформації із зовнішніх джерел, на відміну від даних, які зберігають метадані в моделі даних на основі графів.

У сфері OBDA також широко використовується термін відображення. Відображення встановлюють зв’язок між атрибутами даних і KG, і вони є технічною передумовою для перезапису запитів, наприклад, із SPARQL на SQL. Хоча відображення в OBDA технічно стосуються різних понять, однак на концептуальному рівні їх можна розглядати як тісно пов’язані із семантичними мітками, оскільки по суті вони пов’язують метадані джерела даних із екземплярами концептуалізації «розумного міста». Основна відмінність полягає в тому, що відображення в OBDA вимагають точної семантики, щоб правильно

переписати запити. Семантичні мітки зазвичай не визначаються з урахуванням точної семантики, оскільки вони спрямовані на надання користувачеві додаткової контекстної інформації або на підключення пов’язаних наборів даних «розумного міста».

Визначення 4. Семантична модель визначає ширший контекст джерела даних шляхом розширення семантичних міток за допомогою додаткових значущих понять і забезпечує відповідні зв’язки, які зберігаються між різними поняттями.

Деякі підходи, наприклад [14], встановлюють зв’язок між будь-якими двома стовпцями в таблиці безпосередньо через предикат без проміжного елемента семантичної моделі «розумного міста» між ними. Наприклад, трійка складається з двох стовпців таблиці, і лише предикат походить від концептуалізації «(ci,p,cj)». Семантичні мітки, особливо коли розглядаються як «слабкі» анотації, часто також моделюються як додаткове поле в сховищі метаданих і не виражаються як трійка в RDF, наприклад, у GEMMS. Якщо хтось думає, що предикат є «RelatedTo», дана формалізація все ще застосовується без подальших специфікацій «розумного міста».

1.4 Таксономія управління семантичними даними

На рисунку 1.3 структуровано різні концепції та методи у таксономію «розумного міста». Підходи горизонтально розділені трьома основними напрямками:

– управління даними;
– управління знаннями;
– їхній конкретний перетин SDM.

Розрізняємо п’ять вимірів по вертикалі, які поступово знижують рівень абстракції, щоб завершитися конкретними застосунками «розумного міста». Перший вимір подає концептуальні компоненти вищого рівня.

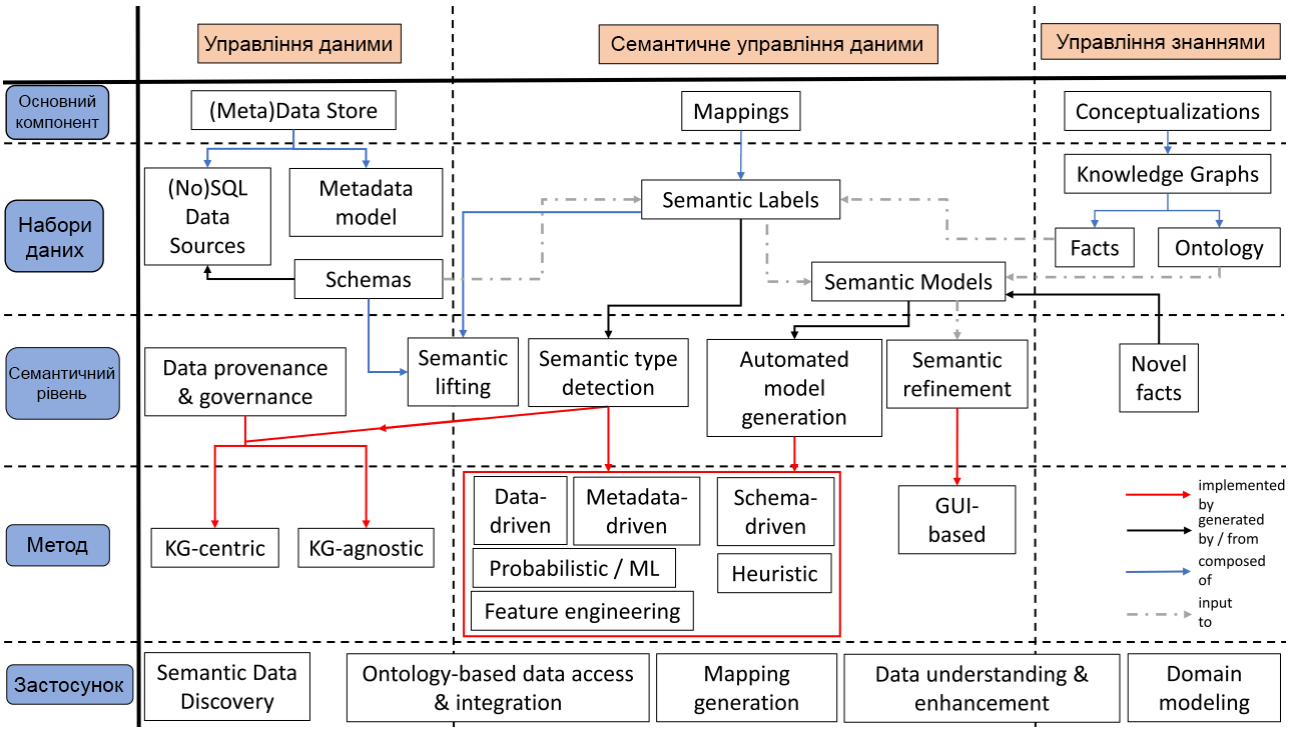

Рисунок 1.3 – Таксономія концепцій і методів [6]

Вони включають систему озера даних, тобто джерела даних «розумного міста» без SQL та їхні метадані), семантичні мітки та моделі, а також серіалізовані знання у формі KG у другому вимірі.

У третьому вимірі семантичний рівень включає в себе функціональність, яку може реалізувати система озера семантичних даних для створення семантики «розумного міста», необхідної для ефективного SDM, яка реалізується різними методами в четвертому вимірі.

Семантичне маркування – це процес створення семантичних міток для джерела даних «розумного міста». У літературі автоматизація цього процесу також позначається STD або STI. Автори [15] визначають три різні напрямки цього завдання:

– керований схемою, тобто використання схеми кожної точки даних у наборі даних, наприклад;

– керований даними, тобто використання фактичних значень даних, що містяться в наборі даних, для призначення відповідних концепцій за допомогою статистики або на основі машинного навчання.

– класифікація керована метаданими, тобто використання всіх доступних додаткових фрагментів інформації про набори даних, які можуть сприяти

семантичному маркуванню даних, наприклад, коментарі або текстова документація даних.

В [16] розділити підходи далі на методи на основі машинного навчання, евристичні та інженерні методи.

Під семантичним моделюванням мається на увазі процес створення зв'язків між семантичними мітками, таким чином описуючи контекст джерела даних «розумних міст». На практиці семантичне моделювання часто залежить від точних семантичних міток, які можна розглядати як елементи семантичної моделі лише першого порядку.

В [17] вводять фазу, яка називається семантичним уточненням. Вони підтверджують думку про те, що автоматизоване створення семантичних міток і моделей є самоочевидною вимогою для будь-якої платформи OBDM «розумних міст». Незважаючи на те, що автоматизовані підходи можуть забезпечити початкове семантичне маркування та моделювання після аналізу схем джерел даних, помилково вибрані або відсутні мітки для концепцій матимуть серйозний вплив на етап семантичного моделювання «розумних міст». Таким чином, оператор-людина повинен мати можливість покращити якість семантичних міток, а також отриманої моделі вручну шляхом перевірки, підтвердження, виправлення, вибору або обміну неоднозначними або взаємовиключними поняттями, а також додавання нових понять, які не були включені спочатку.

В [18] описують перетворення, яке називається семантичним підняттям. Маючи набір семантичних міток і, але не семантичний, набір даних, функція семантичного підйому повертає семантично анотований набір даних із семантичними мітками для сутностей і атрибутів «розумних міст». Вважається семантичним мітками лише в їхній оригінальній пропозиції, однак семантичний підйом потенційно може бути застосований для включення цілих семантичних моделей «розумних міст», а також будь-яких інших описових метаданих. Наприклад, коментарів або документації з текстовими даними, що зберігаються в інструментах документації або Wiki. Попереднє перетворення джерел даних для включення семантичної інформації разом із необробленими значеннями, щоб підготувати їх для наступного семантичного доступу, обов'язково

відповідає підходу ETL, що суперечить принципу ELT, властивому озерам даних «розумних міст». При роботі з великими наборами даних це перетворення може бути складним і дорогим, особливо коли важлива жвавість даних.

1.5 Моделі метаданих для озер даних, які враховують семантику

Моделі метаданих для озер даних «розумних міст» були активною сферою досліджень в останні роки, і багато існуючих пропозицій розглядають семантику як фундаментальну проблему для управління гетерогенними джерелами даних.

MEDAL і його наступник goldMEDAL визначають вимоги до загальної моделі метаданих для озер даних «розумних міст» і розглядають семантичне збагачення як основну вимогу. Попередні пропозиції HANDLE, а також GEMMS уже дають змогу ручне семантичне маркування [19].

Інші підходи моделюють свої метадані безпосередньо в мові RDF як онтологію «розумних міст». Ранній підхід, де всі вихідні дані та метадані зберігаються у форматі RDF. Автори [20] обирають мережеве та кероване семантикою подання для своєї моделі, яке можна розширити зовнішніми KG. Вони описують іншу модель метаданих на основі онтології, адаптовану до аналітики на основі запитів. Тут особлива увага приділяється ключовим індикаторам ефективності та статистичним показникам, типовим для науки про дані, де користувач визначає цікаві показники під час запиту, а також пов’язані параметри аналізу з точки зору концепцій KG «розумних міст». Потім структура виконує необхідне виявлення даних та інтеграцію на вимогу, щоб забезпечити запитані індикатори.

Подібного підходу дотримуються автори [21], які використовують індикатори для персоналізованого дослідження озера даних «розумних міст». Індикатори зазвичай обчислюються на основі централізації зберігання даних відповідно до менш гнучкого підходу ETL, ніж традиційні озера даних «розумних міст». Крім того, експерти домену «розумних міст», які знають дані, що зберігаються в озері даних, зазвичай відрізняються від аналітиків даних, які визначають індикатори, і користувачів, які використовують індикатори. Таким

чином, автори дають змогу експертам предметної області збагачувати різнорідні джерела даних в озері даних «розумного міста» семантичними моделями з використанням онтологій домену, а також пропонують онтологію, яка використовується аналітиками для визначення показників і параметрів аналізу, з точки зору концепцій у семантичних моделях «розумних міст», а також формул для їх агрегування. Продемонстровано ефективність цієї моделі в середовищі «розумних міст», розширеної теоретичними міркуваннями про моделювання уподобань користувачів для забезпечення персоналізованого дослідження даних.

Потенціал семантичних технологій у Bosch вже широко використовується у виробництві. Розробка KG потребувала величезних початкових зусиль від компанії, але обіцяє довготермінове вирішення низки складних проблем. В [22] описують модель метаданих, подану як онтологія DC-PAC на основі рекомендацій W3C, зокрема, Data3 і онтологія PROV, Catalogue Vocabulary DCAT4, для керування каталогом даних «розумного міста» і походження. Отримані ресурси даних «розумного міста» збагачуються семантикою шляхом вирівнювання, анотування та збагачення вхідних даних концепціями DCPAC. Крім того, вони використовують нову концепцію, яка називається онтологічно керованими, самоадаптивними інтерфейсами, у яких графічний інтерфейс каталогу даних «розумного міста» може динамічно відтворювати інформацію з KG. Це корисно, оскільки зміни в базовій онтології не вимагають жодних змін у інтерфейсі.

1.6 Існуючі системи семантичних озер даних «розумних міст»

Існує декілька пропозицій щодо систем озер даних «розумного міста», які включають функції для SDM.

Goods запускає свій каталог даних, скануючи системи зберігання та журнали Google, щоб виявити, які набори даних існують. Він перевіряє їхній вміст і метадані та зіставляє їх із графом знань Google (GKG), щоб ідентифікувати сутності, і це, можливо, є найочевиднішою демонстрацією семантики потужності для керування даними «розумного міста».

Окрім моделі метаданих, GEMMS [19] є інструментом для автоматичного вилучення метаданих в озерах даних «розумного міста». Автори визнали, що часто в якості назви елемента схеми використовуються акроніми або синоніми. Використовуючи семантичні анотації, можна уникнути таких двозначностей і чітко визначити семантику елементів. «Constance» підтримує цю ідею і є однією з перших систем озер даних, яка підтримує семантичні мітки на рівні схеми джерела даних.

Apache ATLAS – це автономна структура керування даними для Hadoop, яка була успішно інтегрована в озера даних [23], що дає змогу семантичне маркування. goldMedal демонструє свою метамодель для озер даних за допомогою Apache ATLAS. Отже, вона не розглядається ні як окрема система озера даних, ні як метамодель, але може використовуватися як основна технологія, яку варто згадати.

Система виявлення даних Aurum має KG як центральний елемент. Система створює так званий корпоративний KG за один прохід через джерела даних, використовуючи підсумовування даних і хешування для фіксації зв’язків між ними. Aurum описує проблему виявлення даних за допомогою властивостей і зв’язків джерел даних, що зберігаються в озері даних «розумного міста». Властивості включають мітки схеми та показники якості, зв’язки можуть включати схожість і зв’язки первинного та зовнішнього ключа. У KG кожен вузол подає стовпці джерел даних, ребра подають зв’язки між двома вузлами, а гіперребра з’єднують будь-яку кількість ієрархічно пов’язаних вузлів, наприклад стовпців однієї таблиці або таблиць одного джерела даних «розумного міста». Aurum дає змогу відображати окремі записи, але припускає, що додаткові витрати на зберігання не компенсуються збільшенням виразності. Щоб обчислити вагові коефіцієнти, пов’язані з кожною властивістю, він враховує подібність вмісту, значення стовпців подібні, схожість схеми, назви атрибутів подібні, або існування зв’язку зовнішнього ключа між ними. Таким чином, основний алгоритм для створення гіперграфа включає два кроки: по-перше, кожне джерело даних профілюється, а по-друге, зв’язки обчислюються на основі профілів. Запити виконуються за допомогою мови SRQL, спеціально

розробленої для продукування KG, яка значно відрізняється від SPARQL. KGLac дотримується подібного підходу, оскільки це зовнішня служба, яка генерує центральний KG на основі вилучених профілів для керування всіма даними «розумного міста». Профілювання даних підтримується моделлю машинного навчання для подання кожного джерела даних у просторі вбудовування для виконання пошуку схожості між таблицями чи стовпцями без розкриття необроблених даних. Знову ж таки, у KG вершини подають джерела даних, тобто таблиці або стовпці, тоді як ребра подають зв'язки між цими вузлами. На відміну від Aurum, KGLac покладається на спеціальне розширення стандарту RDF. Спеціальна онтологія використовується для запитів за допомогою SPARQL, щоб дозволити навігацію та підтримку центральної KG. Система не має інтерфейсу користувача, але покладається на Python API для запитів до бази даних RDF із прямою інтеграцією в загальні канали обробки даних.

CoreKG – це служба з відкритим вихідним кодом, розроблена для озер даних для керування збереженими джерелами даних через REST API, яка складається з вилучення метаданих, їх збагачення, зв'язування та анотації. Зв'язок у CoreKG означає встановлення зв'язку із зовнішнім KG, таким як WikiData, а також з іншими джерелами даних «розумного міста». Це не повне озеро даних, оскільки воно не надає жодних можливостей зберігання чи обробки, але його функціональні можливості можна легко додати до існуючих озер даних за допомогою REST API.

Таблиця 1.1 – Порівняння озер даних із семантичним шаром

Система	**SL**	**SR**	**C/E**	**MI**	**IAL**	**TA**	**C**	**O**
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Goods	✓	✓	X	X	✓	X	X	X
Constance	✓	X	X	X	X	X	X	X
Aurum	✓	✓	✓	X	✓	X	X	✓
KGLac[47]	X	✓	✓	X	✓	✓	✓	X

Продовження таблиці 1.1

1	2	3	4	5	6	7	8	9
CoreKG[48]	✓	✓	✓	✓	✓	✓	✓	✓
BARENTS[49]	X	✓	✓	✓	X	✓	✓	X
Enterprise Knowldege Graph Plattform by Stardog at Bosch	✓	✓	✓	✓	✓	✓	✓	X
Semantic Integrator from the Semantic Web Company	?	✓	?	?	✓	✓	?	X
Anzo Semantic data lake	?	✓	?	?	✓	✓	?	X
Semantic Layer by AtScale	X	✓	X	✓	✓	✓	X	X
Semantic Layer Dremio LakeHouse Plattform	X	✓	X	✓	✓	✓	X	X

BARENTS [24] подає метод, заснований на онтології, який дає змогу аналітикам моделювати, як набори даних мають бути попередньо оброблені, щоб перетворити їх у значущі знання «розумного міста». Вони вводять онтологію не для виявлення даних, а явно для попередньої обробки. Експерти домену «розумного міста» можуть використовувати цю онтологію, щоб описати, які перетворення даних потрібно застосувати до необроблених даних із вихідних систем, щоб зробити їх придатними для використання у своїх випадках використання.

1.7 Порівняння та обговорення семантичних озер даних «розумних міст»

Ми визначаємо критерії для порівняння та перевірки повноти існуючих систем, узагальнених у табл. 1.1. Таблиця обмежена конкретними системами озер даних і не містить переліку підходів, які просто пропонують модель метаданих «розумного міста»:

1. «SL» – семантичні мітки вказують на можливість зв’язувати записи схеми з екземплярами у зовнішньому КГ, або КГ домену, який потрібно завантажити, або онлайн-КГ, наприклад schema.org.

2. «SR» – семантичні зв’язки, тобто будь-яка форма ієрархічних, загальних або попередньо визначених семантичних зв’язків, семантичні зв’язки між наборами даних, наприклад, щодо походження чи управління. Слід зазначити, що жодна з існуючих систем озер даних не забезпечує розширених функцій моделювання, таких як ті, що зображені на рис. 1.1.

3. «CE» – центральні графіки KG проти зовнішніх KG. Тут розрізняємо, чи сама базова метамодель моделюється безпосередньо як KG, тобто зв’язки обмежені тими, що визначені в цьому KG. Більш гнучка система дозволила б використовувати зовнішню КГ, і будь-які поняття та зв’язки з цієї КГ можна було б використовувати в семантичній моделі «розумного міста». Позначене поле вказує на те, що модель метаданих використовує центральний, попередньо визначений KG.

4. «MI» – сумісність метаданих Це вказує на те, чи розглядали автори можливість імпорту чи експорту метаданих. тобто серіалізація семантичної моделі. Це тісно пов’язане з критерієм C, оскільки серіалізовані моделі легше переносяться на різні системи.

5. «IAL» – початкове автоматичне семантичне маркування вказує на функціональні можливості для автоматичного створення семантичних міток зі схеми під час прийому.

6. «TA» – технічна абстракція. Це вказує на те, чи автори явно розглянули технічну абстракцію семантичного моделювання як вимогу для підвищення зручності використання для незнайомих користувачів. Це означає, що технічні деталі приховані від користувачів, наприклад, користувачеві не потрібно знати RDF.

7. «C» – сумісність із технологією Semantic Web. Цей критерій перевіряє, чи семантичні мітки змодельовані мовою семантичної мережі, наприклад, RDF, OWL.

8. «О» – відкритий код. Цей критерій перевіряє наявність доступної реалізації «розумного» застосунку з відкритим кодом.

1.8 Висновок до першого розділу

В першому розділі кваліфікаційної роботи освітнього рівня «Магістр» описано міські джерела семантичних даних. Проаналізовано стан та досліджень в царині озер даних «розумних міст». Наведено ключові терміни та визначення управління семантичними даними в озерах даних. Розроблена таксономія управління семантичними даними. Розглянуті моделі метаданих для озер даних, які враховують семантику. Описані існуючі системи семантичних озер даних «розумних міст». Виконано порівняння та обговорення семантичних озер даних «розумних міст».

2 СЕМАНТИЧНЕ МОДЕЛЮВАННЯ ДЛЯ ІНТЕГРАЦІЇ ДАНИХ «РОЗУМНИХ МІСТ»

Семантичне моделювання спрямоване на надання семантичного опису джерел даних «розумних міст» і не обмежується озерами даних, але може бути застосоване до керування даними в цілому. Тим не менш, методи збагачення метаданих дуже актуальні в цьому контексті, оскільки вони підвищують зручність використання озера даних «розумного міста». Семантична модель діє як міст між джерелом даних «розумного міста» [25] і семантично багатим KG, який можна застосувати для інтеграції даних або доступу до даних на основі онтології. Спочатку зосередимося на підходах до створення семантичних моделей напівавтоматично, потенційно підтримуваних людським введенням у графічному інтерфейсі користувача (GUI). Визначення семантичних моделей вручну можливе, але це неможливо для великої кількості різнорідних наборів даних, як очікується в озерах даних «розумного міста». Таким чином, автоматизація генерації семантичних міток і моделей допомагає зменшити ручні зусилля в системі озера семантичних даних [26]. Дослідимо поточний стан таких пропозицій і проведемо їх порівняння та обговорення.

2.1 Системи семантичного моделювання «розумних міст»

Розглянуті в попередньому розділі системи, зосереджені на семантичному маркуванні, тобто дають змогу анотувати набори даних елементами з внутрішніх і зовнішніх онтологій. Семантичне моделювання є наступним кроком у семантичному збагаченні при якому пов’язані елементи онтології збагачуються додатковою семантикою для відображення зв’язків, поданих у наборах даних «розумного міста» (див. рис. 1.1).

ESKAPE – це платформа семантичних даних, яка обробляє повний процес керування даними від прийому до вилучення, підтримуючи різнорідні джерела даних. Вона пропонує інтуїтивно зрозумілий інтерфейс, де користувачі можуть створювати семантичні моделі «розумного міста». У [10] автори повідомляють

про інтеграцію ESKAPE в систему озера даних у умовах «розумного міста». Замість того, щоб просто скидати необроблені дані безпосередньо в озеро даних, вони використовують ESKAPE під час прийому даних, щоб увімкнути семантичну анотацію наборів даних для виконання семантичної інтеграції даних «розумного міста». Щоб об'єднати визначені семантичні моделі, ESKAPE також дає змогу інтегрувати зовнішні KG, зокрема, WordNet, BabelNet або доменоспеціалізовані онтології «розумного міста». Елементи KG можна повторно використовувати під час процесу моделювання і, отже, слугувати якорями при зв'язуванні різних різних семантичних моделей і наборів даних. Автори [27] – це інструмент інтеграції даних, який дає змогу користувачам отримувати дані, а також словник, визначений як онтологія OWL, для створення відповідної семантичної моделі «розумного міста». «Karma» підтримує декілька форматів даних і пропонує початкову семантичну модель, яку користувачі можуть редагувати за допомогою графічного інтерфейсу користувача. «Karma» є важливою прогресивною розробкою, яка однією з перших запровадила ключові технології, зокрема, умовні випадкові поля (CRF) і алгоритм вирішення проблеми дерева Штейнера для вивчення семантичних міток і виявлення зв'язків між елементами схеми джерела «розумного міста». Після завершення семантичного моделювання на основі онтології «Karma» інтегрує набори даних на основі визначеної семантичної моделі «розумного міста». Отримані семантичні моделі можна експортувати в файл RDF. На відміну від ESKAPE, «Karma» надає початкові пропозиції семантичного маркування, але також вимагає від користувачів глибоких знань про технічні деталі, наприклад, формат RDF. Крім того, він фокусується на автоматичному створенні семантичних моделей «розумного міста» на основі фіксованої онтології для кожного випадку використання, де онтологія не розвивається з часом. ESKAPE стверджує, що він простіший у використанні, оскільки він приховує технічні деталі, однак немає доступної реалізації для порівняння зручності використання.

Нещодавно платформа з відкритим вихідним кодом PLASMA з'явилася на основі ESKAPE, а також декілька інших помітних робіт, які більш детально описані нижче. Дослідники запровадити фазу, яка називається семантичним

уточненням (див. рис. 1.3), на якій людина відповідає за покращення якості семантичної моделі шляхом ручної перевірки, підтвердження, виправлення, вибору або обміну поняттями. PLASMA постачається з рекомендаційною структурою семантичних концепцій, щоб пропонувати семантичні концепції, отримані з загальнодоступних KG, а також із будь-якого власного джерела, наприклад, KG «розумного міста». Семантичні моделі можна створювати, видаляти, шукати та експортувати як RDF у синтаксисі Turtle. Центральне місце у всій структурі займає ідея онтологій, що постійно розвиваються. Оскільки джерела даних потрібно додати або оновити, створену семантичну модель необхідно розширити або адаптувати для відображення цих змін. Щоб вирішити цю задачу, автори [28] пропонують підхід, що включає еволюційний KG, який складається з внутрішнього зростаючого універсального KG і відображень для певного джерела даних «розумного міста».

Користувач, який створює семантичну мітку, має змогу визначати та використовувати концепції та зв’язки у своїх семантичних моделях, які раніше були невідомі KG. Це, наприклад, у випадку, якщо концепція, яку користувач хоче використовувати, відсутня в поточних концептуалізаціях «розумних міст». Ці нові знання повинні бути інтегровані в онтологію універсального KG «розумного міста». Для цього було розроблено стратегію контрольованого та еволюційного розвитку KG, яка оцінює зв’язки та концепції, що вибирає або вводить користувач, і перевіряє, якщо користувач вносить суперечності в поточну KG. Фреймворк визначає додаткові зв’язки між концепціями, які покращують щільність основної онтології «розумного міста», і містить логіку перевірки для усунення невідповідностей.

Система управління простором даних On-demand Data INtegration (ODIN) дає можливість інтегрувати дані шляхом віртуального запиту до різнорідних джерел даних «розумного міста», заснованих на KG. ODIN автоматично витягує схеми з напівструктурованих джерел даних, переводить їх у канонічну модель даних, вирівнює їхні схеми та генерує з них цільові метадані. У контексті семантичного вдосконалення він покладається на підтримуваний процес зворотного зв’язку з користувачем, покладаючись на поступове злиття

семантики в онтологію простору даних «розумного міста». ODIN підтримує розширений візуальний запит mech-Web, який є програмою з відкритим кодом для інтерактивної візуалізації онтологій «розумного міста». Цей інтерфейс запиту, опосередкований онтологією, дає змогу вибирати цікаві вузли з будь-якої доступної семантичної моделі, позначаючи їх графічно, щоб створити запит SPARQL. Джерела даних «розумного міста», пов’язані з позначеними вузлами, подають точку підключення до наборів даних, обгортки кодують запит, щоб видобути свої дані та відкрити відношення першої нормальної форми своїх схем.

2.2 Автоматизоване присвоєння семантичних міток «розумного міста»

Техніка завантаження початкової онтології – це реляційні бази даних і безпосереднє встановлення відображення між базами даних і створеними концепціями присутні в декількох системах. Наприклад, Optique, а також його наступник OnTop надають такі можливості. Ранній контрольний показник для оцінки якості згенерованих відображень у контексті доступу до даних на основі онтології (OBDA) поданий RODI [29]. Впливові роботи включають SemanticTyper [30], які застосовують евристичні правила та підхід на основі TFIDF для текстових даних і перевірку статистичної гіпотези Колмогорова Смирнова для числових даних, а також Domain-Independent Semantic Labeler (DSL) [31], який подає підхід машинного навчання щодо мір подібності.

Як згадувалося раніше, зацікавлені читачі направляються до статті [16] для ретельного та всебічного обговорення пропозицій щодо STD, включаючи численні розробки на основі глибокого навчання. Тут хочемо включити один більш свіжий підхід під назвою Tab2KG, який дотримується зовсім іншого підходу в тому сенсі, що він може інтерпретувати раніше невидимі дані на основі KG довільної області «розумного міста». Цікаво, що Tab2KG використовує RML для серіалізації семантичних міток, можливо корисних для виконання OBDA. Вони вводять так звані семантичні профілі для онтологій домену та таблиць даних «розумного міста», щоб полегшити ефективну інтерпретацію семантичної таблиці. Для створення профілю домену KG домену, який містить

репрезентативні значення для зв’язків типів даних у цільовому домені, перетворюється на вектор ознак, що містить набір статистичних даних, обчислених за допомогою всіх літералів, що надходять із набору зв’язків типів даних «розумного міста». Для створення профілю таблиці даних обчислюється вектор ознак описових значень, який включає типи даних і основні статистичні дані, повнота, середнє значення, стандартне відхилення, асиметрія, гістограми тощо. Вибір обумовлений очікуваною ефективністю функції для інтерпретації семантичної таблиці, тобто відповідністю домену та профілів таблиці даних «розумного міста», і може бути розширений, щоб включити відповідні специфічні для домену характеристики. Профілі генеруються автоматично та описуються за допомогою словників DCAT2 і SEAS3, щоб забезпечити їх повторне використання. Потім Tab2KG використовує профілі таблиці домену та даних для визначення семантичного типу за допомогою нового одноразового підходу до навчання. На відміну від предметно-орієнтованих підходів DoDuo [14], Tab2KG використовує сіамську мережу, яка узагальнює зв’язки типів невидимих даних «розумного міста», індукуючи метрику, яка подає незалежну від домену подібність між двома векторами вхідних ознак, наприклад, між невідомим і відомим зразком. Подібність між стовпцем і зв'язком типу даних передбачається на основі досвіду подібності інших профілів, отриманого раніше. Маючи набір варіантів зіставлення стовпців із відповідними оцінками подібності для кожного стовпця в таблиці вхідних даних «розумного міста» [32], вони жадібним чином відображають кожен стовпець таблиці на відношення типу даних.

2.3 Автоматизована генерація семантичної моделі «розумного міста»

У той час як попередні підходи зосереджені виключно на семантичних мітках, існують також деякі, які спрямовані на побудову цілих семантичних моделей «розумного міста», деякі також включають генерацію семантичних міток, і останнім часом ця сфера привернула все більший інтерес.

Автори [33] подають PGM-SM, який використовує імовірнісні графічні моделі для створення семантичних моделей «розумного міста». На основі попередньо згенерованих семантичних моделей і виявлених семантичних міток умовне випадкове поле навчається розрізняти хороші та погані семантичні моделі. Потім граф використовується для ідентифікації можливих семантичних моделей, а найкращі k з них вибираються за допомогою функції підрахунку балів на основі імовірнісних графічних моделей. У подальшій роботі вони визнають, що, незважаючи на гнучкість у виборі цільової онтології, цей підхід вимагає від користувачів позначати достатню кількість джерел даних «розумного міста», перш ніж системи зможуть досягти високої продуктивності, і ця проблема тим глибша, чим більшою є цільова онтологія. З цієї причини вони зміщують контрольовану проблему до використання великомасштабних KG, таких як DBPedia та WikiData. Для цього вони подають новий імовірнісний підхід для автоматичного створення семантичних описів таблиць Вікіпедії, використовуючи гіперпосилання та наявні знання у Вікіданих для побудови графіка можливих зв’язків у таблиці та її контексті. Щоб оцінити ефективність методу, його оцінюють на наборі даних із завдання «SemTab2020», а також на іншому наборі даних із двохсот п’ятидесяти таблиць Вікіпедії з їхніми семантичними описами, створеними за допомогою онтології Wikidata.

Іншими відомими роботами, пов’язаними з проблемою SemTab2020, є MantisTable, bbw і MTab4WikiData [34], кожна з яких спрямована на автоматичне створення семантичних моделей для даних з Вікіпедії або Вікіданих.

У дусі CoreKG, автори [35] подають ASMaaS, автоматичне семантичне моделювання як послугу, яке є зовнішньою службою для постачальників озер даних і користувачів для створення семантичної моделі для своїх наборів даних «розумного міста». У статті описується сервіс-орієнтована архітектура, включаючи анотування навчальних даних, навчання моделям машинного навчання та прогнозування точної семантичної моделі для нових джерел даних «розумного міста».

Дослідники [35] подають підхід під SeMi (англ. SEmantic Modeling machine) на основі графових нейронних мереж (GNN), а також SemanticTyper від

Ramnandan охоплення процесу семантичного моделювання. Вони використовують застосування пов’язаних сховищ даних як фонових знань. Автори замінюють ручне виділення складових ознак, наприклад, складних шаблонів графів для подання семантичних зв’язків різної довжини, нейронною мережею графів, яка автоматично вивчає приховані ознаки для сутностей і властивостей «розумного міста», кодуючи їх у векторному просторі, використовуючи локальні структури сусідства всередині зв'язаний графік даних. У конвеєрі обробки Semantic Model Builder створює початкову семантичну модель. Пропоноване дерево Штейнера в графі містить найкоротший шлях для з’єднання класів семантичного типу. Однак це не обов’язково виражає правильний семантичний опис цільового джерела «розумного міста». З цієї причини необхідний процес уточнення, щоб визначити більш точну семантичну модель. Для уточнення семантичної моделі «розумного міста» застосовано автоматичний процес, на відміну від процесу в PLASMA (див. рис. 1.3) вимагає підготовки наборів даних як вхідних даних для моделі глибинного навчання. Основна мета нейронної мережі графа – реконструювати зв’язані межі даних за допомогою прихованого подання сутностей і властивості.

Автори [36] подають VC-SLAM, корпус, який дає змогу оцінювати та порівнювати підходи до семантичного маркування та моделювання в різних методологіях. Він містить понад сто наборів даних, що складається з міток, даних і метаданих, а також відповідних семантичних міток і семантичної моделі «розумного міста». Семантичні мітки та семантична модель уточнюються експертами-людьми вручну за допомогою онтології, яка була спеціально створена для «розумного міста». У той час як перші два напрямки в процесі семантичного маркування, тобто керовані даними та схемами, були розглянуті в більшості опублікованих досі робіт, третій напрямок поки що привернув лише мало уваги. Семантичне маркування на основі метаданих стосується всіх доступних додаткових частин інформації в наборі даних, які можуть сприяти семантичному маркуванню даних. Наприклад, структуровані дані, як-от стандарти CKAN, і будь-які додаткові метадані в базі даних, як-от коментарі або документація з текстовими даними, що зберігаються в інструментах

документації або Wiki, можуть використовуватися для вдосконалення процесу семантичного маркування. Вони намагалися порівняти SeMi зі своїми наборами даних, але не змогли через те, що вони створили нову онтологію домену, яка недостатньо велика для навчання GNN із фоновими даними. З цієї причини вони включають додаткову документацію текстових даних, яку можна використовувати.

У [37] подано DocSemMap новий підхід, який використовує документацію текстових даних наборів даних як додаткове джерело для створення семантичних відображень, що використовує методи обробки природної мови (NLP) разом зі структурованими даними для виконання семантичного моделювання. Вони показують, що, особливо в порівнянні з методами, керованими даними, існують різні випадки, коли DocSemMap може відображати концепції «розумного міста», які інакше не ідентифікувалися б. Таким чином, існує великий потенціал для виконання семантичного відображення на основі документації текстових даних.

У [38] подає Relational Natural Language Inference (RLNI), що має на меті забезпечити зрозуміле дослідження даних на озерах даних «розумного міста». Вони звертаються до того факту, що багато фреймворків намагаються створювати показники подібності, які не надають чіткого пояснення того, як визначений набір даних пов’язаний із заданою ціллю. Щоб вирішити цю проблему, вони зосереджуються на основі висновку природної мови (NLI), який має обмежений набір зв’язків, тобто еквівалентність, пряме втягнення, зворотне втягнення, заперечення, чергування, прикриття та незалежність, і забезпечує неконтрольоване аналізу, а також контрольована альтернатива, яка потребує навчальних даних на основі попередньо навчених мовних моделей. Цей підхід не включає семантичну виразність, притаманну повноцінним онтологіям, а скоріше забезпечує семантику та пояснюваність проблеми дослідження в озерах даних, яка визнана недостатньою в порівняних роботах [39].

В [40] досліджено проблему під іншим кутом зору. Зокрема, проаналізовано SQL-запити, написані аналітиками даних, які вже розуміють семантичні зв’язки різнорідних джерел даних. Їхній прототип під назвою Pharos підключається до існуючого сеансу SQL будь-якої програми, яка має інтерфейс

JDBC, і записує та аналізує всі запити у фоновому режимі. Він витягує фрагменти знань із запитів SQL і подає їх у KG на основі RDF, для якого розроблено нову онтологію «розумного міста». Приклади включають конструкції зовнішнього або первинного ключа, отримані із запитів JOIN або записів журналу у формі «атрибут salary has been casted to the data type long», з яких можна зробити висновок, що стовпець salary може бути підсумований.

2.4 Порівняння та обговорення семантичних моделей озер даних «розумних міст»

Порівняння ділиться на дві частини. Спочатку розглядаємо повні платформи моделювання, які включають методи семантичного маркування та моделювання, а також графічний інтерфейс користувача, який дає йому змогу адаптувати створені артефакти. Для категорії «розумного міста» визначено декілька критеріїв:

1. «IAM» – це початкова автоматична семантична модель. На відміну від IAL, це позначає тут функціональні можливості для автоматичного створення цілих семантичних моделей зі схеми після прийому джерела даних. Це включає перші три фази на рис. 1.3.

2. «EX» – це підтримка зовнішніх KG, що визначає чи здатний цей підхід обробляти довільні KG чи прив’язаний до попередньо визначеного KG.

3. «M» – це зрілість, що перевіряє, чи було повідомлено про успішне використання системи в проекті «розумного міста» великих даних.

4. «C» – це сумісність із технологією «Semantic Web». Цей критерій перевіряє, чи семантичні мітки змодельовані мовою семантичної мережі, наприклад, RDF або OWL.

5. «O» – доступне програмне забезпечення з відкритим кодом.

Порівняння наведено в таблиці 2.1. Слід зазначити, що всі згадані підходи зосереджені лише на реляційних/табличних або принаймні напівструктурованих даних, включаючи JSON та XML. Неоднорідність, притаманна застосункам

великих даних «розумного міста», ще далеко не охоплена підходами автоматичного семантичного моделювання.

Таблиця 2.1 – Порівняння платформ семантичного моделювання «розумних міст»

Платформа	**IAM**	**SR**	**EX**	**M**	**C**	**O**
ESKAPE [2]	✓	X	✓	✓	X	X
KARMA [74]	X	✓	✓	✓	✓	✓
PLASMA [75]	✓	✓	✓	✓	✓	✓
ODIN [55]	✓	✓	✓	X	✓	✓

З точки зору ручного моделювання на основі графічного інтерфейсу користувача PLASMA демонструє зріле рішення, яке підтвердило свою придатність у виробництві з багатьма вбудованими функціями, які можна протестувати публічно.

Виявлення семантичного типу останнім часом набуло імпульсу досліджень, але ця сфера ще далека від справді зрілого рішення. Найновіші рішення, засновані на глибокому навчанні, досягають вражаючої точності при оцінці на конкретних тестових наборах даних «розумних міст», але враховуючи той факт, що будь-який набір даних із мітками, який виконує контрольоване навчання, наприклад, «GitTables 14», може містити лише обмежену кількість семантичних WikiTables або типів, його практична корисність стає дуже обмеженою. Зрештою, вони страждають від припущення про закритий світ «розумного міста», як у випадку з DoDuo [14], їх остаточна модель може розрізняти лише 255 різних типів.

Щоб зробити модель чутливою до іншої доменної онтології «розумного міста», окрім WikiData, потрібно навчити модель на вручну позначеному наборі даних, що містить відповідні семантичні типи. Крім того, оскільки вони

використовують мовну модель, виявлення поки що погано працює з числовими даними, і більшість підходів обмежені в роботі з багатовимірними та великими наборами даних «розумного міста». Знову беручись за DoDuo, його навчили на таблицях із явним максимумом у 10 стовпців і 32 маркери в послідовності. Навпаки, підхід до вбудовування в DAGOBAH або евристичні стратегії MTab і JenTab [41], імовірно, зможуть передбачати довільні семантичні типи, оскільки тут семантичні типи витягуються віконними запитами, які не обмежуються конкретними цілями.

Однак обидві стратегії оцінюють свої алгоритми на основі фрагментів міждомених KG, таких як WikiData та DBpedia, і не розроблені спеціально для роботи зі спеціалізованою концептуалізацією, такою як доменна онтологія «розумного міста». Це особливо важливо для застосовності до озер даних, оскільки вони часто використовуються в певній сфері, наприклад, «розумне» виробництво або енергетичний сектор [42], а для глибокого розуміння даних потрібні спеціальні знання. З точки зору автоматизованого моделювання, підхід SeMi, який покладається на вбудовування KG, обчислені за допомогою GNN, є передовим рішенням, але його визначення семантичного типу, перший і найважливіший крок у конвеєрі «розумного міста», не є загальним, а адаптованим до конкретних наборів даних.

2.5 Висновок до другого розділу

В другому розділі кваліфікаційної роботи досліджено системи семантичного моделювання «розумних міст». Розглянуто автоматизоване присвоєння семантичних міток «розумного міста». Висвітлена автоматизована генерація семантичної моделі «розумного міста». Виконано порівняння та обговорення семантичних моделей озер даних «розумних міст».

3 ДОСТУП ДО ДАНИХ «РОЗУМНИХ МІСТ» НА ОСНОВІ ОНТОЛОГІЇ

За допомогою семантичних міток і семантичних моделей доступний семантичний опис джерел даних; тому доцільно використовувати цю семантику для інтеграції даних і запитів «розумних міст». Це тема доступу до даних на основі онтології (OBDA), яка була розроблена з середини 2000-х років [6]з метою полегшення доступу до різних типів джерел даних, використовуючи онтологію як загальну модель даних. Налаштування можна описати як набір існуючих джерел даних, які формують рівень даних «розумних міст», і мета полягає в тому, щоб побудувати на його основі службу, спрямовану на подання концептуального уявлення про дані для споживачів. Онтологія описує предметну область і зазвичай розробляється незалежно від рівня даних.

Проблемою в цьому сценарії є поєднання різних рівнів абстракції та формалізмів. У той час як онтології моделюють концептуальний рівень, використовують припущення відкритого світу та базуються на логіці опису, джерела даних описуються схемами на логічному рівні, які зосереджуються на структурі даних і використовують припущення закритого світу. Поставляючи запит до онтології, основна перевага OBDA полягає в тому, що користувачі можуть запитувати різні джерела даних без необхідності знати, як дані організовані та де вони зберігаються.

Автори [43] визначають два основні підходи до використання OBDA:

1) матеріалізація, тобто всі вихідні набори даних перетворюються в загальний формат даних, наприклад, RDF або реляційні таблиці, і зберігаються в одному сховищі даних;

2) переклад на вимогу, або переписування, запитів, зазвичай виражених у SPARQL на необхідну мову запитів джерел даних.

Доцільно наголосити на застосуванні другого підходу до великих даних, оскільки, масштабованість є суттєвою вимогою. Підходи до матеріалізації можуть бути складними та дорогими при роботі з великими наборами даних «розумного міста». Отже, лише підходи до перекладу запитів можливі в озерах даних «розумного міста», хоча той факт, що він значно складніший, оскільки

вимагає глибокого теоретичного розгляду. Це частина більшої області об’єднання даних [6], яка вирішує проблему рівномірного доступу до дають змогу, можливо різнорідних джерел даних «розумного міста», відображаючи їх в уніфіковану схему.

Загальна модель даних, яка використовується для запитів, також може бути виражена іншими мовами схем. Більш ранні підходи до інтеграції даних використовували реляційну модель і переводили джерела даних у загальну реляційну схему. Переписування запитів у цьому реляційному налаштуванні було добре вивчено, тому розглядаємо більш складний сценарій перекладу запиту з онтологіями над різнорідними джерелами.

Трансляція запитів [44] тісно пов’язана і необхідна для будь-якої ефективної структури OBDA різноманітність різнорідних джерел даних в озері даних «розумного міста». В OBDA трансляція запитів на вимогу розкриває вміст джерел даних як RDF-трійки, використовуючи класи та предикати з KG. RDF-трійки не матеріалізуються, вони є частиною віртуального графа знань (VKG), це означає, що дані залишаються в джерелах даних, а не зберігаються в якійсь загальній базі даних. Система VKG має компоненти (див. рисунок 3.1):

– запити, які описують потреби користувача в інформації;
– онтологія з класами та властивостями;
– відображення джерел даних;
– колекція даних джерела.

W3C опублікував рекомендації щодо мов компонентів: SPARQL, OWL 2 та R2RML відповідно.

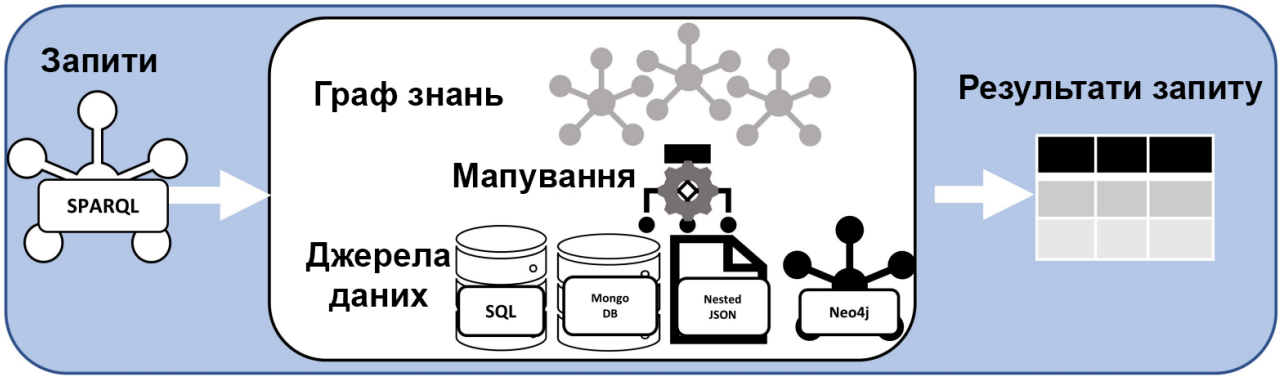

Рисунок 3.1 – Доступ до даних на основі онтології [6]

Створення відображення та керування ним, ймовірно, є найскладнішим завданням під час проектування OBDA, оскільки відображення визначає семантику джерел даних [6] у термінах онтології та заповнює зазвичай великий концептуальний розрив між вихідною схемою та онтологією. Застосовуючи методи подання знань і автоматизованих міркувань, система OBDA використовує онтології та відображення для переформулювання запитів SPARQL у запити, що стосуються джерела. Формальна структура подана в [3], включаючи відповіді на запити, керування відображенням та аналіз, а також розширення класичної структури OBDA. Тут опускаємо теоретичні деталі переписування запитів і зосереджуємося на практичному застосуванні для інтеграції даних у режимі великих даних «розумного міста».

Комплексний огляд методів перекладу запитів, проведений в [36], включаючи класифікацію, карту шляхів перекладу різними мовами та історичну шкалу часу. У той час як дослідники десятиліттями шукали вирішення задачі подання даних, логіка, ER та реляційний підхід, XML, графове подання, NoSQL, їх робота підтверджує думку, що жодна схема подання даних не підходить для всіх випадків використання.

Дослідники визначають декілька цікавих недоліків. По-перше, вони визнають, що багато методів машинного перекладу не можуть підтримувати складніші операції, такі як різні типи з’єднань і тимчасові функції в SQL або порожні вузли, групування та зв’язування в SPARQL. Потім вони стверджують, що існує багато добре обґрунтованих і визначених структур перекладу запитів, від процесу перекладу запиту до різних стратегій оптимізації.

Однак вони навряд чи подаватимуть запити реального світу, а переважно прості запити. Методи перекладу, орієнтовані на випадки використання, були б більш корисними для виявлення корисних шаблонів запитів і оцінки методів перекладу та оптимізації даних «розумних міст». Стверджується, що існує велика різноманітність систем оцінювання, які використовуються кожним із методів перекладу запитів, і потреба в унікальному стандартизованому тесті, що спеціалізується на оцінюванні аспектів перекладу запитів «розумних міст». У заключних зауваженнях обговорюються кандидати на «універсальну» мову

запитів «розумного міста». Ґрунтуючись на своїх висновках, досліджуючи різні методи перекладу запитів, можемо зробити висновок, що SQL і SPARQL як найбільш підходящі мови для роботи в якості «універсальної» мови для реалізації інтеграції гетерогенних даних «розумного міста».

3.1 OBDA над реляційними даними «розумних міст»

Деякі з найперших систем, що виконують OBDA, використовують Mastro, UltraWrap і MorphDB [6]. Optique була першою успішною системою OBDA, розробленою для сценаріїв великих даних, в тому числі «розумних міст». Його базова архітектура в основному використовує Ontop для керування доступом до даних за допомогою відображень R2RML, а також Exareme, дослідницький прототип, який діє як компонент виконання запитів у фоновому режимі, що обробляє великомасштабні завдання обробки даних «розумних міст». Основним недоліком цих ранніх систем є відсутність горизонтальної масштабованості.

Імовірно, що фреймворк Ontop для OBDA, який містить шістнадцять систем OBDA [45], буде найбільш зрілим відкритим вихідним кодом, що пройшов довгий шлях з моменту свого започаткування в 2009 році [6]. Він є результатом активних досліджень та був використаний у багатьох академічних і промислових проектах. Оскільки Ontop перейшов до підтримки рекомендацій W3C для SPARQL і R2RML, виникли нові виклики, які вимагали розробки основної структури даних, яка називається проміжним запитом «IQ», структурою даних на основі алгебри, яка об’єднує як SPARQL, так і реляційну алгебру. «IQ» є уніфікованим поданням як для запитів SPARQL, так і для запитів SQL із відображення. Коли перетворення запиту, його переписування та розгортання, завершено, вираз «IQ» перетворюється на SQL і виконується базовою системою керування реляційною СУБД. OnTop прагне перенести майже всю обробку запитів на СУБД і виконує лише найвищу проекцію, яка зазвичай перетворює значення бази даних у терміни RDF «розумного міста». Складні методи оптимізації, такі як усунення надлишкових з’єднань, використання первинних і зовнішніх ключів, ефективне усунення дублікатів [46] і опускання

з’єднань на рівень даних, а також використання обмежень цілісності для підвищення повноти запиту та продуктивності [6] було ретельно вивчено протягом останнього десятиліття, щоб переконатися, що запити можуть ефективно оброблятися базовою СУБД.

Завдяки нещодавньому розширенню Ontop4theWeb [6] фреймворк тепер можна оцінювати в режимі реального часу інформацію щодо веб-даних за допомогою SPARQL, не матеріалізуючи його спочатку в СУБД. Ontop4theWeb реалізує підхід для створення запитів SPARQL поверх веб-даних, які не є RDF, в режимі реального часу. Щоб досягти цього, оператори віртуальної таблиці вбудовані в SQL-запити, які включені в зіставлення R2RML. Вони вказують, яку частину та джерело веб-даних буде отримано та як вони будуть зіставлені з віртуальними термінами RDF. Специфікація цих відображень створює звичайні накладні витрати, але вони повинні бути. При цьому вказується лише один раз, якщо схема не змінюється. Незважаючи на те, що явна матеріалізація даних не виконується, і можна отримати відповіді веб-інтерфейсів API та NoSQL JSON API, Ontop4theWeb все ще покладається на реляційну модель даних і покладається на систему реляційної бази даних у backend.

Ontopic Studio є комерційним допоміжним продуктом Ontop, який надає графічний інтерфейс для полегшення створення та пошуку відображень за допомогою редактора відображень. Фреймворк Ontop нещодавно привернув інтерес до застосування в озерах даних [6] показують успішну інтеграцію в озеро даних «розумного міста» шляхом додавання семантичного рівня на основі фреймворку Ontop.

Chimera Suite [47] розширює рівні системи OnTop за допомогою двох компонентів:

1. Ontop$Spark$ – застосунок, який дає змогу OnTop виконувати OBDA на реляційних даних за допомогою Apache Spark, ефективно перетворюючи запити SPARQL на SparkSQL.

2. Кінцева точка JDBC для підключення Spark Thrift Server до OnTop, який може направляти запити SparkSQL до файлової системи HDFS.

Дослідники вказують на поточний розрив між технологіями великих даних і семантичним світом «розумних міст». З одного боку, інженери знань працюють з OWL і RDF для створення графів знань «розумних міст». З іншого боку, інженери обробки даних визначають процеси ETL, щоб надати вихідні дані у формі, яку можуть використовувати науковці з даних. Потім дослідникам даних потрібно поєднати ці дві точки зору, щоб отримати результати аналітики, які можна інтерпретувати. Щоб вирішити цю задачу було запропоновано всеохоплюючий конвеєр, де дослідники даних пишуть запити SPARQL і надсилають їх Jena Fuseki у Jupyter Notebook за допомогою бібліотеки PySPARQL, щоб отримати семантично збагачену відповідь.

Використовуючи відображення та онтологію, запит переписується в SQL. OntopSpark отримує доступ до даних, що зберігаються як файли Parquet у HDFS, за допомогою Apache Spark. Відповідь SPARQL, отримана від OntopSpark, додатково збагачується за допомогою графа знань, а остаточні результати надсилаються назад у блокнот Jupyter. В [47] описано реальне розгортання в найбільшій публічній компанії Італії для промислових досліджень в енергетичному секторі «розумних міст». За цим сценарієм було розроблено KG, що містить близько семи мільйонів потрійних елементів, що подають міський район Мілана, де зберігається топологія мережі та вся інформація про різні частини електричного обладнання «розумного міста». Chimera Suite дає змогу команді з обробки даних робити аналітичні прогнози, надсилаючи запити як до KG, так і до сховища великих даних «розумного міста». Інженери знань можуть отримати доступ до графа знань та отримати результати аналізу даних.

Ontario [48] – це механізм обробки запитів над різнорідними джерелами даних в озері даних «розумного міста», що використовує VKG. Ontario перекладає запити з глобального механізму запитів, наприклад, SPARQL, до основного власного механізму запитів джерел даних «розумного міста». Таким чином, система перетворює необроблені дані в RDF на вимогу, застосовуючи правила відображення на основі R2RML. Ontario використовує спеціальні розширення стандартів RDF і SPARQL, де відповідність стандартам W3C.

Obi-Wan [49] створено на основі посередника Tatooine для різнорідних джерел і Graal, інструментарію для відповідей на запити в базах знань «розумних міст». Цей підхід ґрунтується на зіставленнях глобального локального вигляду (GLAV) на відміну від глобального перегляду (GAV), які використовуються для, наприклад, у Ontop, MorphDB і Mastro. Відображення GLAV вважаються гнучкішими в сценаріях з еволюцією джерел даних [50]. У зіставленнях GAV зміна джерела даних може вимагати зміни глобальної схеми та всіх зіставлень «розумного» джерела. За допомогою відображень GLAV джерела описуються незалежно одне від одного щодо глобальної схеми.

Це також було визнано авторами [51], які пропонують структуру для інтеграції даних, повністю засновану на графах, що використовують відображення GLAV. Задіяно прототип алгоритмів перезапису включено в ODIN. Пропонована мова запитів базується на покриттях графа, що подає глобальну схему, і не вимагає від користувачів визначення умов з’єднання. Запит візуально поданий у графічному вигляді. Необхідні операції об’єднання даних генеруються алгоритмом перезапису. Кодування всіх необхідних метаданих, тобто глобальної схеми, описів джерел, зіставлень і запитів, у вигляді графів спрощує взаємодію між ними та дає змогу переписувати алгоритми для ефективної ідентифікації відповідних джерел. PolyWeb покладається на R2RML, а запропонований ним супровід виконує вихідний код на основі предикатів.

3.2 OBDA над нереляційними даними даними «розумних міст»

Існує чотири основні групи NoSQL Database Management Systems [6]:

– орієнтовані на документ;
– орієнтовані на стовпці;
– орієнтовані на графи;
– сховища ключ-значення.

Кожна з цих СУБД покладається на різні моделі даних і механізми запитів, що ускладнює завдання створення єдиної системи OBDA в цьому випадку. Через зростання популярності та відсутність уніфікованих мов запитів і моделей даних

для цих систем зростає інтерес до застосування OBDA в цьому контексті. Системи, засновані на VKG, обіцяють забезпечити бажане уніфіковане семантичне уявлення про набори даних. Ранні статті [52] включають різні мови, що були запропоновані для розширення мови відображення R2RML. RML є таким запропонованим розширенням для роботи з різнорідними джерелами даних та може бути застосована для потреб «розумних міст». Існують пропозиції розширити RML однозначно описаними функціями та операціями та незалежно від технології, яка використовується для реалізації. Такі приклади включають онтологію функцій (FnO) і FunUL [6]. Тоді як RML розширює R2RML до перетворень схеми, комбінація RML з FnO розширює R2RML щодо перетворень даних. RML широко поширений і реалізований в існуючих системах, і існує багато пропозицій, спрямованих на полегшення його використання. YAML [6] спрощує використання R2RML та і RML за допомогою зручного для читання текстового подання для правил відображення, виражених у YAML, широко використовуваній зручній для людини мові серіалізації даних. Редактор RML, RMLx, Map-On – тільки для R2RML, та MapVOWL, надають візуальну підтримку, щоб допомогти користувачам генерувати пов’язані дані з необроблених даних «розумних міст» шляхом визначення зіставлення RML. Автори проводять експеримент щодо зручності використання на трьох різних мовах ShExML, YARRRML і SPARQL-Generate [6]. RML було пропущено через занадто багатослівність. Їхній прототип перекладає файли JSON та XML у формат RDF за допомогою однієї з мов відображення та перевіряє зручність використання за допомогою анкети для користувачів, які вперше знайомляться з ними, які мають лише базові навички програмування та пов’язаних даних. Результати показують, що користувачі ShExML працюють краще, ніж користувачі YARRRML і SPARQL-Generate. З SDM-RDFizer, RocketRML [6], існують додаткові конкретні рішення для «розумних міст», які побудовані на RML. Однак ці підходи, а також деякі зі згаданих раніше, дотримуються підходу ETL, тобто перетворення всіх даних у великий RDF-файл на відміну від перезапису запитів на вимогу.

xR2RML – це мова для відображення різних типів баз даних у RDF, що дає змогу перекладати широкий спектр джерел даних за допомогою переписування запитів [53]. Оскільки ці мови не обов'язково сумісні, і багато з них підтримують лише певний механізм, наступне покоління систем OBDA «розумних міст» має розглянути, як звертатися до мов відображення перекладу, на додаток до методів перезапису запитів, які широко розглядалися в SOTA OBDA поки що (див рисунок 3.2).

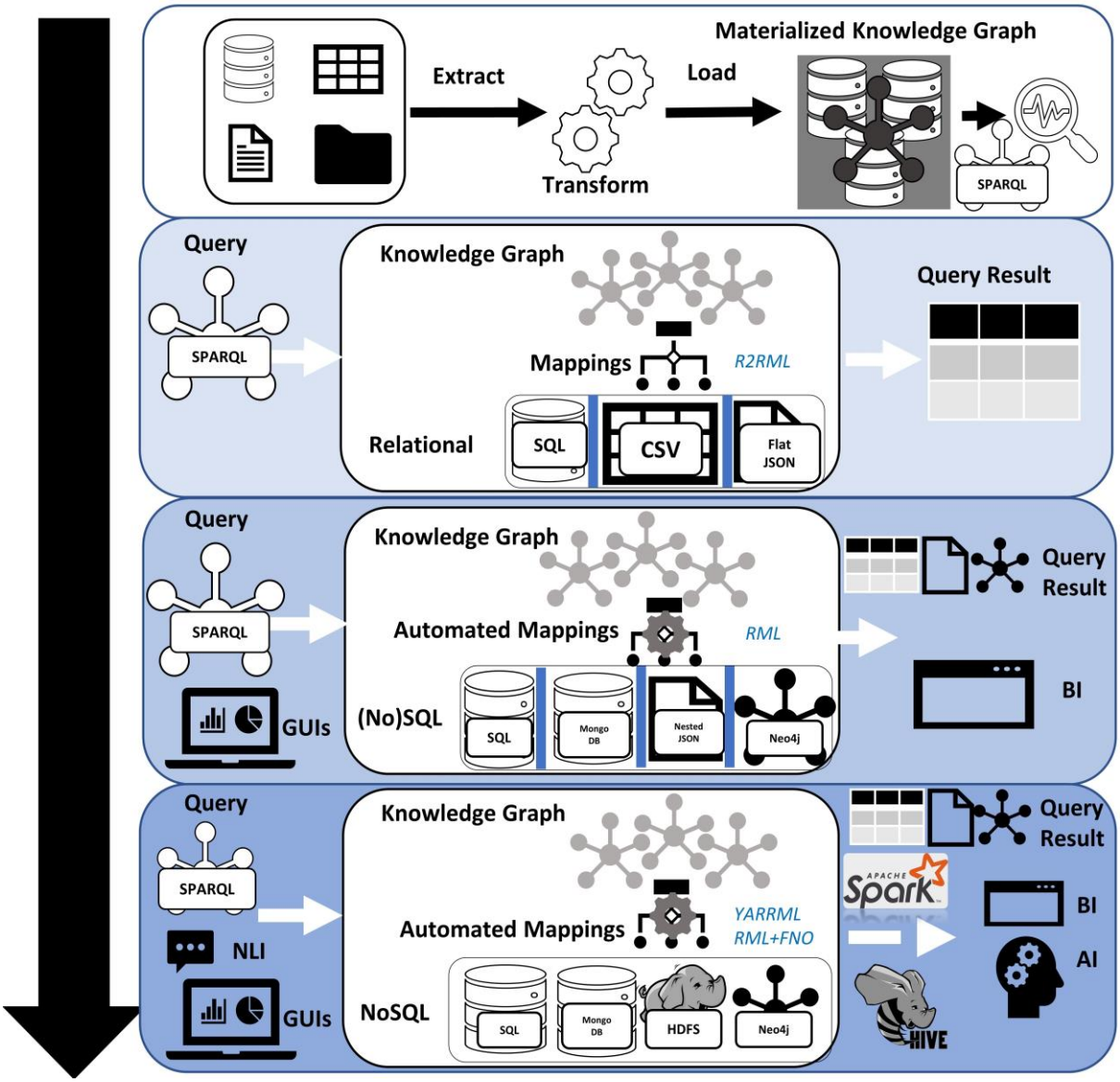

Рисунок 3.2 – Хронологія методів OBDA «розумних міст» [6]

Squerall витягує дані, передає їх у ефективний формат і надсилає запити масштабованим способом за допомогою Apache Spark. Навпаки, Ontop виник ще до публікації Chimera і тому не використовує жодної технології великих даних.

Однак у наданому порівнянні зазначено, що Ontop все ще конкурентоспроможний для більшості оцінюваних типів запитів, їх загалом 12, за винятком запитів, які передбачають агрегування, де Squerall ефективніший. Це може виникнути через те, що OnTop має більш складну оптимізацію запитів техніки, тоді як Squerall покладається виключно на свій механізм обробки. Крім того, Squerall не підтримує всі типи запитів «розумних міст».

Базуючись на наявній літературі, визначаємо історичний розвиток методів OBDA (див. рис. 3.2), який розділено на чотири різні періоди.

По-перше, є пропозиції, що дотримуються підходу ETL, тобто перетворення всіх джерел даних у величезний RDF-граф для запитів. за допомогою SPARQL. Хоча підхід може запропонувати уніфікований метод запиту для різнорідних джерел даних, питання інтеграції даних залишається без відповіді, оскільки дані лише перетворюються в загальний формат, але незрозуміло, як вони семантично пов’язані один з одним [6]. Підхід також має описані раніше недоліки щодо живості даних і накладних витрат, пов’язаних із початковим перетворенням у RDF.

Потім ці слабкі місця були усунені першими платформами, які виконували OBDA над реляційними базами даних за допомогою перезапису запитів, що залишає дані у відповідній СУБД, яка отримує їх за допомогою власних методів оптимізації запитів. Хоча ці рішення демонструють вражаючий прогрес, вони здебільшого обмежені підтримкою лише однієї реляційної СУБД одночасно [6]. Процес створення відображень також є здебільшого ручним, а отже виснажливим, схильним до помилок і адаптованим до єдиної статичної онтології.

Третій період усуває прогалину в доступі до баз даних NoSQL і більш гнучких моделей даних, таких як вкладені структури, графіки або REST API, а також розширює простір результатів запитів до моделей даних NoSQL. У цей період з’являється багато «одноцільових» рішень, які розширюють підтримку однієї СУБД NoSQL або напівструктурованих файлів, таких як JSON і XML, але не існує рішення, яке було б достатньо гнучким для запитів до довільних баз даних «розумних міст». У цей період спостерігаємо перші реальні програми, які

створили GUI для цих методів для ефективного виконання бізнес-аналітики, наприклад, [54]. Паралельно набирає обертів розробка алгоритмів для автоматизованої генерації відображень. Ранні пропозиції з акцентом на OBDA включають вже згаданий BootOX, доповнений MIRROR [6], обидва сприяють автоматичному створенню W3C R2RML з RDBMS «розумних міст». За ними йдуть публікації обширного кола дослідників [55], що подають різні алгоритми для покращення якості відображення, зменшення складності завдання та, отже, зусиль користувача для створення відображень R2RML «розумних міст».

Нарешті, найновіші підходи додають ефективні, масштабовані та розподілені механізми обробки, такі як Apache Spark або Hive і сховища даних «розумних міст», такі як HDFS. Chimera Suite підтримує всі рекомендації W3C щодо мови відображення R2RML для перекладу SPARQL у SQL і усуває важливий розрив між технологіями семантики та великих даних. Однак у той же час він обмежений Hive і реляційною моделлю даних. Крім того, традиційні відображення «розумних міст» на основі GAV мають той недолік, що їх потрібно переглянути, якщо зміни схеми відбуваються на рівні джерела. Наприклад, thedocker container22 потрібно перезапускати щоразу, коли зміни Ontop вносяться до визначень відображення або джерел даних «розумного міста». Цю проблему частково вирішує використання відображень GLAV, але вони ще не привернули широкої уваги, тому мають бути подані масштабовані пропозиції з відкритим кодом. З іншого боку, Squerall підтримує довільні бази даних NoSQL, водночас також використовує масштабовані механізми обробки, але не підтримує складні запити SPARQL. Крім того, спостерігаємо появу методів, керованих даними, зокрема масштабування можливостей використання аналітики машинного навчання (ML), яка вимагає керування великою різноманітністю різнорідних даних «розумного міста» за допомогою єдиного механізму, щоб заощадити час на розробці нових рішень ML та повторно використовувати вже розроблені. Для цього компанії зобов’язуються розробити довгострокові загальні семантичні рамки для управління промисловими даними та аналізу ML [56].

Тісніша співпраця між системами «людина-машина» та «машина-машина» революціонізувала поточну промислову ситуацію, що призвело до розвитку «Індустрії 4.0» – також відомої як «Промисловий Інтернет речей» (IIoT) [57]. Ключові чинники IIoT включають постійне додавання нових типів пристроїв із власними моделями даних «розумного міста». Цей вибух даних потребує високогнучких моделей даних, які дають можливість здійснювати моніторинг, обробку, оптимізацію та аналіз даних для отримання розуміння для кращого прийняття рішень «розумного міста». Однак ефективне використання даних вимагає інтеграції даних, що включає очищення, дедуплікацію та семантичну гомогенізацію. Продовжимо, подаючи два випадки використання, відібрані вручну, де управління семантичними даними «розумного міста» вийшло за межі академічної сфери та застосовувалося до реальних проблем Інтернету речей.

3.3 Виробництво семантичних озер даних «розумних міст»

Компанія Bosch визнала потенціал графів знань і серія публікацій висвітлює значні успіхи компанії в області управління семантичними даними та інтеграції. По-перше, OBDA з використанням SANSA Stack, а також OnTop було описано в області технології поверхневого монтажу (SMT), процесу монтажу електронних компонентів на друкованих платах [58]. Цей процес включає декілька підпроцесів, які виконуються спеціалізованими «розумними» пристроями, кожен з яких генерує значні обсяги даних. Ці дані містять семантичні конфлікти сумісності, такі як варіанти іменування об’єктів. Щоб ефективно досліджувати ці дані, ці конфлікти потрібно вирішити шляхом відображення їх на терміни онтології. Розробка онтології SMT стала результатом спільної роботи, яка тривала декілька тижнів, за участю лінійного інженера, двох лінійних менеджерів, експерта з процесів SMT, менеджера даних SMT, менеджера проекту, двох менеджерів великих даних і чотирьох експертів із семантики. Таким чином, прийняття підходу VKG, який зосереджується на створенні високоякісних відображень, є трудомістким процесом, який спочатку потребує поєднання трьох типів знань: знання предметної області, деталей схеми

бази даних і розуміння підходу VKG. Однак після досягнення інтегроване подання даних, яке забезпечує онтологія SMT, дає змогу експертам домену виконувати складні завдання аналізу продукту на раніше неінтегрованих необроблених даних «розумного міста».

Розробка графів знань «розумних міст» є серйозною справою для компанії, що вимагає значних початкових зусиль, але обіцяє довгострокові рішення для різних складних завдань. По-перше, це служить головній меті зберігання знань «розумних міст» і підтримки інженерів-матеріалів у пошуку інформації [6]. При цьому, основна увага поширюється на методи, керовані даними, зокрема на підвищення зручності використання аналітики машинного навчання (ML). Це передбачає керування неоднорідними даними з широкого спектру джерел за допомогою уніфікованого механізму, економію часу на розробку нових рішень ML і повторне використання раніше розроблених [6]. Їхня система містить декілька семантичних артефактів:

– основну онтологію, яка інкапсулює загальні знання про виробничий процес;

– онтології домену, що відображає домен-специфічні знання;

– відображення онтології даних у домен, які пов’язують атрибути необроблених даних із термінами онтології домену «розумних міст», щоб увімкнути OBDA.

Декілька семантичних модулів підтримують керування трьома фундаментальними будівельними блоками:

– модуль генерації KG;

– резонер чи анотатор відображення;

– модуль інтеграції даних.

Головною метою компанії є використання аналітики машинного навчання для досягнення високого рівня автоматизації виробництва. Усі ці компоненти служать передумовами, кульмінацією яких є платформа, відома як SemML – Semantic Machine Learning [59], яка вирішує три основні проблеми за допомогою семантично вдосконаленого ML:

1) комунікація, оскільки робочий процес передбачає співпрацю між експертами з різних областей, включаючи науковців з обробки даних, інженерів, експертів з процесів і менеджерів із різним досвідом, що робить спілкування трудомістким і схильним до помилок;

2) інтеграція даних;

3) можливість узагальнення моделей ML.

3.4 Використання семантичних озер даних «розумних міст» в розумному місті

Цифрова трансформація «розумних міст» охоплює не тільки промислові програми, але й інфраструктури для послуг загального інтересу, таких як транспортні мережі та пов’язані з ними варіанти транспорту, водо- та енергопостачання, а також утилізація відходів і стічних вод. Ця діяльність зазвичай асоціюється з терміном «розумне місто». У «розумному місті» інтелектуальна інформаційна та комунікаційна технологія використовується для покращення участі, якості життя та створення економічно, екологічно та соціально стійкої спільноти чи регіону.

Автори [21] використовують індикатори, щоб уможливити персоналізоване дослідження озера даних, демонструючи ефективність моделі в середовищі «розумних міст» завдяки різноманітності даних і ролей користувачів. Крім того, дослідники подають інструмент, який реалізує семантичний рівень над гетерогенною екосистемою джерел даних, служачи відправною точкою для програм із підтримкою семантики «розумних міст» або дослідження джерел даних. Їх конвеєр включає три етапи:

1) лексичне збагачення атрибутів;

2) стандартизація назв понять для анотації;

3) семантична анотація атрибутів даних;

4) створення семантичних відношень.

Подібним чином, було описано маркування семантичних даних Ketplace для легкого обміну даними в розумному місті, застосовуючи попередницю

PLASMA для застосунків «розумних міст». Ринок семантичних даних використовувався в проекті, який вимагав від команд розробки. Міські заявки на основі наборів даних, наданих міськими департаментами з трьох «розумних міст». Обидва підходи включають виробника та споживача даних. Виробник даних, експерт із домену, відповідає за завантаження наборів даних і збагачення їх семантичними метаданими з графів знань (див. рисунок 3.3).

Рисунок 3:3 – Діаграма Венна для ілюстрації поточного стану в області управління семантичними даними «розумних міст»

Використовуючи граф знань і пов’язані з ним семантичні моделі, споживачі даних можуть знаходити та розуміти дані шляхом пошуку та перегляду семантичних концепцій. VC-SLAM [15], що складається з понад ста

наборів даних з відповідними семантичними мітками та семантичною моделлю, було створено з використанням даних «розумних міст», отриманих із порталів відкритих даних. Нещодавнє дослідження [60] подає технології семантичного Інтернету речей, які можуть зіграти вирішальну роль у вирішенні важливих проблем «розумного міста», насамперед проблем сумісності.

Незважаючи на те, що розглянули різні інструменти, в яких KG відіграють вирішальну роль, комплексного рішення, яке забезпечує справді масштабований доступ до гетерогенних даних «розумних міст» на основі відображень на базі онтології, включаючи компонент семантичного моделювання для створення та підтримки відображень. Цю прогалину проілюстровано на рис 3.3. На діаграмі Венна пов’язано три сутності: OBDA, семантичне моделювання та озера даних.

Управління семантичними даними «розумних міст» може включати більше підходів, наприклад підходи ETL. Перетин між семантичним моделюванням і OBDA поданий, наприклад, ODIN. Поєднання OBDA для великих даних і озер даних «розумних міст» проілюстровано прикладом використання у виробництві [58]. Однак вона вимагала великої кількості ручної праці експертів. Водночас її структура недоступна ні з відкритим вихідним кодом, ні як комерційна система тому семантичне моделювання має бути включено в озері даних «розумного міста».

Однак використовується спеціальний формат даних Semantic Linked Tree (SLT), який є форматом даних на основі JSON, у якому семантичні мітки додаються до значень необроблених даних «розумного міста». Цю трансформацію можна інтерпретувати як семантичний підйом. Через свою досить ранню розробку система також не містить масштабованих механізмів зберігання та обробки даних, властивих більшості сучасних озер даних «розумного міста».

OBDA розвинулась з традиції, декілька разів повідомлялося про виробництво, а останні пропозиції, такі як SANSA та Chimera демонструють значний прогрес у напрямку доступу до даних на вимогу на основі семантики «розумного міста». Однак створення необхідних відображень є рутинною ручною роботою для користувача.

На даний час досліджувана у кваліфікаційній роботі область стикається з недостатнім рівнем якісних інтерфейсів користувача чи автоматичних процедур. Тому в перспективі вона стикається з декількома викликами:

– Початкові накладні витрати та зручність використання – створення значущих семантичних моделей для великої кількості різнорідних наборів даних «розумних міст» супроводжується величезними початковими накладними витратами на створення KG та відповідних зв'язків. По-перше, створення предметно-специфічних концептуалізацій вже займає багато часу та ресурсів, не дивлячись на те, що це завдання справді є окремою областю. Навіть якщо припустимо, що маємо під рукою різноманітний набір відповідних KG «розумного міста», створення відповідних зіставлень з джерелами даних все одно займає багато часу. Є деякі пропозиції щодо автоматичного створення семантичних моделей і відображень «розумних міст», однак більше уваги слід приділяти необхідному людському внеску.

– Оцінка – впродовж останнього періоду часу значна увага приділяється автоматизованому створенню семантичних міток і моделей «розумних міст», однак незрозуміло, наскільки точні методи насправді, за межами наборів тестів, використаних у конкретній публікації. Для семантичного маркування SemTab описано різні набори даних і тести, і зараз проводяться дослідження для більш складної інтерпретації таблиць. Для семантичних моделей VC-SLAM є першою пропозицією щодо стандартизації набору онтологій і наборів даних поряд із відповідними семантичними моделями. Поява щорічних викликів і публікація контрольних показників «розумних міст» буде корисною і може бути додатково спрощено за допомогою ініціативи, подібної до OAEI25 для узгодження онтології.

– Технічна сумісність – беручи до уваги різноманітність сучасного ландшафту даних «розумних міст», слід зазначити, що більшість методів семантичного маркування та моделювання базуються на табличних даних. Ця сфера ще далека від рівномірного охоплення всіх моделей даних NoSQL. Це також є фокусом сучасних досліджень OBDA «розумних міст». Хоча масштабований доступ до даних за допомогою онтологій можна вважати

значною мірою вирішеним для реляційних даних, увага зміщена більше в бік включення різних мов запитів NoSQL, а також підтримки об’єднаної обробки запитів до баз даних, файлових систем та інших різноманітних джерел даних «розумних міст» одночасно. Будь-яке таке рішення має бути сумісним зі стандартами W3C для семантичної мережі, щоб забезпечити максимальну сумісність між системами.

– Технічна абстракція – загальна якість семантичної моделі повинна бути остаточно вдосконалена людиною-оператором. Цей циклічний процес потребує керівництва користувачів через сильну технічну абстракцію, тобто технічні деталі повинні бути приховані від окремого користувача «розумного міста». Важко уявити, що навіть із дуже вдосконаленими методами штучного інтелекту, перевірка людиною та вдосконалення відображень стануть застарілими. Тому в цьому домені потрібні розширені користувацькі інтерфейси, які зосереджені на зручності використання також і для нетехнічних користувачів «розумних» послуг.

– Застосовність для великих даних – такі рішення, як Squerall і Chimera, є багатообіцяючими підходами до вирішення OBDA «розумних міст» також у сценаріях великих даних. Проте існує багато обмежень у виразності запитів або відображень. Крім того, вони прив’язані до певних версій платформ великих даних «розумних міст», що ускладнює їх розгортання. Зрілість таких рішень можна підвищити, якщо обширніша спільнота продовжить розробку цих прототипів, щоб усунути обмеження та зробити їх застосовними до широкого спектру систем NoSQL великих даних «розумних міст».

– Використання штучного інтелекту – після публікації ChatGPT багато людських завдань обговорювалися, чи можна їх вирішити за допомогою великих мовних моделей (LLM) у майбутньому. Це також актуально для всіх завдань, пов’язаних з інтеграцією даних, оскільки ChatGPT може описувати джерела даних, створювати онтології, знаходити зв’язки між різними наборами даних, щоб назвати лише декілька прикладів того, чого можна досягти за допомогою LLM. З огляду на останній розвиток методів штучного інтелекту, маємо очікувати, що ці системи зможуть вирішувати більш складні завдання інтеграції

даних «розумних міст» у майбутньому [6]. Було б цікаво подивитися, як такі універсальні системи штучного інтелекту можна налаштувати та оптимізувати для конкретних завдань інтеграції даних у масштабних системах озер даних «розумних міст». З іншого боку, нам потрібно вивчити процеси в озерах даних, які обговорюються в цьому опитуванні, щоб побачити, в яких точках і як можна інтегрувати систему штучного інтелекту на основі LLMs для покращення загального процесу.

У кваліфікаційній роботі надано огляд методів, заснованих на семантиці, для керування даними, доступу та інтеграції та пов’язали ці висновки з поточними пропозиціями семантичного озера даних. Нарешті, можемо сказати, що спільнота «розумних міст» стикається з рядом задач та розривом у сучасному ландшафті між існуючими платформами озер даних, OBDA та семантичними технологіями для моделювання контексту «розумних міст», в якому виникають гетерогенні набори даних. Неоднорідність, притаманна застосунками великих даних, все ще викликає занепокоєння в поточних платформах OBDA, хоча останнім часом було досягнуто значного прогресу.

У попередніх параграфах було проаналізовано ряд проблем, які можна вирішити в галузі досліджень озер даних «розумних міст». Тому можна бути впевненим, що технології великих даних і семантичного вебу можуть виграти від комплексної інтеграції в «розумних містах», і що в найближчі роки стануть доступні більш вдосконалені рішення для озер семантичних даних.

3.5 Висновок до третього розділу

В третьому розділі кваліфікаційної роботи описано OBDA над реляційними даними «розумних міст». Розглянуто OBDA над нереляційними даними даними «розумних міст». Досліджено виробництво семантичних озер даних «розумних міст». Описано використання семантичних озер даних «розумних міст» в розумному місті.

4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1 Медичні профілактичні заходи щодо збереження здоров’я та працездатності користувачів комп’ютерів та відеодисплейних терміналів

Кваліфікаційна робота освітньо-наукового рівня «Магістр» присв’ячена дослідженню семантичного управління озерами сховищ даних «розумних міст». Більшість аналітичного опрацювання великих обсягів даних виконується за допомогою персональних комп'ютерів. Тому з погляду охорони праці важливо розглядати медичні профілактичні заходи для збереження здоров'я та працездатності користувачів ПК та відеодисплейних терміналів. Заходи в області охорони праці користувачів ПК варто розглядати з трьох основних поглядів [61]:

– Соціальний: це охоплює оптимізацію умов життя, роботи, відпочинку, харчування, побуту, а також розвиток культури і транспорту.

– Психологічний: розв'язання цих проблем пов'язане з формуванням раціональних колективів, де відсутня психологічна несумісність, що сприяє зниженню нервово-психічного напруження та підвищенню працездатності та ефективності праці.

Особливу увагу серед користувачів відеодисплейних терміналів приділяється психоемоційному стресу, який виявляється у кожного в різному ступені.

Існує набір профілактичних заходів для користувачів ПК, який включає як елементи первинної профілактики здоров'я (професійний відбір), так і вторинної, спрямованої на зменшення ймовірності виникнення перевтоми та перенапруження. Ці комплексні заходи націлені на відновлення функціонального стану органів зору та опорно-рухового апарату.

Медичні заходи для підтримки здоров’я та підвищення працездатності користувачів комп’ютерів включають:

– Медичні огляди.

– Раціональне та профілактичне харчування.

– Спеціальні вправи.

– Самомасаж.

– Психофізіологічне розвантаження.

Режим роботи і відпочинку розробляється з урахуванням особливостей праці робітників. Перш за все, надають оптимальні режими для тих, хто працює з підвищеними фізичними та нервово-емоційними навантаженнями, особливо в умовах монотонності та впливу небезпечних і шкідливих виробничих факторів.

Лікувально-профілактичне обслуговування охоплює профілактичні та регулярні медичні огляди працівників, забезпечення лікувально-профілактичного харчування та проведення заходів для запобігання захворюванням серед працівників [62]. Санітарно-побутове обслуговування включає забезпечення санітарно-побутовими приміщеннями та пристроями для працівників і їх експлуатацію відповідно до чинних норм і правил. Професійний відбір за конкретними спеціальностями передбачає визначення професійної (фізіологічної та психофізіологічної) придатності працівників для безпечного виконання робіт.

В даному розділі було розглянуто медичні профілактичні заходи для збереження здоров'я та підтримки працездатності користувачів комп'ютерів та відеодисплейних терміналів, оскільки кваліфікаційна робота присвячена вивченню засобів та методів машинного навчання для аналізу великих обсягів даних.

4.2 Особливості заходів електробезпеки на підприємствах

Дана кваліфікаційна робота присв’ячена дослідженню семантичного управління озерами сховищ даних "розумних міст". Оскільки при запровадженні інновацій в галузі «розумних міст» активно використовуються різноманітні електронні та електричні пристрої та обладнання, то доцільно розглянути особливості заходів електробезпеки на підприємствах. В умовах сучасного виробництва, яке нерозривно пов’язане з використанням електроенергії, особливого значення набуває питання електробезпеки.

Електробезпека визначається як комплекс організаційних, технічних заходів і засобів, що гарантують захист людей від шкідливого та небезпечного впливу електричного струму, електричної дуги, електромагнітного поля і статичної електрики [63]. Забезпечення електробезпеки на підприємстві визначається додержанням вимог, які викладені у законодавчих актах:

– Правила безпечної експлуатації електроустановок споживачів (далі – ПБЕЕС), затверджені наказом Держнаглядохоронпраці від 09.01.1998 № 4, вимоги яких поширюються на працівників, що обслуговують діючі електроустановки споживачів напругою до 220 кВ включно і є обов'язковими для всіх споживачів та виробників електроенергії, незалежно від їх відомчої належності і форм власності на засоби виробництва.

– Правила безпечної експлуатації електроустановок, дія яких поширюються на працівників, що виконують роботи в електроустановках Міністерства енергетики України (наказ Держнаглядохоронпраці України від 06.10.1997 № 257).

– Правила технічної експлуатації електроустановок споживачів (ПТЕЕС), затверджені наказом Мінпаливенерго України від 25.07.2006 № 258 (у редакції наказу від 13.02.2012 № 91), якими унормовано організаційні й технічні вимоги щодо експлуатації електроустановок споживачів.

– Правила експлуатації електрозахисних засобів, затверджені наказом Міністерства праці та соціальної політики України від 05.06.2001 № 253, в яких наведено перелік засобів захисту, вимоги до них, обсяги і норми випробувань, порядок застосування, зберігання їх, а також норми комплектування засобами захисту електроустановок і виробничих бригад.

– Правила улаштування електроустановок (ПУЕ), які визначають будову, принципи улаштування, особливі вимоги до окремих систем, їх елементів, вузлів і комунікацій електроустановок. Наказ Мініністарства енергетики та вугільної промисловості України від 24.07.2017 № 476.

– ДСТУ 2843-94 «Електротехніка. Основні поняття. Терміни та визначення», який установлює терміни та визначення основних понять в галузі електротехніки.

– Правила пожежної безпеки в Україні, затверджені наказом МВС від 30.12.2014 № 1417.

Побутові електроприлади в умовах виробництва експлуатують відповідно до експлуатаційної документації підприємств-виробників і ПТЕЕС, п. 1.5 розд. I ПТЕЕС. Відповідальність за організацію безпечної експлуатації електроустановок ПБЕЕС покладають на роботодавця, який повинен [64]:

– призначити відповідального за справний стан і безпечну експлуатацію електроустановок;

– створити і укомплектувати електротехнічну службу з числа осіб, які досягли 18-річного віку, мають відповідну освіту та пройшли медичний огляд і не мають протипоказань;

– розробити і затвердити Положення про енергетичну службу підприємства, посадові інструкції працівників та інструкції з безпечного виконання робіт;

– забезпечити навчання і перевірку знань працівників, своєчасний огляд електроустановок, проведення профілактичних, протиаварійних та приймально-здавальних випробувань;

– встановити такий порядок, щоб працівники, на яких покладено обов’язки з обслуговування електроустановок, вели ретельні спостереження за дорученим їм обладнанням і мережами.

Для прямого виконання завдань з організації функціонування електроустановок призначається особа, яка має відповідальність за управління електроенергетичною системою.

4.3 Висновок до четвертого розділу

В четвертому розділі кваліфікаційної роботи описано медичні профілактичні заходи щодо збереження здоров’я та працездатності користувачів комп’ютерів та відеодисплейних терміналів. Розглянуто особливості заходів електробезпеки на підприємствах.

ВИСНОВКИ

У цьому опитуванні ми розглядаємо останні підходи з особливим акцентом на застосуванні в системах озер даних і масштабованості до великих даних. Ми класифікуємо підходи на базове управління семантичними даними, підходи до семантичного моделювання для збагачення метаданих в озерах даних і методи доступу до даних на основі онтології. У кожній категорії ми розглядаємо основні методи та їх передумови, а також порівнюємо останні дослідження. Нарешті, ми вказуємо на виклики для майбутньої роботи в цій галузі досліджень, яка потребує більш тісної інтеграції великих даних і технологій семантичного вебу.

В першому розділі кваліфікаційної роботи освітнього рівня «Магістр»:

– Описано міські джерела семантичних даних.

– Проаналізовано стан та досліджень в царині озер даних «розумних міст».

– Наведено ключові терміни та визначення управління семантичними даними в озерах даних.

– Розроблена таксономія управління семантичними даними.

– Розглянуті моделі метаданих для озер даних, які враховують семантику.

– Описані існуючі системи семантичних озер даних «розумних міст».

– Виконано порівняння та обговорення семантичних озер даних «розумних міст».

В другому розділі кваліфікаційної роботи:

– Досліджено системи семантичного моделювання «розумних міст».

– Розглянуто автоматизоване присвоєння семантичних міток «розумного міста».

– Висвітлена автоматизована генерація семантичної моделі «розумного міста».

– Виконано порівняння та обговорення семантичних моделей озер даних «розумних міст».

В третьому розділі кваліфікаційної роботи:

– Описано OBDA над реляційними даними «розумних міст».

– Розглянуто OBDA над нереляційними даними даними «розумних міст».

– Досліджено виробництво семантичних озер даних «розумних міст».

– Описано використання семантичних озер даних «розумних міст» в розумному місті.

У розділі «Охорона праці та безпека в надзвичайних ситуаціях» проаналізовано описано медичні профілактичні заходи щодо збереження здоров’я та працездатності користувачів комп’ютерів та відеодисплейних терміналів. Розглянуто особливості заходів електробезпеки на підприємствах.

ПЕРЕЛІК ДЖЕРЕЛ

1 E. Bionda, C. Tornelli, M. Brambilla, M. Balduini, G. Mauri, D. D. Giustina, F. Garrone, E. D. Valle, The smart grid semantic platform: Synergy between iec common information model (cim) and big data, in: Proc. IEEE EEEIC/I&CPS Europe), 2019, pp. 1–6. doi: 10.1109/EEEIC.2019.8783632.

2 A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, A. N. Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, J. F. Sequeda, S. Staab, A. Zimmermann, Knowledge graphs, ACM Comput. Surv. 54 (4) (2022) 71:1–71:37. doi:10.1145/3447772. URL https://doi.org/10.1145/3447772.

3 G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati, M. Zakharyaschev, Ontology-based data access: A survey, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden, ijcai.org, 2018, pp. 5511–5519.

4 Duda, O., Kunanets, N., Martsenko, S., Matsiuk, O., Pasichnyk, V., Building secure Urban information systems based on IoT technologies. CEUR Workshop Proceedings 2623, pp. 317-328. 2020.

5 A. Paulus, A. Burgdorf, A. Pomp, T. Meisen, Recent advances and future challenges of semantic modeling, in: Proc. 15th IEEE ICSC, IEEE, 2021, pp. 70–75. doi:10.1109/ICSC50631.2021.00016.

6 Hoseini, Sayed, Johannes Theissen-Lipp, and Christoph Quix. "Semantic Data Management in Data Lakes." arXiv preprint arXiv:2310.15373 (2023).

7 A. Pomp, J. Lipp, T. Meisen, You are missing a concept! enhancing ontology-baseddataaccesswithevolvingontologies,in:Proc.ICSC, IEEE, 2019, pp. 98–105. doi:10.1109/ICOSC.2019.8665620.

8 Duda, O., et al, Selection of Effective Methods of Big Data Analytical Processing in Information Systems of Smart Cities. CEUR Workshop Proceedings 2631, pp. 68-78. 2020.

9 P. N. Sawadogo, J. Darmont, On data lake architectures and metadata management, J. Intell. Inf. Syst. 56 (1) (2021) 97–120. doi:10.1007/ s10844-020-00608-7.

10 A. Pomp, A. Paulus, A. Kirmse, V. Kraus, T. Meisen, Applying semantics to reduce the time to analytics within complex heterogeneous infrastructures, Technologies 6 (3) (2018) 86.

11 G. Xiao, L. Ding, B. Cogrel, D. Calvanese, Virtual knowledge graphs: An overview of systems and use cases, Data Intell. 1 (3) (2019) 201–223.

12 R. Hai, C. Koutras, C. Quix, M. Jarke, Data lakes: A survey of functions and systems, IEEE Transactions on Knowledge and Data Engineering (2023) 1–20doi:10.1109/TKDE.2023.3270101.

13 É. Scholly, P. N. Sawadogo, P. Liu, J. Espinosa-Oviedo, C. Favre, S. Loudcher, J. Darmont, C. Noûs, Coining goldmedal: A new contribution to data lake generic metadata modeling, in: Proc. DOLAP, Vol. 2840 of CEUR WS, 2021, pp. 31–40. URL http://ceur-ws.org/Vol-2840/paper5.pdf.

14 Y. Suhara, J. Li, Y. Li, D. Zhang, Ç. Demiralp, C. Chen, W. Tan, Annotating columns with pre-trained language models, in: Proc. ACM SIGMOD, 2022, pp. 1493–1503. URL https://doi.org/10.1145/3514221.3517906.

15 A. Burgdorf, A. Paulus, A. Pomp, T. Meisen, VC-SLAM - A handcrafted data corpus for the construction of semantic models, Data 7 (2) (2022) 17. doi:10.3390/data7020017.

16 J. Liu, Y. Chabot, R. Troncy, V.-P. Huynh, T. Labbé, P. Monnin, From tabular data to knowledge graphs: A survey of semantic table interpretation tasks and methods, Journal of Web Semantics (2022) 100761.

17 A. Paulus, A. Burgdorf, L. Puleikis, T. Langer, A. Pomp, T. Meisen, PLASMA: platform for auxiliary semantic modeling approaches, in: Proc. ICEIS, SCITEPRESS, 2021, pp. 403–412. doi:10.5220/0010499604030412.

18 M. N. Mami, S. Scerri, S. Auer, M. Vidal, Towards semantification of big data technology, in: Proc. DaWaK 2016, Vol. 9829 of LNCS, Springer, 2016, pp. 376–390. doi:10.1007/978-3-319-43946-4\. URL https://doi.org/10.1007/978-3-319-43946-4_25.

19 C. Quix, R. Hai, I. Vatov, Metadata extraction and management in data lakes with GEMMS, Complex Syst. Informatics Model. Q. 9 (2016) 67–83. doi:10.7250/csimq.2016-9.04.

20 C. Diamantini, P. L. Giudice, L. Musarella, D. Potena, E. Storti, D. Ursino, A new metadata model to uniformly handle heterogeneous data lake sources, in: Proc. ADBIS Short Papers & Workshops, Vol. 909 of CCIS, Springer, 2018, pp. 165–177. doi:10.1007/978-3-030-00063-9_{1}{7}. URL https://doi.org/10.1007/978-3-030-00063-9_17.

21 A. Bagozi, D. Bianchini, V. D. Antonellis, M. Garda, M. Melchiori, Personalised exploration graphs on semantic data lakes, in: Proc. OTM Conf., Vol. 11877 of LNCS, Springer, 2019, pp. 22–39. doi: 10.1007/978-3-030-33246-4_2. URL https://doi.org/10.1007/978-3-030-33246-4_2.

22 E. G. Kalayci, I. Grangel-González, F. Lösch, G. Xiao, A. ul Mehdi, E. Kharlamov, D. Calvanese, Semantic integration of bosch manufacturing data using virtual knowledge graphs, in: Proc. ISWC, Vol. 12507 of LNCS, Springer, 2020, pp. 464–481.

23 P. Liu, S. Loudcher, J. Darmont, E. Perrin, J. Girard, M. Rousset, Metadata model for an archeological data lake, in: Proc. Digital Humanities Conf., 2020.URL https://dh2020.adho.org/wp-content/uploads/2020/07/206_Metadatamodelfora narcheologicaldatalake.html.

24 C. Stach, J. Bräcker, R. Eichler, C. Giebler, B. Mitschang, Demanddriven data provisioning in data lakes: BARENTS - A tailorable data preparation zone, in: Proc. iiWAS, ACM, 2021, pp. 187–198. doi:10.1145/3487664.3487784.

25 Bodnarchuk I., Duda O., Kharchenko A., Kunanets N., Matsiuk O., Pasichnyk V. Choice method of analytical information-technology platform for projects associated to the smart city class. ICTERI 2020 ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer Proceedings of the 14th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume I: Main Conference p.317-330.

26 G. A. Braun, L. A. Cecchi, P. R. Fillottrani, Towards All-In-One OBDA systems, in: Proc. Styrian Autumn of Ontology, Vol. 2518 of CEUR WS, 2019. URL http://ceur-ws.org/Vol-2518/paper-DAOSI1.pdf.

27 S. Gupta, P. A. Szekely, C. A. Knoblock, A. Goel, M. Taheriyan, M. Muslea, Karma: A system for mapping structured sources into the semantic web, in: ESWC Satellite Events - Revised Selected Papers, Vol. 7540 of LNCS, Springer, 2012, pp. 430–434. doi: 10.1007/978-3-662-46641-4_40. URL https://doi.org/10.1007/978-3-662-46641-4_40.

28 A. Pomp, J. Lipp, T. Meisen, Enabling the continuous evolution of ontologies for ontology-based data management, Intl. Journal of Robotic Computing (2019).

29 C. Pinkel, C. Binnig, E. Jiménez-Ruiz, E. Kharlamov, W. May, A. Nikolov, A. S. Bastinos, M. G. Skjæveland, A. Solimando, M. Taheriyan, C. Heupel, I. Horrocks, RODI: benchmarking relational-to-ontology mapping generation quality, Semantic Web 9 (1) (2018) 25–52. doi:10.3233/SW-170268.

30 S. K. Ramnandan, A. Mittal, C. A. Knoblock, P. A. Szekely, Assigning semantic labels to data sources, in: Proc. ESWC, Vol. 9088 of LNCS, Springer, 2015, pp. 403–417. doi:10.1007/978-3-319-18818-8_{2}{5}. URL https://doi.org/10.1007/978-3-319-18818-8_25.

31 M. Pham, S. Alse, C. A. Knoblock, P. A. Szekely, Semantic labeling: A domain-independent approach, in: Proc. ISWC, Vol. 9981 of LNCS, 2016, pp. 446–462.

32 Duda, O., Pasichnyk, V., Kunanets, N., Antonii, R., Matsiuk, O. Multidimensional Representation of COVID-19 Data Using OLAP Information Technology. International Scientific and Technical Conference on Computer Sciences and Information Technologies, 2020, 2, pp. 277–280, 9321889.

33 B. Vu, C. A. Knoblock, J. Pujara, Learning semantic models of data sources using probabilistic graphical models, in: Proc. WWW, 2019, pp. 1944–1953. doi:10.1145/3308558.3313711.

34 R. Avogadro, M. Cremaschi, Mantistable V: A novel and efficient approach to semantic table interpretation, in: Proc. SemTab, Vol. 3103 of CEUR WS, 2021, pp. 79–91. URL http://ceur-ws.org/Vol-3103/paper7.pdf.

35 G. Futia, A. Vetrò, J. C. D. Martin, Semi: A semantic modeling machine to build knowledge graphs with graph neural networks, SoftwareX 12 (2020) 100516. doi:10.1016/j.softx.2020.100516.

36 M. Taheriyan, C. A. Knoblock, P. A. Szekely, J. L. Ambite, Leveraging linked data to discover semantic relations within data sources, in: Proc. ISWC, Vol. 9981 of LNCS, 2016, pp. 549–565. doi:10.1007/978-3-319-46523-4_{3}{3}. URL https://doi.org/10.1007/978-3-319-46523-4_33.

37 A. Burgdorf, A. Paulus, A. Pomp, T. Meisen, Docsemmap: Leveraging textual data documentations for mapping structured data sets into knowledge graphs, in: Proc. IEEE ICSC, 2022, pp. 209–216. doi:10.1109/ICSC52841.2022.00042.

38 M. Ramirez, A. Bogatu, N. W. Paton, A. Freitas, Natural language inference over tables: Enabling explainable data exploration on data lakes, in: Proc. ESWC, Vol. 12731 of LNCS, Springer, 2021, pp. 304–320. doi:10.1007/978-3-030-77385-4_{1}{8}. URL https://doi.org/10.1007/978-3-030-77385-4_18.

39 R. C. Fernandez, E. Mansour, A. A. Qahtan, A. K. Elmagarmid, I. F. Ilyas, S. Madden, M. Ouzzani, M. Stonebraker, N. Tang, Seeping semantics: Linking datasets using word embeddings for data discovery, in: Proc. IEEE ICDE, 2018, pp. 989–1000. doi:10.1109/ICDE.2018.00093.

40 D. Haller, R. Lenz, Discovery of ontologies from implicit user knowledge, in: Proc. Conf. on "Lernen, Wissen, Daten, Analysen", Vol. 2738 of CEUR WS, 2020, pp. 241–245. URL http://ceur-ws.org/Vol-2738/LWDA2020_paper_4.pdf.

41 N. Abdelmageed, S. Schindler, Jentab: A toolkit for semantic table annotations, in: Proc. Workshop on Knowledge Graph Construction, Vol. 2873 of CEUR WS, 2021. URL http://ceur-ws.org/Vol-2873/paper5.pdf.

42 Kunanets N. et al. (2021) Designing the Repository of Documentary Cultural Heritage. In: Shakhovska N., Medykovskyy M.O. (eds) Advances in Intelligent Systems and Computing V. CSIT 2020. Advances in Intelligent Systems and Computing, vol 1293, pp 1034-1044. Springer, Cham. ISBN978-3-030-63270-0.

43 N. Fathy, W. Gad, N. Badr, A unified access to heterogeneous big data through ontology-based semantic integration, in: Proc. ICICIS, IEEE, 2019, pp. 387–392.

44 M. N. Mami, D. Graux, H. Thakkar, S. Scerri, S. Auer, J. Lehmann, The query translation landscape: a survey, CoRR abs/1910.03118 (2019). arXiv:1910.03118.URL http://arxiv.org/abs/1910.03118.

45 T. Bagosi, D. Calvanese, J. Hardi, S. Komla-Ebri, D. Lanti, M. Rezk, M. Rodriguez-Muro, M. Slusnys, G. Xiao, The ontop framework for ontology based data access, in: Proc. CSWS, Revised Selected Papers, Vol. 480 of CCIS, Springer, 2014, pp. 67–77. doi:10.1007/978-3-662-45495-4_6. URL https://doi.org/10.1007/978-3-662-45495-4_6.

46 D. Bilidas, M. Koubarakis, Efficient duplicate elimination in SPARQL to SQL translation, in: Proc. Description Logics Workshop, Vol. 2211 of CEUR WS, 2018.

47 M. Belcao, E. Falzone, E. Bionda, E. D. Valle, Chimera: A bridge between big data analytics and semantic technologies, in: Proc. ISWC, Vol. 12922 of LNCS, Springer, 2021, pp. 463–479. doi: 10.1007/978-3-030-88361-4_{2}{7}. URL https://doi.org/10.1007/978-3-030-88361-4_27.

48 K. M. Endris, P. D. Rohde, M. Vidal, S. Auer, Ontario: Federated query processing against a semantic data lake, in: Proc. DEXA, Vol. 11706 of LNCS, Springer, 2019, pp. 379–395. doi:10.1007/978-3-030-27615-7_{2}{9}. URL https://doi.org/10.1007/978-3-030-27615-7_29.

49 M. Buron, F. Goasdoué, I. Manolescu, M. Mugnier, Obi-wan: Ontology-based RDF integration of heterogeneous data, Proc. VLDB Endow. 13 (12) (2020) 2933–2936. doi:10.14778/3415478.3415512 .URL http://www.vldb.org/pvldb/vol13/p2933-buron.pdf.

50 M. Lenzerini, Data integration: A theoretical perspective, in: Proc. ACM PODS, 2002, pp. 233–246. doi:10.1145/543613.543644. URL https://doi.org/10.1145/543613.543644.

51 S. Nadal, A. Abelló, O. Romero, S. Vansummeren, P. Vassiliadis, Graph-driven federated data management, IEEE Trans. Knowl. Data Eng. 35 (1) (2023) 509–520. doi:10.1109/TKDE.2021.3077044.

52 O. Curé, R. Hecht, C. L. Duc, M. Lamolle, Data integration over nosql stores using access path based mappings, in: Proc. DEXA, Vol. 6860 of LNCS, Springer, 2011, pp. 481–495. doi:10.1007/978-3-642-23088-2_{3}{6}. URL https://doi.org/10.1007/978-3-642-23088-2_36.

53 F. Michel, C. Faron-Zucker, J. Montagnat, Bridging the semantic web and nosql worlds: Generic SPARQL query translation and application to mongodb, Trans. Large Scale Data Knowl. Centered Syst. 40 (2019) 125–165.

54 E. Kharlamov, T. Mailis, G. Mehdi, C. Neuenstadt, Ö. L. Özçep, M. Roshchin, N. Solomakhina, A. Soylu, C. Svingos, S. Brandt, M. Giese, Y. E. Ioannidis, S. Lamparter, R. Möller, Y. Kotidis, A. Waaler, Semantic access to streaming and static data at siemens, J. Web Semant. 44 (2017) 54–74. doi:10.1016/j.websem.2017.02.001.

55 S. Mathur, D. O'Sullivan, R. Brennan, Milan: Automatic generation of R2RML mappings, in: Proc. AIAI Irish Conf. on AI & Cognitive Science, Vol. 2259 of CEUR WS, 2018, pp. 90–101. URL http://ceur-ws.org/Vol-2259/aics_10.pdf.

56 B. Zhou, Y. Svetashova, A. Gusmao, A. Soylu, G. Cheng, R. Mikut, A. Waaler, E. Kharlamov, Semml: Facilitating development of ML models for condition monitoring with semantics, J. Web Semant. 71 (2021) 100664. doi:10.1016/j.websem.2021.100664.

57 A. Ustundag, E. Cevikcan, Industry 4.0: Managing the Digital Transformation, Springer, 2018.

58 M. N. Mami, I. Grangel-González, D. Graux, E. Elezi, F. Lösch, Semantic data integration for the SMT manufacturing process using SANSA stack, in: ESWC Satellite Events, Revised Selected Papers, Vol. 12124 of LNCS, Springer, 2020, pp. 307–311. doi:10.1007/978-3-030-62327-2_{4}{7}. URL https://doi.org/10.1007/978-3-030-62327-2_47.

59 B. Zhou, Y. Svetashova, T. Pychynski, I. Baimuratov, A. Soylu, E. Kharlamov, Semfe: Facilitating ML pipeline development with semantics, in: Proc. ACM CIKM, 2020, pp. 3489–3492. doi:10.1145/3340531.3417436.

60 S. Nahhas, Potentials of semantic internet of things in smart cities: an overview and roadmap, in: 2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC), 2023, pp. 1–6. doi:10.1109/ICAISC56366.2023.10085121.

61 Основні правила дотримання охорони праці при роботі на персональних ЕОМ. URL: https://www.victorija.ua/dovidnik/osnovni-pravyla-dotrymannya-ohorony-pratsi-pry-roboti-na-personalnyh-eom.html.

62 КУРС ЛЕКЦІЙ. ОХОРОНА ПРАЦІ В ГАЛУЗІ. URL: https://www.uzhnu.edu.ua/uk/infocentre/get/36621.

63 Електробезпека: охорона праці. URL: https://www.sop.com.ua/article/745-elektrobezpeka.

64 Лекція. Заходи електробезпеки на підприємствах галузі. URL: http://opcb.kpi.ua/wp-content/uploads/2014/09/%D0%9B%D0%B5%D0%BA%D1%86%D1%96%D1%8F-8.pdf.

ДОДАТКИ

Додаток А

Тези конференції

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ПУЛЮЯ

М А Т Е Р І А Л И

XI НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ

«ІНФОРМАЦІЙНІ МОДЕЛІ, СИСТЕМИ ТА ТЕХНОЛОГІЇ»

13-14 грудня 2023 року

ТЕРНОПІЛЬ
2023

УДК 004.9

Лебідко Д.М., Онуферко В.А., Перетятко Т.П.
Тернопільський національний технічний університет імені Івана Пулюя, Україна

ВЕЛИКІ ЗА ОБСЯГОМ ДАНІ, РЕЛЯЦІЙНІ ТА НЕРЕЛЯЦІЙНІ МОДЕЛІ

Lebidko D.M., Onuferko V.A., Peretiatko T.P.

BIG DATA, RELATIONAL AND NON-RELATIONAL MODELS

В кінці минулого століття наука та технології вступили в нову еру – еру великих за обсягом даних. Науковий прогрес призвів до різкого збільшення кількості даних, що продукуються щодня. Наукові дані, зібрані на даний час, здебільшого залежать від ряду різних змінних. Тому їх можна представити у вигляді багатовимірних масивів або гіперкубів даних [1]. Впорядкування даних усередині гіперкубів є важливою задачею, і для цього було розроблено багатий спектр різноманітних інструментів.

Більшість сучасних програмних архітектур забезпечують підтримку обробки структур даних на базі гіперкубів [2]. Однак у кожній з цих програм гіперкуби даних можуть отримувати доступ до даних лише «ортогонально» до своїх «осей». Це означає, що можна вибирати лише певні значення або діапазони вздовж заданих розмірів. Це обмежує можливості аналізу даних, оскільки часто потрібно отримувати доступ до даних у неортогональних напрямках. Наприклад, може бути потрібно порівняти дані з різних осей або визначити, як зміна однієї змінної впливає на інші.

Поширення великих за обсягом даних створило потребу в масштабованих системах, які можуть ефективно їх обробляти. Реляційні бази даних, які базуються на SQL, є стандартним підходом до зберігання та керування даними. Вони добре працюють для структурованих даних, зокрема, таблиць, але мають обмеження щодо масштабованості. Бази даних NoSQL, з іншого боку, не використовують реляційну модель даних. Вони зазвичай краще підходять для керування великими обсягами даних, оскільки можуть масштабуватися по горизонталі, тобто формувати додаткові вузли для обробки даних.

Бази даних NoSQL – це нереляційні бази даних, які використовують різні моделі зберігання даних, що краще підходять для сучасних типів даних. На даний час існує чотири основні типи баз даних NoSQL: Бази даних типу «ключ-значення» зберігають дані у вигляді пари ключ-значення. Ключ ідентифікує запис, а значення містить дані про запис. Ці бази даних добре підходять для зберігання великих обсягів даних, які не мають чіткої структури. Бази даних документів зберігають дані у вигляді документів, які є JSON-подібними об'єктами. Ці бази даних добре підходять для зберігання неструктурованих даних, таких як текст і код. Бази даних стовпців зберігають дані у вигляді стовпців. Ці бази даних добре підходять для зберігання даних, які часто запитуються за певними стовпцями. Графові бази даних зберігають дані у вигляді графів. Граф складається з вузлів та ребер, які подають об'єкти та відношення між ними. Ці бази даних добре підходять для зберігання даних, що мають багато зв'язків між сутностями.

Захист особистої інформації користувачів є важливою задачею при передачі хмарних сховищ. Це пов'язано з тим, що хмарне сховище містить чутливі дані, зокрема, імена, адреси електронної пошти, номери кредитних карток та інші особисті дані.

Література

1. J. Gray, S. Chaudhuri, A. Bosworth, A. Layman, D. Reichart, M. Venkatrao, F. Pellow, and H. Pirahesh, "Data cube: A relational aggregation operator generalizing group-by, cross-tab, and subtotals," Data mining and knowledge discovery, vol. 1, pp. 29–53, 1997.

2. Xtensor Stack, "Xtensor Documentation," In *Xtensor* (Version 0.24.6). Retrieved from Read the Docs: [https://xtensor. readthedocs.io/en/latest/], 2023.

УДК 004.9

Лебідко Д.М., Онуферко В.А., Перетятко Т.П.
Тернопільський національний технічний університет імені Івана Пулюя, Україна

ХМАРНІ ПЛАТФОРМИ, ОБЧИСЛЕННЯ ТА ІНТЕРНЕТ РЕЧЕЙ

Lebidko D.M., Onuferko V.A., Peretiatko T.P.

CLOUD PLATFORMS, COMPUTING AND THE INTERNET OF THINGS

Керування хмарною інфраструктурою та платформами є важливим та складним завданням, яке вимагає ретельного планування та поетапного виконання, а розуміння внутрішніх складнощів, які потрібно враховувати, допоможе установам та організаціям розробити ефективні плани управління. Існує багато внутрішніх складнощів, які потрібно враховувати.

Хмарні платформи повинні керувати хмарним програмним забезпеченням, яке використовується в установі чи організації. Це включає в себе такі завдання щодо встановлення, конфігурування та оновлення програмного забезпечення. Хмарна платформа повинна керувати транзакціями, передбаченими з постачальниками хмарних послуг. Це включає в себе завдання моніторингу відповідності транзакцій та вирішення спорів. Хмарна платформа повинна керувати хмарними програмними службами, які використовуються в організації. Це включає в себе завдання планування, розгортання та управління послугами. Хмарна платформа повинна сегментувати хмарну інфраструктуру на основі планових транзакцій. Це допомагає забезпечити, щоб дані, які підпадають під різні вимоги, були ізольовані одні від одних.

Хмарні обчислення – це модель обробки даних, при якій ресурси, сервери, сховища та програми, доступні через Інтернет. Це надає користувачам можливість отримувати доступ до цих ресурсів за потреби, не маючи необхідності інвестувати в власну інфраструктуру. Хмарні обчислення мають ряд переваг перед традиційними моделями обробки даних. Хмарні обчислення можуть бути більш економічно ефективними, ніж власна інфраструктура, оскільки користувачі сплачують лише за ті ресурси, які вони використовують. Хмарні обчислення можна легко масштабувати вгору або вниз відповідно до потреб, що може допомогти організаціям заощадити гроші. Хмарні обчислення доступні з будь-якого місця, де є Інтернет.

Інтернет речей – це мережа фізичних пристроїв, які підключені до Інтернету і можуть збирати та обмінюватися даними. Ці пристрої можуть використовуватися для широкого спектру цілей, таких як моніторинг стану обладнання, управління ланцюгами поставок та надання послуг клієнтам. Хмарні обчислення та IoT тісно пов'язані між собою. Хмарні обчислення можна використовувати для зберігання даних, які збирають пристрої IoT, а також для обробки цих даних [1]. Це дозволяє організаціям отримувати цінну інформацію з пристроїв IoT, яка може бути використана для покращення їхньої ефективності та продуктивності.

Хмарні обчислення дають користувачам змогу отримувати доступ до програм та даних через Інтернет. Це означає, що користувачі можуть використовувати свої «розумні» пристрої, смартфони та планшети, для доступу до своїх IT-ресурсів. Користувачі можуть отримувати доступ до своїх IT-ресурсів з будь-якого місця, де є Інтернет. Користувачі можуть отримувати доступ до своїх IT-ресурсів за допомогою одного інтерфейсу.

Література

1. Khadidos, A.O.; Shitharth, S.; Khadidos, A.O.; Sangeetha, K.; Alyoubi, K.H. Sangeetha, and Khaled H. Alyoubi. Healthcare Data Security Using IoT Sensors Based on Random Hashing Mechanism. J. Sens. 2022, 2022, 8457116.

Корба Д., Мудрик І.
ПРОЕКТУВАННЯ ТА РОЗРОБКА СИСТЕМИ МОНІТОРИНГУ РУХОМИХ ОБ'ЄКТІВ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ JAVA, SPRING ТА ПРОТОКОЛУ GTFS
Korba D., Mudryk I.
DESIGN AND DEVELOPMENT OF MOBILE OBJECTS MONITORING SYSTEM USING JAVA, SPRING AND GTFS PROTOCOL TECHNOLOGIES 63

Віталій Кравчук
ПРОБЛЕМА ЗАХИСТУ КІБЕРПРОСТОРУ МАЛОГО ТА СЕРЕДНЬОГО БІЗНЕСУ
Vitaliy Kravchuk
CYBERSECURITY ISSUES FOR SMALL AND MEDIUM-SIZED BUSINESSES 64

О. Крамар, К. Козачук; Ю. Лавришук
КОНЦЕПТ VR-ПРОСТОРУ ЦЕНТРУ НАУКИ ТЕРНОПОЛЯ
O. Kramar, K. Kozachuk; Yu. Lavryshchuk
THE CONCEPT OF THE TERNOPIL SCIENCE CENTER'S VR SPACE 66

Т.О. Крамар, О.М. Дуда
ТЕХНОЛОГІЇ ДОПОВНЕНОЇ РЕАЛЬНОСТІ В «РОЗУМНОМУ МІСТІ»
T.O. Kramar, O.M. Duda
AUGMENTED REALITY TECHNOLOGIES IN THE SMART CITY 67

Крисюк М.В., Закопець А.І., Дуда Х.О.
СТАН ТА ПЕРСПЕКТИВИ ОБЧИСЛЮВАЛЬНИХ ПЛАТФОРМ ДЛЯ МІСЬКИХ ЗАДАЧ
Krysiuk M.V., Zakopets A.I., Duda Kh.O.
STATUS AND PROSPECTS OF COMPUTING PLATFORMS FOR URBAN TASKS 68

Крисюк М.В., Закопець А.І., Дуда Х.О.
ІНФОРМАЦІЙНО-ТЕХНОЛОГІЧНІ ПЛАТФОРМИ ТА ІМІТАЦІЙНЕ МОДЕЛЮВАННЯ ДЛЯ ПОТРЕБ «РОЗУМНИХ МІСТ»
Krysiuk M.V., Zakopets A.I., Duda Kh.O.
INFORMATION TECHNOLOGY PLATFORMS AND SIMULATION FOR THE NEEDS OF SMART CITIES 69

Кубарич З.П., Скарга-Бандурова І.С.
ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ЕФЕКТИВНОГО РЕАГУВАННЯ НА ІНЦИДЕНТИ У SIEM СИСТЕМІ
Z.P. Kubarych, I.S. Skarga-Bandurova
USING ARTIFICIAL INTELLIGENCE FOR EFFECTIVE INCIDENT RESPONSE IN SIEM SYSTEM 70

О.П. Кузьмич, Я.В. Литвиненко
МЕТОДИ СТАТИСТИЧНОГО ОПРАЦЮВАННЯ МЕДИЧНИХ СИГНАЛІВ
O.P. Kuzmych, Ia.V. Lytvynenko
METHODS OF STATISTICAL PROCESSING OF MEDICAL SIGNALS 71

О.А. Кучеренко, О.О. Кучеренко
ОСОБЛИВОСТІ ПЕРЕДОБРОБКИ ДАНИХ ДЛЯ МЕТОДІВ ПРОГНОЗУВАННЯ
O.A.Kucherenko, O.O.Kucherenko
FEATURES DATA PREPROCESSING FOR FORECASTING METHODS 72

Лебідко Д.М., Онуферко В.А., Перетятко Т.П.
ВЕЛИКІ ЗА ОБСЯГОМ ДАНІ, РЕЛЯЦІЙНІ ТА НЕРЕЛЯЦІЙНІ МОДЕЛІ
Lebidko D.M., Onuferko V.A., Peretiatko T.P.
BIG DATA, RELATIONAL AND NON-RELATIONAL MODELS 73

Лебідко Д.М., Онуферко В.А., Перетятко Т.П.
ХМАРНІ ПЛАТФОРМИ, ОБЧИСЛЕННЯ ТА ІНТЕРНЕТ РЕЧЕЙ
Lebidko D.M., Onuferko V.A., Peretiatko T.P.
CLOUD PLATFORMS, COMPUTING AND THE INTERNET OF THINGS 74