Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)
Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження методу побудови онтології для моделювання предметної галузі
(тема)

Виконав:
студент 2 курсу, групи СШІм-22-2
Хрущ Д.О.
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)
Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)
Освітня програма Системи штучного інтелекту
(повна назва спеціалізації)
Керівник доц. Кудрявцева М.С.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____ В.О. Філатов
(підпис) (прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Штучного інтелекту
(повна назва)

Рівень вищої освіти другий (магістерський)

Спеціальність 122 Комп'ютерні науки
(код і повна назва)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Системи штучного інтелекту
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«_____» _____ 20 ___ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Хрущу Денису Олеговичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методу побудови онтології для моделювання предметної галузі

затверджена наказом університету від 1 квітня 20 24 р. № 260Ст
2. Термін подання студентом роботи до екзаменаційної комісії 7 червня 20 24 р.
3. Вихідні дані до роботи Науково-технічні публікації, дані статей, результати експериментальних досліджень по технологіям, методам, моделям

4. Перелік питань, що потрібно опрацювати в роботі
1) Вступ, мета роботи та постановка задачі, визначення бізнес-логіки
2) Теоретичні дослідження
3) Аналіз технологій та засобів реалізації для побудови онтології

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	01.04.2024	виконано
2	Анализ предметної галузі та постановка задачі	10.04.2024	виконано
3	Теоретичні дослідження	20.04.2024	виконано
4	Проектування веб-додатку	30.04.2024	виконано
5	Розробка методу побудови онтології	10.05.2024	виконано
6	Підготовка пояснювальної записки	31.05.2024	виконано
7	Попередній захист кваліфікаційної роботи	05.06.2024	виконано
8	Захист перед ЕК	12.06.2024	

Дата видачі завдання 1 квітня 20 24 р.

Студент _____
(підпис)

Керівник роботи _____ доц. Кудрявцева М.С.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 116 с., 10 рис., 4 табл., 2 дод., 11 джерел.

КОНЦЕПТИ, НАПОВНЕННЯ, ОНТОЛОГІЇ, ОНТОЛОГІЧНЕ НАВЧАННЯ, FCA, RCA.

Тема кваліфікаційної роботи: онтологічне навчання із застосуванням методу аналізу формальних концептів.

Об'єктом дослідження є метод аналізу формальних концептів на навчання онтологій.

Метод роботи: теоретичні та експериментальні дослідження.

Результат роботи: розроблено метод побудови онтологій, їх навчання та «заселення» на основі FCA.

Метою даної роботи є розвиток підходу до побудови онтологій, їх навчання (ontology learning) і «заселення» (population), тобто. наповненню конкретними екземплярами концептів, на основі методу FCA.

ABSTRACT

Master`s thesis contains: 116 pp., 10 fig., 4 tabl., 2 ann., 11 references.

CONCEPTS, CONTENT, FCA, ONTOLOGICAL LEARNING CONTENT ONTOLOGY CONCEPTS, RCA.

The topic of the Master`s thesis: ontological teaching using the method of formal concept analysis.

The object of research is the method of analyzing formal concepts for teaching ontologies.

Method of work: theoretical and experimental research.

Result: a method of building ontologies, their training and “populating” based on FCA.

The purpose of this work is to develop an approach to building ontologies, their learning and “population”, i.e. filling them with specific instances of concepts, based on the FCA method.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛОВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

БЗ – база знань;

API – Application Programming Interface – інтерфейс прикладного програмування;

FCA – Formal Concept Analysis – аналіз формальних понять;

RCA – Relational Concept Analysis – реляційний аналіз концептів.

ВСТУП

Аналіз формальних понять як напрямок у математиці було запроваджено професором Рудольфом Вілле і надалі суттєво розвинений ним, його колегами та учнями. Аналіз формальних понять можна, зокрема, розглядати як метод аналізу та графічної інтерпретації знань, представлених за допомогою двовимірних таблиць «об'єкт-властивість». Ідеї цього напряму виявилися настільки привабливими, що аналіз формальних понять висувається деякими його апологетами на роль одного з основних методологічних принципів у побудові теорій (у тому числі математичних), в описі навколишнього світу, а також у викладі математики.

У кваліфікаційній роботі було продемонстровано можливість застосування методів формального аналізу понять до даних, отриманих політичними блогами. Як результат було отримано демонстрацію наочного подання процесів, які відбуваються в політиці в ході виборів на пост президента в США. Наочне представлення великих даних, отриманих по блогах, може допомогти експерту зосередити увагу на найцікавіших фактах та подіях.

Ґрунтуючись на отриманих результатах, було також зроблено спроби спрогнозувати подальші шляхи розвитку подій, які підтвердилися надалі. Крім цього, було представлено метод, що дозволяє виявляти зв'язки між політиками, напрямами їх роботи та проводити оцінку таких зв'язків. У роботі описані методи, що дозволяють побудувати онтологію на основі текстового корпусу, який представляє певну предметну область. Показано, що на основі методу FCA можна коректно групувати об'єкти, розглядаючи їх загальні властивості, та будувати таксономію концептів, пов'язаних з транзитивним ставленням підпорядкованості (is-a).

Другий пропонований метод дозволяє отримувати зовнішні відносини між концептами за допомогою реляційного аналізу концептів і, таким чином, розширити можливості використання онтології, наприклад,

отримувати відповіді на більш складні запити. Дескрипторна логіка була обрана мовою опису онтології завдяки відносній простоті побудови правил виведення. Перевагою методу є його універсальність та незалежність від сфери застосування.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ ТА ПОСТАНОВКА ЗАВДАННЯ

1.1 Поняття онтологій

На сьогоднішній день онтології є найбільш ефективним засобом для представлення та обміну знаннями у Web-просторі. Найчастіше онтології визначаються як експліцитна специфікація концептуалізації. Зазвичай це означає, що онтологія описує концепти (поняття) та відносини на виявленій множині концептів, релевантних предметній області (ПрО). Однак, для забезпечення можливості обміну знаннями необхідно описати ці концепти та стосунки більш ретельно, ніж просте впорядкування їх у таксономію. Проте зазвичай розробка онтологій починається і закінчується побудовою таксономій. У зв'язку з цим є доцільним застосування для побудови, уточнення, поповнення та «заселення» онтологій методу аналізу формальних концептів (Formal Concept Analysis; FCA), що спочатку призначений для аналізу даних і дозволяє ідентифікувати концептуальні структури на безлічі даних. FCA також дозволяє виявляти необхідність у нових онтологічних концептах і відносинах, що, у свою чергу, призводить до більш повної онтології, що містить опис цих сутностей таким чином, щоб уможливити обмін знаннями та спільне використання (sharing) розробленої онтології зацікавленими групами користувачів, що працюють у однієї чи подібних ПрО. Наприклад, концепт може бути описаний не лише його позицією у таксономічній (is-a) ієрархії, але також за допомогою відносин, які можуть бути застосовані до цього концепту. Аналогічним чином, відношення може бути описане концептами, пов'язаними цим ставленням

Аналіз формальних понять (АФП) є прикладною гілкою алгебраїчної теорії решіток, в рамках якої запропоновано математичний формалізм, що описує мовою алгебри поняття та ієрархії понять. Основні ідеї АФП було сформульовано Рудольфом Вілле у його роботі, а найповнішою монографією з АФП є книга Гантера і Вілле.

Фактично аналіз формальних понять має справу з даними в об'єктно-ознаковій формі, а формальні поняття, визначені за допомогою відповідності Галуа, є парою множин виду (обсяг, зміст), їм точно до перестановки рядків і стовпців відповідають максимальні прямокутники в таблиці об'єкт-ознака. Основними перевагами такого визначення поняття є відповідність традиційним уявленням про поняття, що використовуються у філософії: 1) поняття – це пара виду (обсяг, зміст); 2) при зменшенні обсягу поняття збільшується його зміст і навпаки; 3) поняття ієрархічно впорядковані по відношенню «бути більше загальним поняттям».

За останні 30 років АФП пройшов значний шлях від початкових теоретичних досліджень до різноманітних численних додатків (тільки англійською мовою видано близько 900 наукових праць з тематики АФП, більше половини з яких присвячені додаткам), що дозволяє повноправно назвати його прикладною математичною дисципліною. Основними додатками АФП, яким ми приділимо увагу в цій роботі, є аналіз даних (машинне навчання та розробка даних), подання знань (онтології та таксономії), інформаційний пошук, аналіз неструктурованих даних (зокрема текстів), програмна інженерія, соціологія та освіта . В даний час існують три найбільш репрезентативних міжнародних конференції з тематики АФП: International Conference on Formal Concept Analysis, International Conference on Concept Lattices and Their Applications і International Conference on Concept-Tual Structures. Перша в списку конференція є найбільш представницькою і служить для обговорення значних теоретичних і практичних результатів в області, друга присвячена переважно додаткам АФП, а третя, крім АФП-спільноти, покликана зібрати дослідників у галузі представлення знань та онтологічного моделювання (наприклад, співзасновником цієї серії конференцій є творець понятійних графів Джон Сова).

Онтологія – це формальний опис знань як набору понять у межах предметної області та зв’язків, які існують між ними. Він забезпечує

загальне розуміння інформації та робить чіткі припущення щодо домену, що дозволяє організаціям краще зрозуміти свої дані.

Онтологія – це формальний опис знань як набору понять у межах предметної області та зв’язків, які існують між ними. Щоб увімкнути такий опис, нам потрібно формально визначити такі компоненти, як індивідууми (екземпляри об’єктів), класи, атрибути та відносини, а також обмеження, правила та аксіоми. Як наслідок, онтології не тільки вводять загальне та багаторазове подання знань, але також можуть додавати нові знання про предметну область.

Онтологічну модель даних можна застосувати до набору окремих фактів для створення графа знань – набору сутностей, де типи та зв’язки між ними виражаються вузлами та ребрами між цими вузлами. Описуючи структуру знань у домену, онтологія готує основу для графа знань для збору даних у ньому. Існують, звичайно, інші методи, які використовують формальні специфікації для представлення знань, такі як словники, таксономії, тезауруси, тематичні карти та логічні моделі. Однак, на відміну від таксономій або схем реляційних баз даних, наприклад, онтології виражають зв’язки та дозволяють користувачам пов’язувати кілька концепцій з іншими концепціями різними способами.

Будучи одним із будівельних блоків семантичної технології, онтології є частиною стеку стандартів W3C для семантичної мережі. Вони надають користувачам необхідну структуру для зв’язування однієї частини інформації з іншими частинами інформації в мережі пов’язаних даних. Оскільки вони використовуються для визначення загальних представлень моделювання даних з розподілених і гетерогенних систем і баз даних, онтології забезпечують взаємодію баз даних, пошук між базами даних і плавне управління знаннями.

Місце онтологій в структурі Semanic Web представлено на рисунку 1.1.

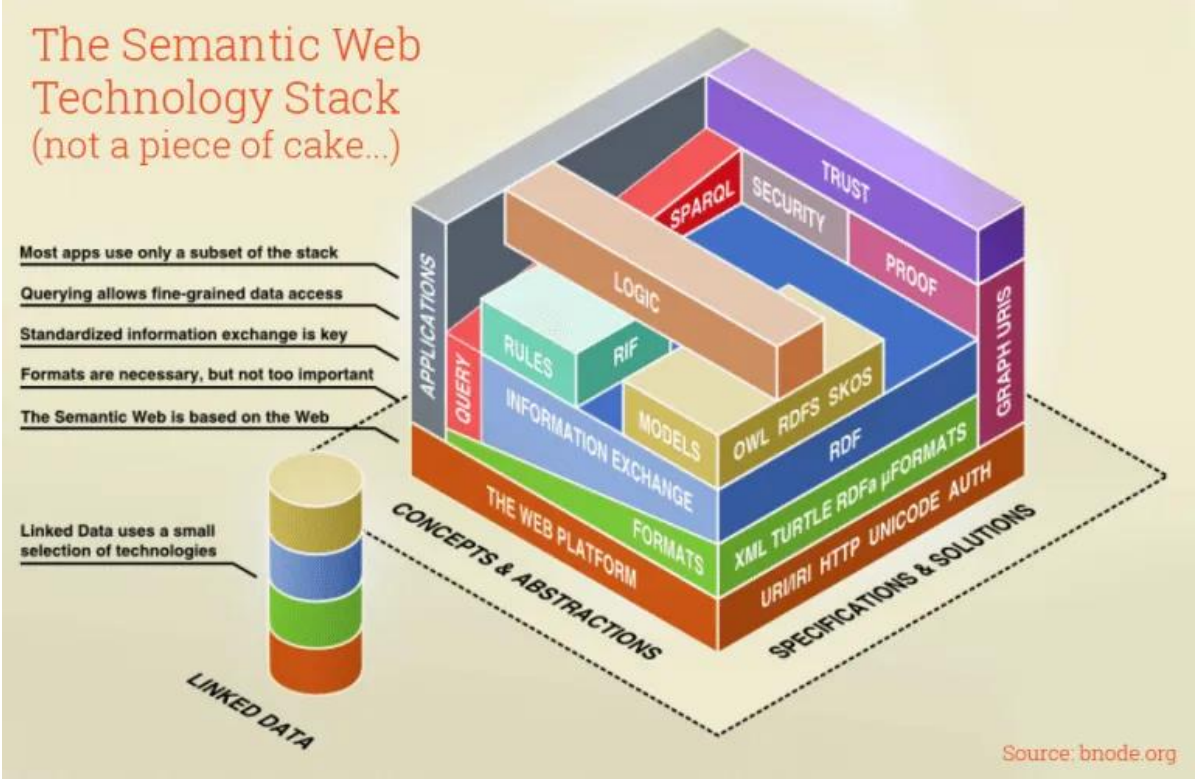

Рисунок 1.1 – Місце онтологій в стеку Semantic Web

1.2 Онтології для управління даними

Деякі з основних характеристик онтологій полягають у тому, що вони забезпечують загальне розуміння інформації та роблять чіткі припущення щодо домену. Як наслідок, взаємозв'язок і сумісність моделі роблять її безцінною для вирішення проблем доступу до даних і запитів до даних у великих організаціях. Крім того, покращуючи метадані та походження, а отже дозволяючи організаціям краще розуміти свої дані, онтології покращують якість даних.

В останні роки відбулося поширення вираження онтологій за допомогою мов онтології, таких як мова веб-онтології (OWL). OWL – це семантична веб-обчислювальна логічна мова, призначена для представлення багатих і складних знань про речі та зв'язки між ними. Він також надає детальні, послідовні та значущі відмінності між класами, властивостями та зв'язками.

Вказуючи як класи об’єктів, так і властивості зв’язків, а також їх ієрархічний порядок, OWL збагачує онтологічне моделювання в базах даних семантичних графів, також відомих як RDF triplestores. OWL, що використовується разом із резонером OWL у таких потрійних сховищах, дозволяє перевіряти узгодженість (щоб знайти будь-які логічні неузгодженості) і забезпечує перевірку задовільності (щоб знайти, чи є класи, які не можуть мати екземпляри).

Крім того, OWL оснащений засобами для визначення еквівалентності та відмінності між примірниками, класами та властивостями. Ці зв’язки допомагають користувачам зіставляти концепції, навіть якщо різні джерела даних описують ці концепції дещо по-різному. Вони також забезпечують усунення неоднозначності між різними примірниками, які мають однакові імена чи описи.

Онтології для кращого управління даними.

Деякі з основних характеристик онтологій полягають у тому, що вони забезпечують загальне розуміння інформації та роблять чіткі припущення щодо домену. Як наслідок, взаємозв’язок і сумісність моделі роблять її безцінною для вирішення проблем доступу до даних і запитів до даних у великих організаціях. Крім того, покращуючи метадані та походження, а отже дозволяючи організаціям краще розуміти свої дані, онтології покращують якість даних.

Стандарт OWL та моделювання онтології.

В останні роки відбулося поширення вираження онтологій за допомогою мов онтології, таких як мова веб-онтології (OWL). OWL – це семантична веб-обчислювальна логічна мова, призначена для представлення багатих і складних знань про речі та зв’язки між ними. Він також надає детальні, послідовні та значущі відмінності між класами, властивостями та зв’язками.

Вказуючи як класи об’єктів, так і властивості зв’язків, а також їх ієрархічний порядок, OWL збагачує онтологічне моделювання в базах даних

семантичних графів, також відомих як RDF triplestores. OWL, що використовується разом із резонером OWL у таких потрійних сховищах, дозволяє перевіряти узгодженість (щоб знайти будь-які логічні неузгодженості) і забезпечує перевірку задовільності (щоб знайти, чи є класи, які не можуть мати екземпляри).

Однією з головних особливостей онтологій є те, що завдяки вбудованим у них суттєвим зв’язкам між поняттями вони дозволяють автоматизувати міркування щодо даних. Таке міркування легко реалізувати в базах даних семантичних графів, які використовують онтології як свої семантичні схеми.

Більше того, онтології функціонують як «мозок». Вони «працюють і міркують» з поняттями та відносинами у спосіб, близький до того, як люди сприймають взаємопов’язані поняття.

На додаток до функції міркування, онтології забезпечують більш узгоджену та просту навігацію, коли користувачі переходять від однієї концепції до іншої в структурі онтології.

Ще одна цінна особливість полягає в тому, що онтології легко розширити, оскільки зв’язки та відповідність понять легко додати до існуючих онтологій. Як наслідок, ця модель розвивається зі зростанням даних, не впливаючи на залежні процеси та системи, якщо щось піде не так або потребує змін.

Онтології також надають засоби для представлення будь-яких форматів даних, включаючи неструктуровані, напівструктуровані або структуровані дані, забезпечуючи більш плавну інтеграцію даних, легшу концепцію та аналіз тексту, а також аналітику на основі даних.

1.3 Постановка задачі

Метою даної роботи є розвиток підходу до побудови онтологій, їх навчання (ontology learning) і «заселення» (population), тобто. наповнення

конкретними екземплярами концептів, на основі методу FCA. Такий підхід дозволить виявляти нові концепти та відносини, не задані явним чином, та вбудовувати їх в онтологічну структуру, представляючи більш повний та адекватний опис концептуальної моделі ПрО для вирішення комплексу поставлених завдань. Для досягнення поставленої мети у цій роботі планується вирішення наступних завдань:

– аналіз описів онтологій у загальному вигляді;

– опис методу FCA та його можливостей для онтологічного навчання;

– розробка методу побудови онтологій, їх навчання та «заселення» на основі FCA;

– практична апробація запропонованого методу онтологічного навчання;

– перспективи подальшого використання запропонованого підходу в онтологічному інжинірингу під час вирішення прикладних завдань

2 ДОСЛІДЖЕННЯ ПОЛІТИЧНИХ БЛОГІВ НА ОСНОВІ ФОРМАЛЬНОГО АНАЛІЗУ ДАНИХ

Метою дослідження було застосування методів формального аналізу понять (ФАП) до обробки інформації, отриманої політичними блогами США. Кошти ФАП використовувалися раніше в аналізі інформації про відвідуваність інтернет ресурсів, структури аудиторій сайтів та виділення різних груп серед цільової аудиторії. Ця стаття є спробою застосувати такі методи політичних даних. Перевагою даних методів є наочне та зручне для вивчення уявлення результатів у вигляді решіток. Оскільки дані надані за період, у якому відбувалися передвиборчі перегони, одним із наших завдань стало визначення найбільш обговорюваних політиків та зміна складу лідерів цих перегонів. На відміну від статистичних методів, формальний аналіз понять дозволяє будувати структури інтересів блогерів у вигляді решіток формальних понять, що дає можливість наочно показати всю структуру інтересів

2.1 Дані, їх властивості

Дані були надані кампанією RTGI. Ця французька аналітична кампанія займалася аналізом американських політичних блогів. Для аналізу з текстів блогів було обрано 79 слів, які, на думку фахівців RTGI, мають відображати тематику блогів. Потім було зібрано дані за період з 1 листопада 2007 року по 29 травня 2008 року.

Оскільки дані мають тимчасову характеристику, ми можемо не лише побудувати тематичну структуру політичних блогів, а й простежити зміни цієї структури з часом.

Основним методом аналізу блогів буде формальний аналіз понять, створення формальних контекстів та побудова за ними решіток формальних понять.

2.2 Створення формального контексту

Формальний контекст − це трійка K = (G, M, I):

– *G* − безліч об'єктів;

– *M* − безліч ознак;

– *I* − відношення володіння ознакою, $I \cap G \cap M$;

– *I* = {(gi, mj)}. Пара *(gi, mj)* показує, що об'єкт *gi* має ознаку *mj*.

Часто формальний контекст подають у вигляді бінарної матриці (таблиця 2.1).

Таблиця 2.1 Приклад формального контексту

	M1	M2	M3	M4
G1			x	x
G2		x	x	
G3	x			x
G4	x	x	x	

У нашому випадку об'єктами будуть блогери, а ознаками – 79 ключових слів.

Насамперед, оберемо тимчасовий період (день, тиждень). До кожного блогера порахуємо, скільки за цей період блогер вживав кожне слово, тобто. ми отримуємо матрицю об'єкти-ознаки Q, в якій елемент q_{ij} дорівнює кількості вжитків і-м блогером j-го слова.

З цієї матриці необхідно отримати бінарну. І тому ми встановлювали поріг на кількість вживань слова. Якщо блогер використовував слово більше заданого порога, вважаємо, що блогер активно використовував це слово, отже, обговорював тему, до якої належить це слово. Поріг необхідний, щоб відсікти випадки випадкового вживання слова.

2.3 Побудова решітки формальних понять

Тепер, маючи контекст, необхідно побудувати структуру інтересів, виділити групи блогерів зі схожими інтересами. Для цього побудуємо решітки формальних понять.

У визначенні формального поняття використовують оператори Галуа. Для A∩G та B∩M:

– $A' = \{m \cap M \mid g \cap A: gIm\}$;

– $B' = \{g \cap G \mid m \cap B: gIm\}$.

Іншими словами A' – безліч ознак, якими мають всі об'єкти з множини A. B' – безліч об'єктів, які мають всі ознаки з множини B.

Формалінне поняття *(A,B)* складається з безлічі об'єктів A∩G та безлічі ознак B∩M, таких що B'=A та A'=B. A називається обсягом, а B – змістом поняття. У матриці контексту формальне поняття є підматрицею, що складається з одиниць. Приклад формальне поняття ([g2, g4], [m2, m3]) наведено у таблиці 2.2.

Таблиця 2.2 Приклад формального поняття

	M1	M2	M3	M4
G1			x	x
G2		x	x	
G3	x			x
G4	x	x	x	

Безліч понять контексту K утворюють решітку формальних понять, оскільки створюють частковий порядок вкладення обсягів понять і мають найменше і найбільше за вкладенням поняття.

Приклад решітки формальних понять наведено на рисунку 2.1.

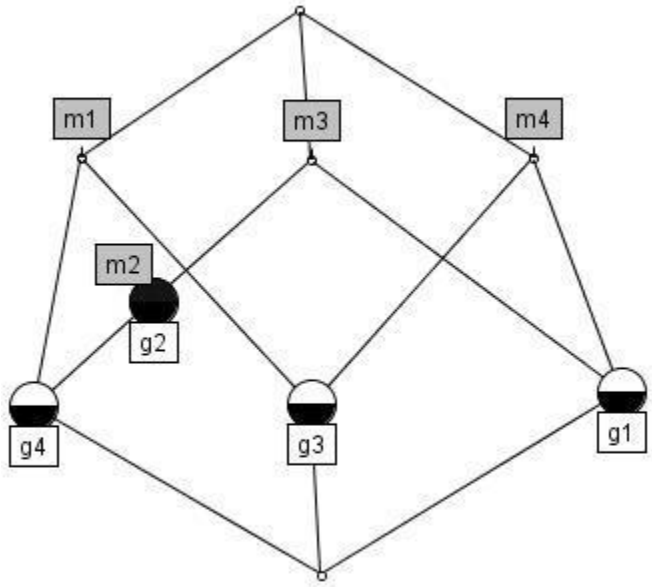

Рисунок 2.1 −Приклад решітки формальних понять

Як читаються ці решітки? Кожна вершина решітки – формальне поняття. Поруч із поняттям пишуться об'єкти, яких немає у менш загальних поняттях (що знаходяться під даним поняттям), та ознаки, яких немає у більш загальних поняттях. Тоді обсяг формального поняття – всі об'єкти, написані навпроти даного поняття та всіх понять, менш загальних, ніж воно. Зміст − ознаки, написані навпроти даного поняття та більш загальних понять.

Наприклад, поняття з підписами «m2» та «g2» має обсяг [g2, g4] та зміст [m2, m3].

Знаходяться такі формальні поняття алгоритмом «замикай по одному». Функція починає працювати з найзагальнішого формального поняття, яке містить усі об'єкти і найчастіше жодної ознаки. Потім перебувають усі інші поняття рекурсивним додаванням ознак.

Але використовувати всю решітку формальних понять не завжди зручно через її громіздкість. Кількість формальних понять експонентно залежить від розміру матриці. Наприклад, контекст, складений за даними за

1 день (1 листопада), складається з 49 блогерів та 65 слів. Грати, побудовані за цим контекстом, мають 202 формальні поняття (рисунок 2.2).

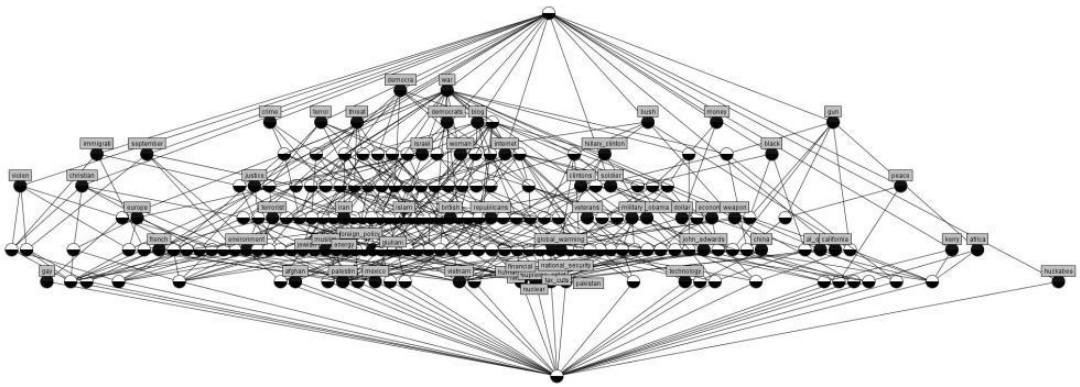

Рисунок 2.2 – Ґрати формальних понять за 1 листопада 2007

Контексти за більші періоди (наприклад, тиждень) можуть містити понад 1 млн. понять. Звичайно, аналізувати їх неможливо.

Ґрати громіздкі, але багато формальних понять не несуть практично ніякої інформації або виникли через шум. Тому раціонально знехтувати малозначущими формальними поняттями для отримання більш простої решіток.

Один із способів відбору найбільш важливих понять – індекс стійкості.

2.4 Дослідження блогів

Тепер ми можемо знаходити всі формальні поняття, вибирати найбільш стійкі з них, і отримати компактніші решітки, які при цьому зберігають основну структуру інтересів.

Наприклад, виберемо з решіток понять, побудованої за даними за 1 листопада, поняття, з інтенсіональною стійкістю більше 0.9. Отримуємо цілком читабельну та зручну для аналізу решітку (рисунок 2.3) [2].

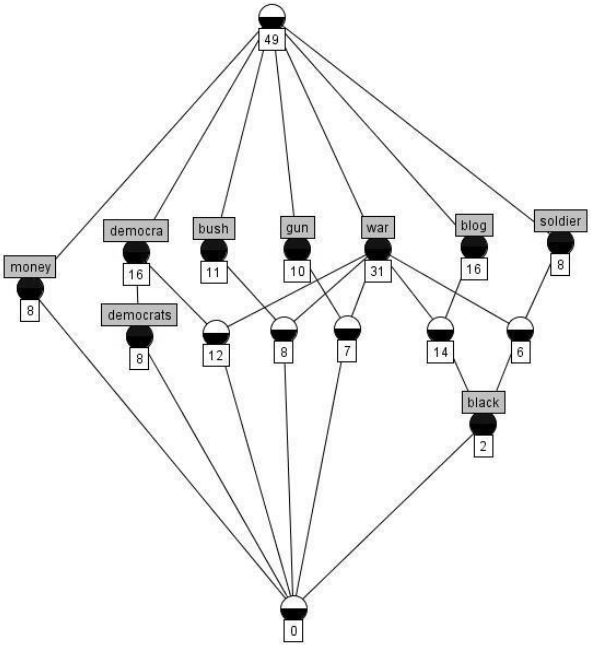

Рисунок 2.3 − Грати зі стійкими поняттями за 1 листопада 2007

Можемо помітити, що найпопулярніше слово − War. Мабуть, у цей день активно обговорювалися військові дії, які велися на той час у світі. Так само з війною пов'язане слово «bush», мабуть, ці люди обговорювали війни за участю США та зовнішню політику, яку вів Буш.

Тепер побудуємо решітки за сотий день, коли блогери були активнішими (рисунок 2.4).

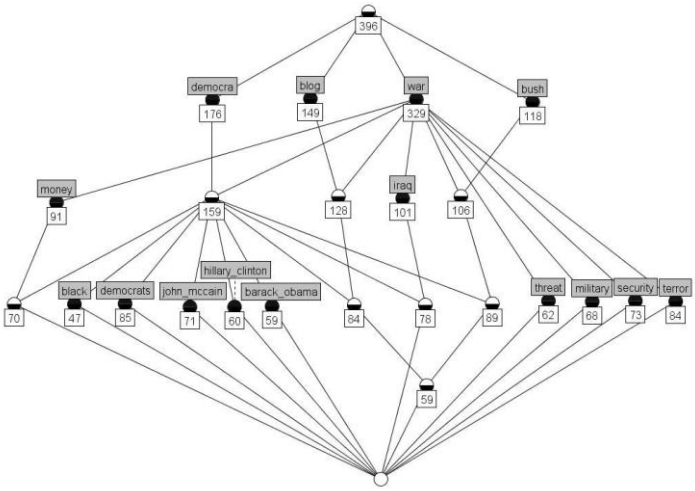

Рисунок 2.4 − Решітки формальних понять за 8 лютого 2023 року

По решітках видно, що головним чином у блогах найбільш популярні 2 області: війна (права частина решіток) і тема майбутніх виборів (ліва частина решіток)

Для того, щоб подивитися, як змінювалися переваги блогерів під час передвиборчих перегонів, залишимо в нашому контексті лише імена кандидатів на пост президента США. Таких серед 79 слів сім: Барак Обама, Хіларі Клінтон, Джон Едвардс, Джон Маккейн, Мітт Ромні, Руді Джуліані та Майк Хакабі.

Розіб'ємо наші дані на 17 періодів по 7 днів. Складемо контексти за кожним періодом та залишимо в них лише по 7 ознак, імен політиків. Побудуємо решітки по кожному із 17 періодів. За листопадовими решітоками видно чіткий поділ політиків на демократів та республіканців (рисунок 2.5).

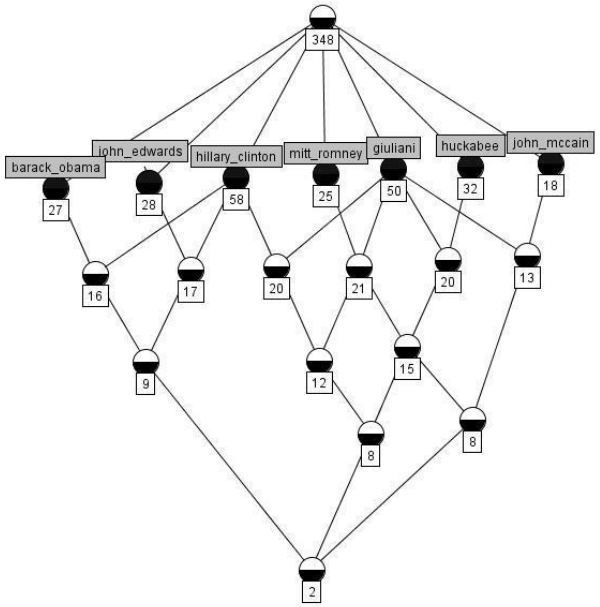

Рисунок 2.5 − Обговорення політиків, листопад

При цьому можна помітити, що на відміну від інших демократів, Хіларі Клінтон має багато спільних з республіканцями формальних понять. Звідси можна припустити, що на початку передвиборних перегонів Клінтон

була найочікуванішим претендентом від демократів. При цьому серед республіканців найбільш популярним є Джуліані.

Найбільш сильні зміни в структурі відбулися на першому тижні лютого, різко впав інтерес до Джуліані та Едвардсу (рисунок 2.6), і в середині лютого, впав інтерес до Мітт Ромні (рисунок 2.7).

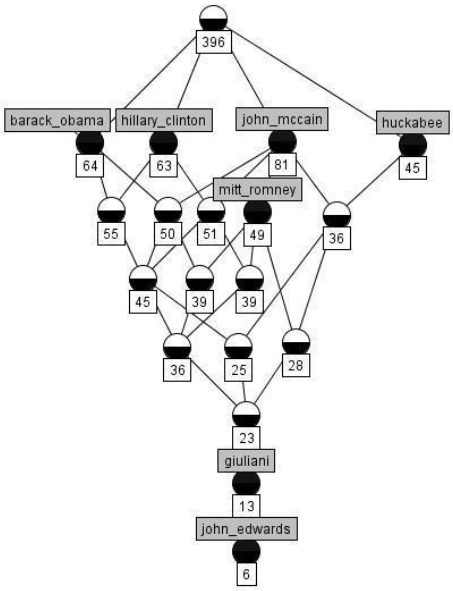

Рисунок 2.6 − Обговорення політиків, початок лютого

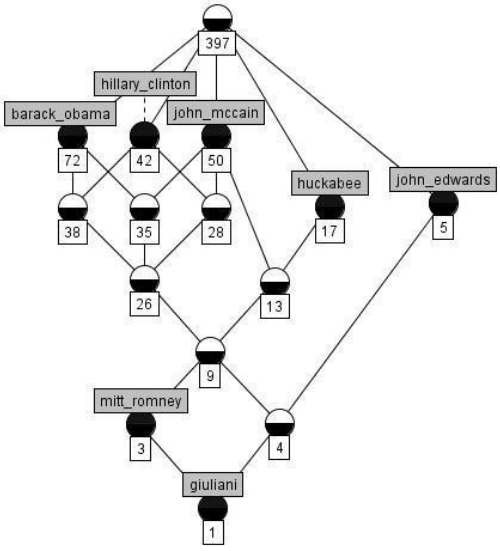

Рисунок 2.7 − Обговорення політиків, середина лютого

Причиною такої зміни став вихід цих політиків із передвиборчих перегонів.

При цьому явно видно, що серед претендентів, що залишилися, Обама, Клінтон і Маккейн досить активно обговорюються блогерами, в той час, як Хаккабі набагато менш популярний, тобто. можна припустити, що Хаккабі наступним закінчить гонку. Як показує історія, він закінчив свою передвиборчу кампанію 4 березня 2008 року, через місяць.

Таким чином, по решітоках можна з точністю до тижня сказати, коли який політик вибув із перегонів, але постає питання, якого політика стали підтримувати виборці, які підтримували політиків?

2.5 Оптимальні набори ознак

Крім індексу стійкості, можна спробувати скоротити сам контекст. Для цього ми можемо використовувати вже готові методи оптимізації, що виникли у різних наукових галузях. Але, по-перше, необхідно визначитися за якими критеріями слід відбирати ознаки (далі розумітиметься саме вибір оптимального набору ознак, у нашому випадку це слова, які вжили блогери). Наприклад, можна шукати такий мінімальний набір ознак, якими будуть володіти всі об'єкти, оскільки слова такого набору, як очікується, відіграватимуть ключову роль в описі всієї спільноти. Так само необхідно, щоб перетин об'єктів у них був найменший.

Грунтуючись на цих умовах важливості, були реалізовані та апробовані оптимізаційні методи такі як: метод Монте-Карло, метод якнайшвидшого спуску, генетичний алгоритм. Як експеримент, було реалізовано вибір 20-ти ознак з решіток ФП по американських блогах за весь період. Результатом такої вибірки стали такі слова: 'democra', 'wall_street', 'dollar', 'islam', 'financial', 'iran', 'afghan', 'democrats', 'september', 'national_security', 'money' , 'george_w_bush', 'violen', 'mexico', 'pakistan', 'huckabee', 'immigrati', 'mitt_romney', 'energy', 'palestin'.

Якщо тепер розглянути решітки, що складається лише з цих 20 ознак, а об'єкти залишити без зміни, то можна значно швидше виділити отримати всі формальні поняття з таких даних і побудувати діаграму решітки ФП (рисунок 2.8).

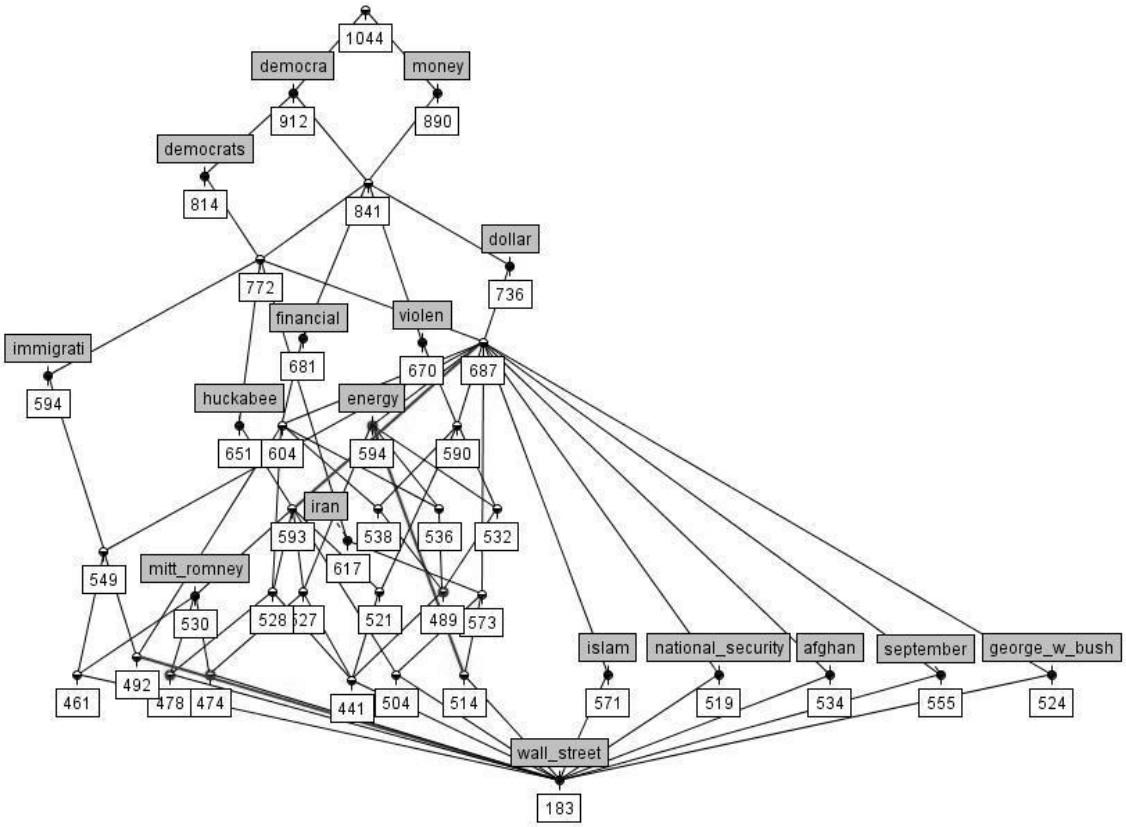

Рисунок 2.8 − Решітки з вибраними ознаками

За такою діаграмою вже скорочених решіток формальних понять, ми можемо, наприклад, перевірити адекватність згаданого підходу до редукування даних. Наприклад, як видно з діаграми, більшість блогерів, говорячи про демократів, згадували також слово демократія. Аналогічно можна сказати і про долар, про який говорили у контексті із грошима. При детальному вивченні діаграми ми бачимо, що слова «іслам», «національна безпека», «афган», «вересень», «Джордж Буш» вжили з 687 осіб приблизно по 500 осіб і щонайменше 183 особи висловили їх в одному контексті. Іншими словами, тепер, відштовхуюсь від отриманих результатів, можна

спробувати сконцентруватися на перевірці гіпотези: «Чи правда, що Дж. Буша обговорювали лише в контексті його політики в галузі національної безпеки та результатів його роботи у цій сфері». Слово 'huckabee' часто вживалося разом з 'money' та 'dollar'. Можна припустити, що блогери цікавилися запропонованими кандидатом варіантами виведення США із кризи, пов'язаної із вливанням грошей в економіку держави.

2.6 Метод k-середніх

Для роботи з політичними блогами також можна застосувати метод k-середніх до проблеми оптимізації даного набору ознак. Усі ознаки розбиваються на кластери. Потім створюється контекст, в якому як ознаки використовуються отримані кластери. Якщо блогер використовував більше половини слів із кластера, то він має ознаку, що відповідає даному кластеру.

Як експеримент ми виділяли за різні періоди часу по 20 кластерів з набору ознак. Як назву кластера використовувалося одне зі слів кластера. Були отримані кластери:

[cluster's name]: [words in the cluster]

pakistan: afghan pakistan

september: nuclear september

palestin: israel palestin

technology: british crime environment internet technology

super_tuesday: democrats primarias republicans super_tuesday

french: europe french middle_east military security soldier terror terrorist threat violen weapon

al_qaeda: al_qaeda dollar iran islam muslim

supreme_court: gun justice supreme_court

та ін.

Тепер побудуємо решітку формальних понять, відбираючи найстійкіші (рисунок 2.9) [3].

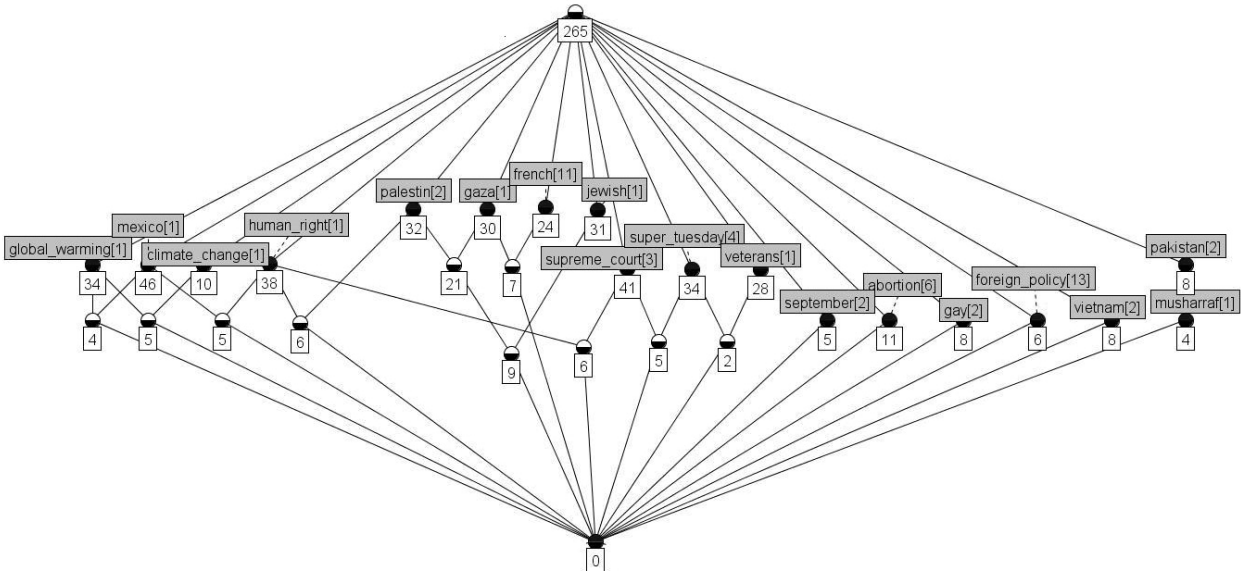

Рисунок 2.9 − Решітки з групами ознак

Після таких перетворень, дані щодо політичних блогів легше піддаються аналізу. З діаграми бачимо формальні поняття зі змістом [«Mexico», «human_right»] і [«Palestin», «human_right»]. Можна розглянути таку гіпотезу: «учасників американських блогів якщо й цікавлять права людини, то в Мексиці та Палестині». Також ми бачимо зв'язок кластера “french”, “gaza” та “palestin”, який, як можна припустити, відповідає за обговорення близькосхідних проблем та участі Європи у їх вирішенні.

3 МЕТОД FCA ДЛЯ ПОБУДУВАННЯ ОНТОЛОГІЙ НА ОСНОВІ ТЕКСТОВОГО КОРПУСУ

Важливим етапом розвитку будь-якої предметної області є структуроване уявлення знань, акумульованих у цій галузі, як бази знань (БЗ). Це дозволяє фахівцям області краще маніпулювати знаннями, виявляти нові закономірності, обмінюватись досвідом. Наприклад, у галузі радіаційного захисту, при медичному опроміненні важливим є розуміння властивостей радіоактивних ізотопів та специфіки їх використання для діагностики та лікування різних захворювань, від чого залежить як терапевтичний ефект для пацієнтів, так і безпека персоналу. Одним з ефективних способів представлення знань є онтології, до розробки яких останнім часом проявляється великий інтерес [1]. Поняття онтології в інтерпретації завдань обробки інформації найкоротше і виразно визначив Т. Грубер [2]. Він визначив онтологію як «явну специфікацію деякої концептуалізації». Під концептуалізацією у Грубера розуміється сукупність загальних понять, об'єктів і зв'язків, притаманних певної предметної області (ПЗ). Побудова онтології передбачає визначення класів об'єктів та опис їхніх відносин за допомогою однієї з формальних мов, наприклад, дескриптивної логіки, що дозволяє відповідати на запити, чи так

Звані компетентні питання (competency questions), які спочатку складаються природною мовою і потім «переводяться» на формальну мову, що використовується. Виходячи з цього, процес побудови онтології («ручний», напівавтоматичний або автоматичний) передбачає вихідне створення базової онтологічної структури, що становить основні поняття ПЗ та зв'язок між ними. Традиційно з цією метою залучаються експерти відповідних предметних областей. Проте останніми роками фахівці у сфері штучного інтелекту (ІІ) ставлять собі завдання знайти ефективні методи добування знань із спеціалізованих текстових корпусів без дорогого залучення фахівців. Підставою для цього може бути те, що

саме в текстах зберігається значна частина всіх знань у стабільній, відносно впорядкованій формі. При цьому частка електронних текстів, доступних для безпосередньої обробки, постійно зростає.

3.1 Вирішення задачі побудови онтології

Метою даної кваліфікаційної роботи є обґрунтування вибору мови опису концептів і розробка методу побудови онтології для деякої предметної області, представленої текстовим корпусом. В даному випадку був використаний текстовий корпус, що складається зі статей, стандартів та інших джерел в галузі радіології та радіаційного захисту, доступних в електронному вигляді, що допускають використання інструментів термінологічного аналізу для автоматичного виявлення та визначення властивостей термінів. Конкретні завдання, які стоять перед авторами, полягають у наступному. По-перше, було цікаво запропонувати опис концептів за допомогою дескриптивної логіки (ДЛ) [4] з подальшим порівняльним аналізом опису. По-друге, запропонувати метод побудови правил виведення для концептів базової онтології, які дозволяють отримати відповіді на компетентні питання наступного типу:

– які об'єкти належать до того ж класу, що і «пухлина печінки»?;

– чи належать об'єкти «бронхогенна карцинома» та «пухлина шлунка» одній групі?;

– чи є об'єкт «легким» органом?

І, нарешті, запропонувати новий підхід – аналіз реляційних понять, що є розширенням аналізу формальних понять (FCA), який дозволяє описати так звані зовнішні, чи трансверсальні, відносини між концептами. Запропонований підхід дозволяє класифікувати об'єкти на підставі відносин, які вони поділяють з іншими об'єктами, щоб отримати відповіді на такі запитання:

– що руйнує радіація?;

– що діагностується за допомогою радіоактивних ізотопів?

Цей розділ організовано в такий спосіб. У розділі описується загальна методика побудови онтології. Далі розглянуто деякі процедури обробки тексту, що дозволяють перейти від текстового корпусу до вхідних даних, придатних для побудови онтології. Описується метод FCA для побудови ієрархії концептів, наведено метод формування зовнішніх відносин між концептами онтології. І, нарешті, розділ завершується висновками про можливі перспективи використання запропонованого методу.

3.2 Методика побудови онтології

У роботі пропонується використовувати технологію «Methontology» [5], [6], розроблену в лабораторії штучного інтелекту Політехнічного університету Мадрида для побудови прикладних онтологій Процес побудови онтології на основі цієї технології випливає з наступних етапів:

– вилучення термінів: цей етап полягає в тому, щоб виявити та витягти з вхідного корпусу терміни та їх властивості. Для цього використовуються спеціалізовані інформаційні ресурси;

– глосарій медичних термінів та термінів з радіаційної безпеки, синтаксичні шаблони. За допомогою синтаксичного аналізатора вилучаються пари (об'єкт, властивість) та триплети (об'єкт, властивість, об'єкт), що відносяться до загальних смислових блоків;

– побудова основи онтології: на цьому етапі знайдені пари (об'єкт, властивість) використовуються для побудови ієрархії концептів на основі методі FCA;

– вилучення зовнішніх відносин: на цьому етапі застосовується реляційний аналіз понять для вилучення зовнішніх відносин.

На завершення результати виконання двох попередніх етапів об'єднуються для отримання більш повної онтології (рисунок 3.1).

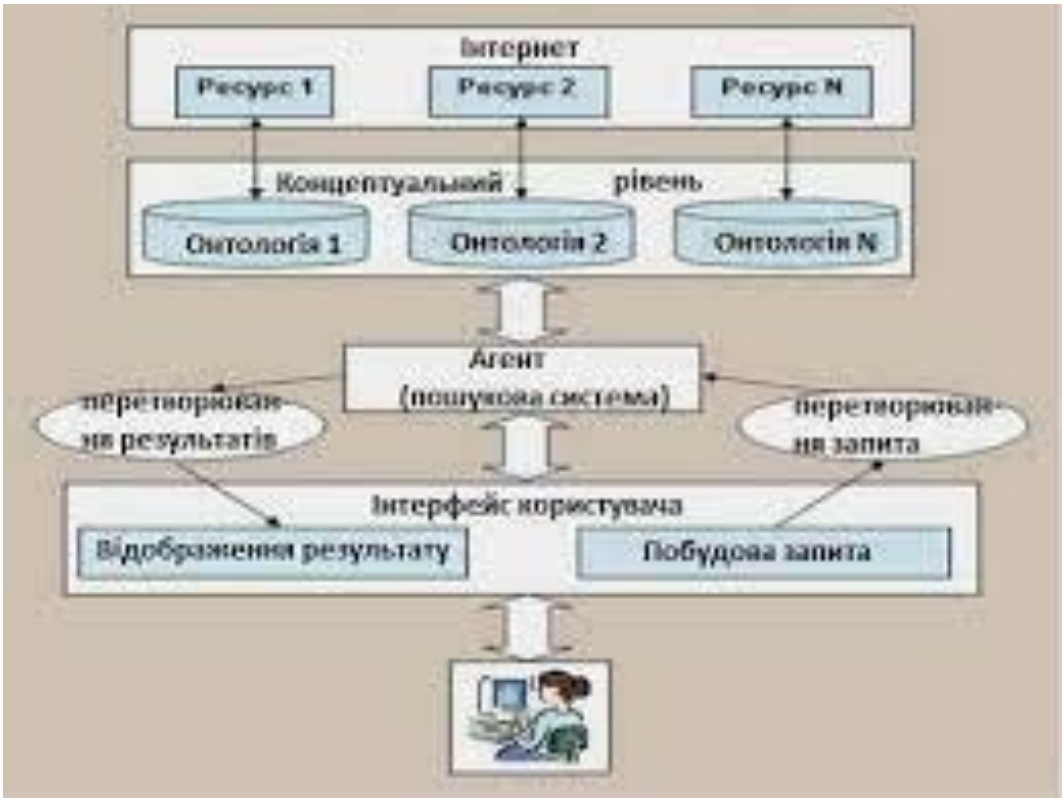

Рисунок 3.1 – Методика побудови онтології

3.3 Методи формування концептів онтології

Незважаючи на різноманітність існуючих підходів, що пропонують автоматичне угруповання понять, це завдання продовжує залишатися в значною мірою відкритою. Одним із раціональних підходів, на думку авторів, є підхід, заснований на гіпотезі Харріса (Harris). Відповідно до нього, вивчення синтаксичних закономірностей у науковому текстовому корпусі, тобто. що складається з текстів, написаних на «Пов'язику», характерному для даного ПЗ, дозволяє виявити специфічні синтаксичні структури, що формуються термінами, які відображають знання досліджуваної ПЗ. Декілька схожих методів заснованих на цій гіпотезі, пропонують групувати терміни в класи на основі їх спільної появи у синтаксичних структурах з однаковими групами дієслів. Використання

одного з цих методів дозволяє об'єднувати назви об'єктів у класи відповідно до груп дієслів, при яких вони виступають як підлягає або доповнення.

Наприклад, об'єкти {щитовидна залоза, шлунок} об'єднані в один клас, так як вони з'являються як підлягає (або суб'єкта) при дієслові {поглинати} і як доповнення (або об'єкту) при дієслові {пошкоджувати}. З іншого боку, даний метод дозволяє також отримувати відносини між різними об'єктами, що з'являються як підлягає або доповнення одного і того ж дієслова.

Наприклад, об'єкт {радіоактивний йод} є доповненням, пов'язаним із об'єктом {Щитоподібна залоза}, який, у свою чергу, підлягає для дієслова {поглинати}. В результаті обробки текстового корпусу потрібно отримати пари (суб'єкт, предикат), (об'єкт, предикат) та триплети (суб'єкт, предикат, об'єкт), що містять ключові слова. Відбір термінів було здійснено за допомогою програми Monoconc. Синтаксичний аналіз виконано з допомогою аналізатора АОТ. Вилучені пари та триплети були запропоновані експерту для відбору серед них найбільш релевантних. Слова у парах та триплетах наведено до нормальної форми.

Представимо кілька прикладів фрагментів тексту в галузі медичної радіології:

1) «Радіонукліди руйнують клітини, ушкоджують органи та тканини і є причиною швидкої загибелі організму, проте вони ж руйнують і злоякісні пухлини». З тексту було вилучено пари (тканини, пошкоджувати), (органи, пошкоджувати), (пухлини, руйнувати), (радіонукліди, руйнувати);

2) «Щитовидна залоза дітей утричі активніше поглинає радіоактивний йод, що потрапив в організм». Витягнуто пари (щитовидна залоза, поглинати) і (радіоактивний йод, поглинати);

3) «Радіоактивний йод використовувався, щоб діагностувати рак щитовидної залози та інші, пов'язані з нею захворювання». Вилучено пари (радіоактивний йод, діагностувати), (рак щитовидної залози, діагностувати).

3.4 Побудова основи онтології за допомогою методу FCA

У роботі було використано ідею Ф. Сіміано, що пропонує будувати онтологію на основі методу FCA [3]. Далі сформульовано перехід від решітки властивостей до онтології та дано визначення кожного концепту за допомогою дескриптивної логіки, що дозволяє отримати відповіді на запитання, сформульовані у розділі 1. За власним зауваженням Ф. Сіміано [7], найбільш вузьким місцем (knowledge acquisition bottleneck) під час побудови онтології є саме моделювання предметної області, тобто. визначення відносин між її концептами. Основою онтології є безліч основних понять чи концептів предметної області, пов'язаних між собою бінарними відносинами. Метод FCA дозволяє візуалізувати залежності між об'єктами за допомогою решіток формальних понять [8]. Його можна використовувати під час аналізу даних виявлення відносин між елементами (у разі концептами) деякої системи (у разі текстового корпусу). Відносини виявляються через атрибути, що описують властивості. У свою чергу, атрибути повинні бути близькими до звичайних категорій людського мислення і повинні допускати наочну та зрозумілу інтерпретацію. Таким чином, FCA можна розглядати як технологію кластеризації концептів, яка дозволяє визначити інтенсіонали окремих блоків даних.

Основними у FCA є поняття формального контексту та формального концепту. Визначення 1 (формальний контекст). Множина (G, M, I) називається формальним контекстом, якщо G та M є множинами, елементи яких пов'язані бінарним ставленням I: $I \subseteq G \times M$. Елементи множини G називаються об'єктами, елементи множини M називаються атрибутами, а елементи множини I визначають інцидентність об'єктів та атрибутів або, іншими словами, належність атрибута об'єкту, визначаючи таким чином формальний контекст.

Визначення 2. Нехай є формальний контекст (G, M, I). Визначимо A' для $A \subseteq G$ наступним чином $A' := \{m \subseteq M \mid \forall g \in A : (g,m) \in I\}$;аналогічно

визначимо B' для B⊆M як B' := {g ⊆G | ∀m∈B: (g, m)∈I}. Простіше кажучи, A' – це безліч атрибутів, загальних для всіх об'єктів безлічі A, і B' – це безліч об'єктів, що мають всі атрибути з B. Оператор «'» називається оператором деривації і застосовується для позначення підмножин множин G і M. На підставі вищесказаного можна визначити поняття формального концепту таким чином Визначення 3 (формальний концепт) Пара (A,B) є формальним концептом (G, M, I) тоді і тільки тоді, коли A G, B M, A'=B і A=B'.

Іншими словами, (A, B) є формальним концептом, якщо безліч всіх атрибутів, що описують об'єкти A, збігається з B і, з іншого боку, всі об'єкти A описуються всіма атрибутами B. У цьому випадку A називають екстентом, а називають формальним інтентом концепту (A, B). Безліч формальних концептів у цьому контексті реалізує відношення підпорядкованості, визначене наступним чином: (,) ( ) ( ) , ABABAABB 1 1 2 2 1 2 2 1 ≤ ⇔ ⊆ ⇔ ⊆ . Відношення підпорядкованості дозволяє організувати формальні концепти на повну матрицю, звану решіткою концептів, яка позначається так: B(G;M; I).

Як приклад у таблиці 1 наведено фрагмент формального контексту (G, M, I) для галузі медичної радіології, де G – це безліч органів людини та його онкологічних патологій, M – безліч їх властивостей. I – це безліч бінарних відносин між M і G, таке, що I(g,m) означає, що g є суб'єктом або об'єктом у текстовому корпусі. Результуюча решітка представлена на рисунку 3.2. Таке розміщення вузлів решітки відбиває успадкування об'єктами властивостей. Властивість позначено над вузлом решіток, а елемент, що володіє даною властивістю – під вузлом.

Усі елементи, розміщені під деяким елементом g, позначеним властивістю m, успадковують цю властивість. Аналогічно, всі послідовники, розміщені під вузлом решітки, поміченим об'єктом g, є більш специфічними поняттями стосовно g.

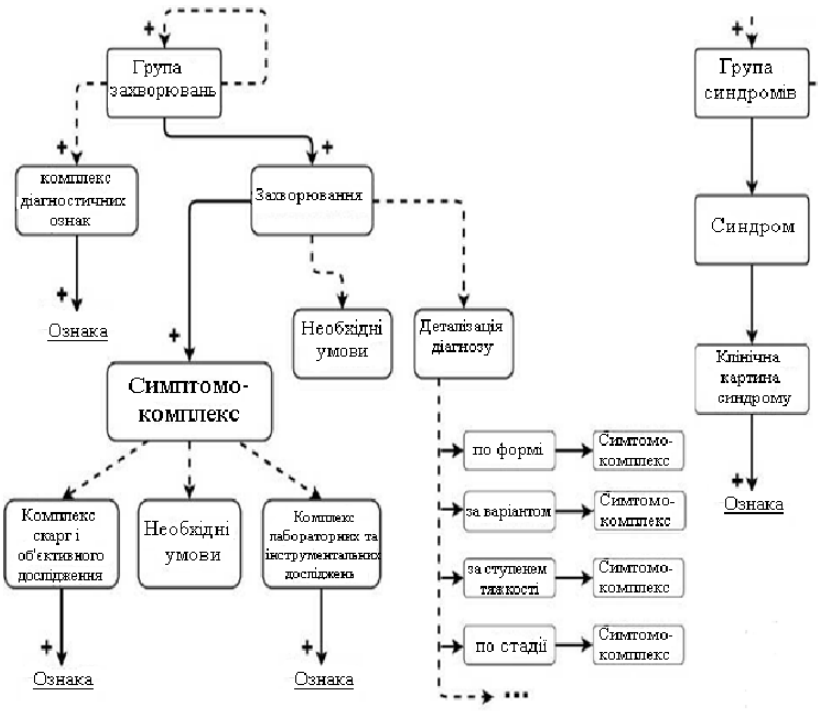

Рисунок 3.2 − Решітка, що відповідають заданому формальному контексту

Таким чином досягається розширення екстенту A концепту (A, B) по відношенню до всіх об'єктів, що з'являються нижче вузла n в решітці, і інтенту по відношенню до всіх атрибутів спадкоємців n. Таке уявлення дозволяє ідентифікувати властивості та екземпляри кожного концепту.

3.5 Перехід від решіток до онтології, розмітка за допомогою експертів

Визначимо функцію перетворення: $\alpha : B(G,M,I) \rightarrow TBox \cup ABox$, де B(G,M,I) – це решітки концептів, отримані на основі FCA. Tbox представляє термінологічний словник ПЗ та Abox – це безліч тверджень про екземпляри, тобто. опис їхніх властивостей. Сукупність Tbox та Abox, термінів та тверджень складає базу знань. [4]. Формально Tbox та Abox визначені таким чином визначення 4,5).

Визначення 4 (основа онтології). Основа онтології є триплетом O:= (C, $\subseteq c$, A) де C – це ансамбль концептів онтології, $\subseteq$ c – відношення

підпорядкованості (is-a) між концептами, що є транзитивним і несиметричним. A – це безліч властивостей чи атрибутів концептів.

Визначення 5 (база знань). База знань для онтології O:= (C, ⊆ c, A) – є структурою: KB:= (I, iC, iA), де I – це безліч екземплярів, iC : C→ 2 I – це функція створення екземплярів, iA – це функція створення атрибутів. Отримана онтологія пропонується експерту, який повинен визначити відповідність загального концепту кожній множині елементів, виходячи з властивостей, якими володіють всі елементи множини. Наприклад, група об'єктів, що володіють властивостями {зруйнований, пошкоджений}, може бути промаркована поняттям орган; а група, що має властивості {руйнується, діагностується} − поняттям пухлина. Таке маркування здійснюється для того, щоб полегшити сприйняття та читання онтології іншими фахівцями.

3.6 Подання концептів за допомогою дескриптивної логіки

Визначення концепту ДЛ виходить шляхом кон'юнкції його атрибутів квантором існування ∃. Результат перетворення решітки понять в онтологію представлено на рисунку 3.3.

Таке подання дозволяє використовувати можливості ДЛ для отримання відповідей на запити наведених нижче типів. Населення онтології (ontology population). Нехай o1 – це об'єкт, що має властивості {a,b}. Примірником якого класу (яких класів) є об'єкт o1? Відповідь може бути наступною: це найвищий клас, який має властивості {a,b}, тобто клас C1 ⊆ ∃ a.T ∩ ∃bT [4]. Наприклад, на запитання «Чи є пухлина щитовидної залози злоякісною пухлиною, якщо цей об'єкт має наступні властивості {досліджуваний, діагностований, виліковуваний}?» Відповідь: відповідний клас C1, що володіє тими ж властивостями – це клас «виліковна злоякісна пухлина» своєю чергою є підкласом класу «злоякісна пухлина». Тому пухлина щитовидної залози є злоякісною пухлиною.

Рисунок 3.3 − Перетворення решіток в онтологію та визначення концептів

Порівняння екземплярів онтології: Нехай є два об'єкти o1 і o2. Чи належить o1 тому ж класу, що і o2? Відповіддю на це питання буде визначення того, який клас C1, якому належить o1, який клас C2, якому належить o2 та наступна перевірка: чи справедливо, що C1 ≡ C2. Наприклад, на запитання, які об'єкти належать до того ж класу, що й «печінка»? Відповідь: це безліч об'єктів того ж класу, що об'єкт «печінка». Об'єкт «печінка» належить класу «орган». До багатьох інших об'єктів цього ж класу належать «шлунок», «щитовидна заліза», «легкі». Питання «Чи належать об'єкти «печінка» та «бронхогенна карцинома» одному й тому класу?». Відповідь: об'єкт «печінка» належить класу орган:= ∃ досліджуваний ∩ ∃ ушкоджуваний; об'єкт «бронхогенна карцинома» належить класу злоякісна пухлина, що виліковується: = ∃досліджуваний ∩ ∃ діагностований ∩ ∃ що виліковується. У свою чергу, орган ∩ лікувана злоякісна пухлина = ⊥. Отже, об'єкти «печінка» і «бронхогенна карцинома»

не належать одному й тому класу. Однак уявлення об'єктів не обмежується лише описом їх властивостей. Вони можуть бути визначені через відносини, які пов'язують їх з іншими об'єктами. Тому виникає потреба розширити основу онтології зовнішніми відносинами між концептами (класами онтології). У цьому випадку пропонується використовувати додатковий метод: реляційний аналіз концептів (RCA) для того, щоб врахувати зовнішні (трансверсальні) відносини в онтології.

3.7 Вилучення зовнішніх відносин

Вилучення зовнішніх (трансверсальних) відносин дозволяє давати концептам більш точні визначення. Завдяки цьому концепт визначається не лише за загальними властивостями окремих екземплярів, а й за відносинами, які його пов'язують з іншими концептами. Осінак-Жилль (Aussenac-Gilles) пропонує використовувати метод навчання онтології для синтаксичних зразків [3]. На основі триплетів (термін1, відношення1, термін2), витягнутих вручну з текстового корпусу, експерти намагаються насамперед узагальнити отношение1, витягуючи триплети типу (термін1, відношення k, термін2). Визначається відношення R найбільш загальне для термінів (термін1, термін2). Потім витягують усі терміни, пов'язані відношенням R у форматі (термін i, R, термін j).

Таким чином, можна замінити одне приватне відношення, що зв'язує екземпляри, іншим більш загальним. Однак [9] не пропонується додатково використовувати деяку базову онтологічну структуру, щоб узагальнити терміни. Інший підхід, запропонований в [10], заснований на вилучення асоціативних правил. З текстового корпусу витягуються пари (термін1, термін2), для них визначається асоціативне правило (термін1, ⇒ термін2) з тим, щоб зберегти лише найчастіше зустрічаються і найбільш достовірні пари. Метод, запропонований у [10], визначає наявність спільних відносин

між концептами онтології, проте самі відносини не визначені. Відомо лише, що концепт C1 пов'язані з концептом C2, але відомо, яким ставленням.

3.7.1 Аналіз формальних концептів

Діяльність пропонується формальний спосіб вилучення зовнішніх відносин, тобто. відносин між концептами. Цей метод дозволяє приписати не який маркер кожному витягнутому відношенню. Він також вимагає наявності базової онтології для узагальнення термінів. Даний метод є розширенням методу FCA, який, як було зазначено вище, дозволяє групувати об'єкти не лише за їхніми загальними властивостями, а й за відносинами, що їх пов'язують. Ідея методу, описаного в [12], полягає в тому, щоб сформулювати формальний контекст для кожного зовнішнього (трансверсального) відношення, витягнутого з текстового корпусу. Елементами кожного контексту є безліч екземплярів, безліч атрибутів, і їхнє бінарне відношення. Таке розширення називається реляційним аналізом концептів (ARC).

Центральним поняттям ARC є реляційна множина контекстів.

Реляційне безліч контекстів є парою (K,R), де – K – це безліч контекстів Ki = (Gi, Mi, Ii), причому кожна безліч екземплярів має один єдиний контекст; – R – це безліч відносин rk⊆ Gi×Gj, де Gi та Gj – це дві множини екземплярів K.

Наведемо приклад для аналізованої предметної галузі – медичної радіології. Нехай є два контексти, сконструйованих на основі методу FCA: K1 = (G1, M1, I1) безліч органів людини та їх пухлинних патологій, і K2 = (G2, M2, I2) безліч радіонуклідів, які застосовуються в медицині. Є також два відношення r1 та r2.

Інтеграція відносин r1 і r2 і решіток властивостей для органів виконана на основі процесу зважування, докладно описаного в [12]. Результуюча решітка демонструє такі зовнішні відносини: діагностується за

допомогою між екземплярами «пухлина щитовидної залози» і «йод 131» і виліковується за допомогою між екземплярами «бронхіальна карцинома» і «кобальт 60». На основі зовнішніх відносин структура онтології може бути розширена та перетворена на більш складну онтологію, описану за допомогою ДЛ.

Визначення 7 (повна онтологія). Повна онтологія представлена наступним набором із п'яти елементів O:=(C, ⊆ C, A, R, σ), де (C, ⊆ C, A) є основою онтології, R – це безліч відносин і σ – це сигнатури відносин.

Визначення 8 (база знань). База знань, що відповідає повній онтології типу O:=(C, ⊆ C, A, R, σ) – це структура CB:=(I; iC; iA; iR), де I – безліч екземплярів, iC : C → 2I – так звана функція встановлення концептів, iA – функція встановлення властивостей та iR: C → 2I+ це функція визначення відносин.

3.7.2 Подання концептів у дескриптивній логіці

Вилучення зовнішніх (трансверсальних) відносин дозволяє поліпшити та збагатити визначення концептів на основі ДЛ. Два концепти «злоякісна пухлина» і «злоякісна пухлина, що виліковується» були перевизначені за допомогою двох відносин, що діагностується за допомогою і виліковується за допомогою:

– злоякісна пухлина: = ∃ досліджуваний ∩ ∃ діагностований ∩ ∃ діагностований_с_допомогою. Радіонукліди короткоживучі;

– злоякісна пухлина, що виліковується: = ∃ досліджуваний ∩ ∃ діагностований ∩ ∃ що виліковується ∩ ∃ що виліковується за допомогою. Радіонукліди довгоживучі. Наведемо ще один приклад запиту: Що діагностується за допомогою короткоживучих радіонуклідів? Відповідь може бути отримана з онтології, збагаченої зовнішніми відносинами між концептами. Примірники концепту «злоякісна пухлина» є підмножиною відносини діагностується за допомогою і є відповіддю на даний запит.

ВИСНОВКИ

Аналіз формальних понять як напрямок у математиці запроваджено професором Рудольфом Вілле і надалі суттєво розвинений ним, його колегами та учнями. Аналіз формальних понять можна, зокрема, розглядати як метод аналізу та графічної інтерпретації знань, представлених за допомогою двовимірних таблиць «об'єкт-властивість». Ідеї цього напряму виявилися настільки привабливими, що аналіз формальних понять висувається деякими його апологетами на роль одного з основних методологічних принципів у побудові теорій (у тому числі математичних), в описі навколишнього світу, а також у викладі математики. З основними поняттями та ідеями цієї теорії можна познайомитися з монографії.

У кваліфікаційній роботі продемонстровано можливість застосування методів формального аналізу понять до даних, отриманих політичними блогами. Як результат було отримано демонстрацію наочного подання процесів, які відбуваються в політиці в ході виборів на пост президента в США. Наочне представлення великих даних, отриманих по блогах, може допомогти експерту зосередити увагу на найцікавіших фактах та подіях.

Ґрунтуючись на отриманих результатах, було також зроблено спроби спрогнозувати подальші шляхи розвитку подій, які підтвердилися надалі. Крім цього, було представлено метод, що дозволяє виявляти зв'язки між політиками, напрямами їх роботи та проводити оцінку таких зв'язків. У роботі описані методи, що дозволяють побудувати онтологію на основі текстового корпусу, який представляє певну предметну область.

Показано, що на основі методу FCA можна коректно групувати об'єкти, розглядаючи їх загальні властивості, та будувати таксономію концептів, пов'язаних з транзитивним ставленням підпорядкованості (isa).

Другий пропонований метод дозволяє отримувати зовнішні відносини між концептами за допомогою реляційного аналізу концептів і, таким чином, розширити можливості використання онтології, наприклад,

отримувати відповіді на більш складні запити. Дескрипторна логіка була обрана мовою опису онтології завдяки відносній простоті побудови правил виведення. Перевагою методу є його універсальність та незалежність від сфери застосування.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Vilem Vychodil: New Algorithm для Computing Formal Concepts. Binghamton, University – SUNY, Binghamton, USA, 2008.

2. Camille Roth, Sergei Obiedkov, Derrick Kourie: Towards Concise Representation for Taxonomies of Epistemic Communities.

3. Baader і B. Sertkaya. Застосовуючи formal concept analysis to description logics. E. P. Eklund, editor, 2D International Conference on Formal Concept Analysis (ICFCA 2004), volume 2961 Lecture Notes in Computer Science, 261-286. Springer-Verlag, 2004.

4. Онтологічний інжиніринг: навчальний посібник / Т. М. Басюк, Д. Г. Досин, В. В. Литвин; Міністерство освіти і науки України, Національний університет «Львівська політехніка». Львів: Видавництво Львівської політехніки, 2017. 224 с.

5. Belohlavek, Radim і De Baets, Bernard і Outrata, Jan і Vychodil, Vilem. Inducing decision trees via concept lattices. *J. International Journal of General Systems*. 2009. Volume 38. 4. P. 455–467.

6. Carpineto, C., Romano, G. (2004b) Випускаючи потенційний Concept Lat-tices for Information Retrieval with CREDO. J. of Universal Computing, 10, 8, 985–1013.

7. Басюк Т. М. Мови опису онтологій: навч. посіб. / Т. М. Басюк, В. В. Литвин; М-во освіти і науки України, Нац. ун-т «Львів. політехніка». Львів : Вид-во Львів. політехніки, 2020. 276 с.

8. Cimiano, P.; Hotho, A. & Staab, S. Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis. *Journal of Artificial Intelligence Research*. 2005. Vol. 24. P. 305–339.

9. Буров Є. В. Концептуальне моделювання інтелектуальних програмних систем : монографія. Львів : Видавництво Львівської політехніки, 2012. 432 с

10. Dau, F., Ducrou, J., Eklund, P. (2008) Concept Similarity and Related Cate-gories в SearchSleuth. P. Eklund та ін. (Eds.): ICCS. LNAI 5113, 255-268. Springer.

11. Vincent Duquenne: Latticial Structures in Data Analysis. Theor. Comput. SCI. 217(2): 407-436 (1999).