

DOI: <https://doi.org/10.15407/kvt218.04.029>

ХАЛА К.О.,

наук. співроб. відд. комплексних досліджень
інформаційних технологій та систем
<https://orcid.org/0000-0002-9477-970X>, e-mail: khala@irtc.org.ua

ГЛАДУН А.Я., канд.тех.наук, доцент,

провідн. наук. співроб. відд. комплексних досліджень
інформаційних технологій та систем
<https://orcid.org/0000-0002-4133-8169>, e-mail: glanat@yahoo.com

Міжнародний науково-навчальний центр інформаційних
технологій та систем НАН України та МОН України,
пр. Академіка Глушкова, 40, м. Київ, 03187, Україна

РОЗШИРЕННЯ МОЖЛИВОСТЕЙ ОНТОЛОГІЧНОГО МОДЕЛЮВАННЯ ПРАВОВИХ ЗНАТЬ З ДОПОМОГОЮ ЕЛЕМЕНТІВ НЕЧІТКОЇ ЛОГІКИ

***Вступ.** Онтологічний аналіз є важливим напрямом в сфері інтелектуальних інформаційних технологій, зокрема для розроблення правових систем, де постійно виникає потреба в ефективному керуванні та обміні юридичними знаннями. Оскільки правові системи відзначаються високою складністю, то використання семантичних технологій у цій сфері дає змогу формалізувати правові поняття. Це сприяє розроблення онтологічних моделей подання правових знань і забезпечує інтеграцію гетерогенних правових інформаційних систем. Водночас важливим завданням є врахування нечіткої логіки для роботи з невизначеністю та неповнотою правової інформації.*

***Мета.** Метою цього дослідження є розроблення моделі правової онтології, здатної ефективно опрацьовувати нечіткі правові терміни та концепти, а також автоматизувати процеси класифікації й аналізування правових документів. Основне завдання полягає у створенні гнучкої системи формалізування юридичних знань, яка враховує специфіку правових актів, сприяє покращенню процесу правозастосування та обґрунтованому прийняттю рішень.*

***Методи.** У дослідженні застосовуються методи семантичного моделювання для створення правових онтологій, а також методи нечіткої логіки для оброблення нечітких та неповних даних. Для реалізування підходу використовуються сучасні інструменти та мови розроблення онтологій, такі як Protege та OWL (Web Ontology Language), а також методи машинного навчання для процесів класифікування та аналізування правових тек-*

стів. Крім того, розглядається використання елементів нечіткої логіки для оцінювання подібності правових документів та подання складних правових концептів.

Результати. Розроблено модель правової онтології, яка дає змогу автоматизувати процеси класифікування та аналізування правових термінів, концептів і відношень між ними. В результаті реалізації запропонованої методології, система здатна отримувати інформацію з різних правових джерел, а також виконувати аналіз правових документів з урахуванням нечітких даних. Під час тестування системи було продемонстровано покращення точності класифікації документів та підвищення ефективності пошуку правових норм у великих масивах слабо структурованих даних.

Висновки. Запропонована модель правової онтології з використанням елементів нечіткої логіки значно покращує процеси подання та оброблення правових знань. Запропонована методологія, яка охоплює аналіз граматики документа та побудову онтологічної моделі документа, і дає змогу точніше порівнювати подібність документів та відмінності між ними. Розроблена методологія на основі семантичного підходу є ефективнішою за метод *k*-середніх для кластеризації ключових фраз. Розглянуто застосування нечітких множин до онтологічної моделі, що дає змогу описувати неточну інформацію та здійснювати нечіткі міркування з певним рівнем повноти. Наразі тривають роботи з розширення української версії юридичної онтології для вирішення практичних завдань правових систем, заснованих на знаннях. Отриманий результат слугуватиме основою для подальшого розвитку інтелектуальних інформаційних систем у правовій сфері.

Ключові слова: нечітке логікове виведення, нечітка логіка, онтологія, аналізування тексту, текстовий контент юридичних знань, подання знань, OWL, Protege.

ВСТУП

Онтологічний аналіз у сфері комп'ютерних наук та інформатики — це відносно нова галузь, що досліджує методи та підходи до створення онтологій, тобто формального представлення сукупності понять і відношень між ними.

Онтологічний аналіз — це методологічний підхід, який використовується для вивчення, структурування та формалізування знань у певній предметній області (ПрО). Він дає змогу визначити ключові концепти, їхні властивості та взаємозв'язки, створюючи основу для розроблення формалізованих моделей знань, зокрема онтологій. Основні аспекти онтологічного аналізу: 1) ідентифікація концептів — виділення ключових понять (об'єктів, явищ, процесів) у досліджуваній області; 2) детермінація відношень — визначення взаємозв'язків між концептами, наприклад, ієрархічних (клас-підклас) або причинно-наслідкових; 3) аналіз властивостей — виявлення характеристик та атрибутів концептів; 4) формалізація — подання знань як вигляді формальних моделей, які можуть бути оброблені комп'ютерними системами. Онтологічний аналіз допомагає перетворити розрізнені знання на чітко структуровану систему, що сприяє їх кращому розумінню, передачі та використанню.

Онтології забезпечують уніфікований словник ПрО та визначають, з різним рівнем формалізації, зміст термінів і зв'язки між ними. Застосування онтологій поширюється на програмні рішення, пов'язані з управлінням знаннями, юриспруденцією, оброблення природної мови, електронною комерцією, інтелектуальною інтеграцією інформації, біоінформатикою, освітніми системами тощо. Онтологічний аналіз є новою сферою досліджень, яка охоплює процеси розроблення онтологій, онтологічний життєвий цикл, а також методи й методології, що використовуються для побудови онтологічних структур [1, 2].

Правова ПрО є однією з найскладніших порівняно з іншими, оскільки вона охоплює знання про фізичний та соціальний світ, а також специфічні правові знання, які фактично формують додатковий шар соціальної реальності.

Актуальним завданням для моделювання та опису правових явищ і процесів є розроблення систем, заснованих на знаннях. Такі системи, що зосереджені на долученні елементів експертної аналітики та інтелектуального пошуку, слугують важливим інструментом у правотворчій та правозастосовній діяльності, а також є основою для розвитку різноманітних семантичних застосунків. Окрім того, подібні системи можуть використовуватися для освітніх цілей у галузі юриспруденції.

Початковим етапом створення системи, що базується на знаннях, є розроблення онтології. Вона слугує джерелом структурованої та документованої сукупності термінів правової ПрО, а також правил, які визначають їхнє використання для формування достовірних правових тверджень. Окрім цього, онтологія забезпечує можливість санкціонованого логікового виведення нових знань, сприяючи автоматизуванню процесів аналізу та прийняття рішень. Онтологія сприяє ефективній взаємодії між користувачем і системою, забезпечуючи обмін інформацією на основі семантичного аналізу, а не лише синтаксичного узгодження єдиної термінологічної бази.

Поняття онтологія охоплює різноманітні рівні формалізації: від простих словників із визначеннями до складних структур. Навіть базовий неструктурований словник може вважатися найпростішою формою онтології. Оскільки в текстах правових актів широко використовуються визначення (наприклад, у преамбулах законів), то створення словника термінів визначеної правової сфери не є складним завданням. Проте, така базова онтологія не забезпечить ефективності у практичному застосуванні.

У роботі [3] представлено перелік характеристик, притаманних лише онтологічним моделям. Розглянемо основні з них:

1. *формальність* — це властивість, яка передбачає точне визначення ієрархічних відношень між класами (is-a), із суворим дотриманням транзитивності цих відношень;

2. *відношення клас-екземпляр* дає змогу містити в онтології не лише назви класів, але й екземпляри (індивіди) на найнижчих рівнях структури;

3. *властивості концептів* дають змогу описувати через концепти інформацію про певні атрибути (слоти), ці властивості особливо корисні, коли їх визначено на верхніх рівнях і успадковано підкласами;

4. *обмеження області значень властивостей* дає змогу визначити додаткові обмеження, які регламентують екземпляром якого класу може бути значення тієї чи іншої властивості;

5. *диз'юнктивні класи та обернені властивості* — диз'юнктивні класи не мають спільних екземплярів, а обернені властивості — це пари властивостей у двох різних класах, що описують взаємозв'язки між їхніми екземплярами;

6. *аксіоми* дають змогу накладати довільні логікові обмеження, розширюючи можливості онтології.

Обґрунтування необхідності розроблення та застосування онтологій викладено в [4], зокрема:

1. забезпечення спільного розуміння структури інформації між людьми або програмними агентами;
2. можливість повторного використання знань у певній ПрО;
3. формулювання явних припущень у ПрО;
4. відокремлення знань ПрО від оперативних знань;
5. забезпечення можливостей об'єктивного аналізу знань у ПрО.

Зазначені аспекти використання онтологічних структур застосовують застосування під час аналізування правових ситуацій, дослідження юридичних процесів та ухвалення рішень у правозастосовній практиці.

Цей напрямок почав активно розвиватися в останні десятиліття, однак одну з методологій вже закріплено в стандартах сімейства IDEF (ICAM/Integrated Definition), розроблених у межах програми комп'ютеризації промисловості США ICAM. Стандарт IDEF5 (Ontology Description Capture) стосується онтологічного опису складних систем. З допомогою методології IDEF5 онтологію системи може бути описано через визначений словник термінів і правил, на базі яких можна формувати достовірні твердження про поточний стан системи. Такі твердження дають змогу робити висновки щодо подальшого розвитку системи та її оптимізації [5].

Серед інших відомих методологій онтологічного аналізу можна виділити такі, як METHONTOLOGY [6], DILIGENT [7], NeOn [8].

ПОСТАНОВКА ЗАВДАННЯ

У рамках онтологічного аналізу в правовій сфері передбачається вирішення низки складних завдань, що стосуються розроблення та інтегрування юридичних онтологій. Основне завдання полягає в створенні базової юридичної онтології, яка буде основою для подальших специфікованих розроблень у різних галузях права, зокрема для моделювання та розроблення онтологій під конкретизовані правові системи. Це охоплює формалізацію правових понять та створення моделей для подання знань у системах, які забезпечують ефективний обмін правовими знаннями.

Однією з основних проблем є інтегрування елементів нечіткої логіки в правову онтологію для забезпечення адекватного опису неточної та неповної інформації, яку часто можна зустріти у правових текстах. Запропонована методологія передбачає використання нечіткої логіки для модифікування онтологічних понять і автоматичного їх оновлення на основі нових знань, що їх отримують в результаті аналізування правових документів.

Метою дослідження є розроблення моделі юридичної онтології, яка враховує всі особливості правових актів, такі як терміни, концепти та відношення, встановлювані між ними. Це передбачає вибір відповідних інструментів для моделювання онтології, зокрема використання онтологічної парадигми, а також розроблення методів для оцінювання подібності правових документів з допомогою елементів нечіткої логіки.

ІНФОРМАЦІЙНИЙ ПОШУК ТА ГРУПУВАННЯ ДОКУМЕНТІВ З ДОПОМОГОЮ МЕТОДІВ АНАЛІЗУВАННЯ

Останнім часом значно зросла кількість наукових публікацій, щодо стосуються розроблення та використання онтологій у різних застосунках.

Розроблення онтологій у правовій сфері має низку специфічних рис [9]:

1. велика кількість загальноприйнятих понять з особливостями їх використання;
2. суттєві відмінності у структурах різних правових галузей;
3. наявність проміжного загальнотеоретичного правового рівня між онтологією верхнього рівня та предметною онтологією;
4. значна кількість теоретичних допущень та абстрактних конструкцій, що залежать від правових концепцій.

Правова галузь характеризується об'ємним документообігом, представленим природномовними текстами, що ускладнює пошук інформації. У сфері інформаційного пошуку постає завдання отримання релевантної інформації з великих джерел, таких як WWW, інтранет і електронних бібліотеки. Підходи до видобутку знань з текстових корпусів, що є частиною корпусної лінгвістики, часто базують на виділенні ключових фраз для індексування та пошуку. Зокрема, наявні методології виділення ключових фраз документа, їхньої частотності та обчислення взаємозв'язків між ними [9].

Оброблення текстів природною мовою за допомогою комп'ютера часто базується на аналізованні процесів мовного оброблення, які відбуваються у людей. Основний підхід полягає у синтаксичному та семантичному аналізованні тексту, де комп'ютерні програми розбивають текст на речення за допомогою граматичних правил, хоча семантичний аналіз залишається складним завданням.

Знання Про зазвичай подають як сукупність об'єктів, понять, сутностей та зв'язків між ними. Обсяги знань у WWW постійно зростають, що підвищує потребу в структурованому поданні — онтології, яка дає змогу описувати та робити ці знання доступними для використання.

Для моделювання метаданих вебресурсів та створення онтології було розроблено Resource Description Framework (RDF). RDF, розроблений Консорціумом WWW, — це модель для подання даних, особливо метаданих, у формі, придатній для оброблення машинами, яка є частиною концепції семантичної павутини. Основою RDF є «трійки»: «суб'єкт» пов'язаний з «об'єктом» через «предикат», який визначає властивість відношень між ними. Завдяки RDF користувачі онтологій можуть легко обмінюватися знаннями, і деякі дослідники використовують RDF як інструмент для підвищення ефективності методів аналізування тексту. Так Палмер запропонував алгоритм, заснований на обчисленні відстані для визначення схожості пар ключових слів в онтології [10]. Крім того, було розроблено алгоритм, який автоматично генерує онтологію, класифікує інформацію з використанням нечітких нейронних мереж, а також метод класифікації документів, який враховує частоту ключових понять для забезпечення точності класифікації. Нечітка логіка надає дослідникам можливість відтворювати експертні правила класифікації.

В обчислювальних системах знання зберігаються та описано з допомогою спеціалізованих мов, для використання яких встановлено певні правила та домовленості. Ці мови спрямовані на створення строго формалізованих структур. Тоді як знання, подані природною мовою, містять лінгвістичні конструкції, які не завжди є чіткими, однозначними або несуперечливими. Тому для таких застосунків доцільно застосовувати методи нечіткої математики та інші методи м'яких обчислень.

Кластеризація — це метод, який дає змогу створювати множини, що є гомогенними всередині груп, але гетерогенними між групами. Математична основа кластеризації полягає в максимізуванні відмінностей між групами та мінімізуванні різниць їх усередині кожної групи. Методи кластеризування знайшли успішне застосування у текстовому обробленні. Так, Ранклер і Бездек [11] виконали кластеризування текстів вебсторінок і послідовностей відвідуваних вебсторінок, а Сюєм [12] продемонстрував застосування методу К-середніх для об'єднання наявних документів у кластери.

Фрази, отримані з текстів, часто використовують для встановлення відношень подібності між текстами документів, що надалі слугує основою для їх кластеризування. Проте статистичний аналіз ключових фраз не завжди є раціональною основою для кластеризації текстів правової Про. З огляду на це, доцільніше застосовувати онтологічну схему заданої Про, ніж підхід до кластеризування лише на основі ключових фраз. Для цього підходу необхідно, щоб експерти побудували онтологічну схему, яка відображатиме структуру знань у цій Про, а також навчали систему на основі заданих шаблонів. У цьому разі випадку природна мова використовується для побудови онтології доступних документів, а нечітка логіка допомагає оцінювати онтологічну подібність між документами, що дає змогу згрупувати їх у відповідні кластери.

Методологія дослідження охоплює описані далі етапи. Спершу моделюється онтологія Про, використовуючи базу знань та інструменти редагування RDF, такі як Protege [13]. Після цього слова та фрази (мовні конструкції, фрагменти тексту та анотації) зіставляють із відповідними концептами онтології Про. Також створюється навчальний набір шаблонів за допомогою інструмента MontyLingua [14]. На основі цих шаблонів у текстових фрагментах документа обчислюють ймовірності концептів. Ймовірності концептів визначають для кожного конкретного документа й надалі використовують для групування шаблонів із застосуванням елементів для нечітких логікових виведень. На завершальному етапі алгоритм ієрархічної кластеризації адаптується, щоб забезпечити інтегрування нечітких множин у процес виведення онтологічних понять.

Введемо уточнення для визначення деяких понять, які можуть мати в цій статті відмінне від загального значення:

- *шаблон* — повторювана конструкція, яка виступає узагальненим рішенням певної проблеми в конкретному контексті.
- *корпус* — сукупність текстів, підібраних для певної дослідницької мети й спеціально оброблених, розмічених, структурованих і поданих в єдиному форматі.
- *блок* — фрагмент документа, виділений інструментом MontyLingua як мінімальна одиниця знань для аналітичного оброблення.

На першому етапі моделювання онтології використовується інструмент для редагування онтологій Protege. Цей редактор дає змогу експертам формувати структуру онтології з допомогою графічного інтерфейсу. Нові та Макгінес вперше запропонували методологію, засновану на експертних знаннях, для розроблення онтологій [9]. Онтологія, створена у Protege, є гнучкою структурою, до якої можна легко додавати нові програмні розширення (плагіни), інтегровані в загальну систему [13]. Таку онтологію може бути автоматично збережено у файли стандартних форматів даних (XML, RDF, OWL) для подальшого використання, аналізування та синтезування знань.

Другим етапом є оброблення текстів природною мовою та «навчання» термінології. Для аналізування поданих у документах знань на предмет певної структури онтології система навчається на основі текстів, доступних у навчальних документах [15]. Кожне речення з навчальних документів обробляється з метою видобування частин мови, фрагментів та анотацій з допомогою MontyLingua. Після цього у таблиці відображення зіставляють видобуті терміни з відповідними концептами онтології. Система обчислює ймовірності концептів, встановлюючи зв'язок між словами та концептами у кожному шаблоні. Умовна ймовірність P (явний концепт | слово S у блоці B корпусів) обчислюється під час навчальної сесії.

Щоб підтримати повноту слів-понять системи, до процесу дослідження залучено повторюваний механізм «переучування». Для долучення нових слів, які входять у поточну базу термінології, коли виявлено новий термін, він спочатку зберігається у базі даних термінології. Після цього системний адміністратор визначає відповідний онтологічний концепт для цього терміну. Це дає змогу системі автоматично повторно обчислити та оновити онтологічну термінологію понять бази знань.

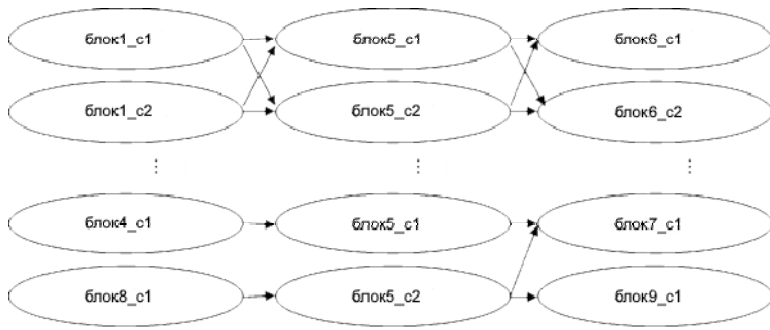
Третім етапом є аналізування термінології, де після оброблення приrodnомовних текстів та «навчання» термінології, всі виведені концепти речень є ймовірнісною величиною, тобто ймовірності концептів кожного блоку обчислені.

Четвертим етапом є видобування знань. Після проведеного аналізу термінології обчислюються потенційні ймовірності концептів для кожного блоку. Насамперед до онтології долучають блоки, що містять концепти як предикати. На рисунку 1 показано, що блок 5 містить два концепти-кандидати, які виконують функцію предикатів в онтології.

Наступним кроком обираються блоки, що містять концепти як суб'єкти онтології, від першої пропозиції до наступної. Таким чином, концепти, пов'язані з блоками 1, 4 і 8, є кандидатами на роль суб'єктів. Той самий підхід застосовується для визначення кандидатів на роль об'єктів (блоки 6, 7 і 9).

За наявності десяти кандидатів на суб'єкт, двох кандидатів на предикат і десяти кандидатів на об'єкт формується $200(10 \times 2 \times 10)$ можливих комбінацій кандидатів. Запити, яких немає в онтології, відсіюються. Підсумкові результати генеруються з використанням ймовірностей, обчислених за наведеним рівнянням:

$$\text{Max}_{\text{для_всіх_значень_блоку}} \times \frac{\text{кандидат(суб'єкт)} + \text{кандидат(предикат)} + \text{кандидат(об'єкт)}}{3} \quad (1)$$



блок i_c_j : поняття j , яке мається на увазі в блоці i

Рис. 1. Фільтрування значень онтології

Табл. 1. Нечіткі правила для виведення подібності документів

№	Якщо два документи, що складаються із пропозицій (отриманих у відповідь на запит з онтології) з ...	то повнота подібності цих двох документів
1	багатьма збігами головних понять та багатьма збігами докладних описів	висока
2	багатьма збігами головних понять та деякими збігами докладних описів	висока
3	багатьма збігами головних понять та кількома збігами докладних описів	середня
4	деякими збігами головних понять та багатьма збігами докладних описів	висока
5	деякими збігами головних понять та деякими збігами докладних описів	середня
6	деякими збігами головних понять і кількома збігами докладних описів	середня
7	кількома збігами головних понять та збігами докладних описів	середня
8	кількома збігами головних понять та деякими збігами докладних описів	низька
9	кількома збігами головних понять та кількома збігами докладних описів	низька

Процес, описаний раніше, використовують для блоків, які містять концепти предикату в документі онтології. Таким чином, документ перетворюють на набір операторів в онтології. Ці запити є індексами документа та основою подібності під час порівняння з іншими документами.

Наступним етапом є *визначення подібності*. Для визначення подібності між документами використовуються елементи нечіткої логіки. Документи розділяють на набір основних концептів та набір деталей. У таблиці 1 подано цю інформацію.

Перед залученням у модель виведення, документи переводяться в онтологічний формат, з долученням головних понять та їх деталей. Головні поняття складаються з верхніх трійок, а деталей з нижніх трійок

$$S = \frac{TT}{ST} \quad (2)$$

де S — міра подібності документа 1 і документа 2; TT — схожі трійки в документі 1 та в документі 2; ST — суми трійок у документі 1 та у документі 2.

Нечіткі логікові подання «багато збігів», «кілька збігів» і «небагато збігів» визначено функціями граничності [16, 17].

Модель нечіткого виведення Мамдані [17] застосовує правила умовного типу «якщо-то» для керування нечіткими вхідними та вихідними даними. Її популярність пояснюється прозорістю формулювання, простою обчислень і зрозумілістю в лінгвістичному поданні, що робить цей підхід ефективним для моделювання. Модель Мамдані використовує «мін-мін-макс» операцію, аналізуючи два базові правила: «прийнятий» та «змінений». Початкова операція Мамдані [17] передбачала підхід із двома правилами, однак у даній статті пропонується розгляд дев'яти правил одночасно (див. таблицю 1). Процедура охоплює такі дії:

1. обчислення подібності документів, які відповідають основним концептам (X_{mc}), та подібності документів, що відповідають детальним описам (X_{dd});

2. визначення оцінок X_{mc} та X_{dd} відповідно до правил таблиці 1 для отримання значень належності до повноти подібності;

3. порівняння значень приналежності та вибір мінімального значення для кожного поняття (висока, середня, або низька подібність) для кожного правила.

4. агрегування значень приналежності для однакових понять в одному наборі.

5. визначення максимального значення належності для кожного набору та обчислення підсумкового результату виведення.

Останнім етапом є *дефазифікування та об'єднання у кластери*. Процедури нечіткого виведення генерують значення належності, які відповідають різним рівням подібності. Оскільки ці значення залишаються нечіткими, вони потребують дефазифікування для отримання чітких показників, що відображають ступінь подібності між документами. Процес дефазифікування відбувається у два етапи: спочатку визначається, яка з оцінок подібності («висока», «середня» або «низька») найкраще передає взаємозв'язок між двома документами; потім здійснюється перетворення оцінок належності на конкретні значення подібності.

Розглянемо це перетворення детальніше для трьох можливих випадків з високою, середньою та низькою подібністю.

Випадок 1 — висока подібність ($U_H > U_L$ і $U_H > U_M$). Якщо значення, обчислене за вищезгаданою процедурою (нечіткого виведення Мамдані) походить від поняття «високої подібності», то для визначення подібності значень документів i і j застосовують таке рівняння:

$$r_{ij}(U_H) = \left\{ \frac{2+U_H}{3} \right\} \quad (3)$$

де U_H — функція належності для «високої подібності»; U_M — функція належності для «середньої подібності»; U_L — функція належності для «низької подібності», з $0 \leq U_H, U_M, U_L \leq 1$.

Випадок 2 — середня подібність ($U_M > U_H$ і $U_M > U_L$). Якщо значення, обчислене за описаною процедурою, походить від «середньої подібності», то для визначення значення подібності використовують рівняння наведене далі. Під час визначення значення подібності для «середньої подібності», взаємозв'язок між «високою подібністю» і «низькою подібністю» впливає на зсув значення дефазифікування. У результаті отримуємо три рівняння для відповідності різним відношенням між «високою подібністю» і «низькою подібністю».

$$r_{ij}(U_M) = \begin{cases} \frac{2+U_M}{6}, U_L > U_H \\ \frac{4-U_M}{6}, U_H > U_L \\ \frac{1}{3}, U_L = U_H \end{cases} \quad (4)$$

Випадок 3 — низька подібність ($U_L > U_H$ і $U_L > U_M$). Якщо значення, обчислене за описаною процедурою, походить від поняття «низької подібності», використовують таке рівняння:

$$r_{ij}(U_L) = \left\{ \frac{1-U_L}{3} \right\} \quad (5)$$

Після того, як всі міри подібності обчислено, генерується матриця подібності. Ієрархічний алгоритм кластеризування використовується для послідовного пошуку різних кластерів відповідно до різного ступеня зв'язку між об'єктами, як подано в матриці:

$$\begin{bmatrix} 1 & \cdots & r_{1j} \\ \vdots & \ddots & \vdots \\ r_{i1} & \cdots & 1 \end{bmatrix} \quad (6)$$

де r_{ij} — подібність документа i та документа j , отже $r_{ij} = r_{ji}$.

Застосування ієрархічного алгоритму кластеризації полягає в наступному:

- 1) віднайденні в матриці максимального значення (r_{ij}) та згрупованні документів i та j у нову групу;
- 2) обчисленні відношення між новими кластерами та іншими документами;
- 3) повторенні процедури, доки не залишиться один остаточний кластер.

НЕЧІТКІ ЗНАЧЕННЯ ТА ОБРОБЛЕННЯ НЕВИЗНАЧЕНИХ ДАНИХ

Концептуальний формалізм стандартної онтології часто є недостатнім для відображення неточної інформації, яка характерної для багатьох прикладних областей. Наприклад, ключові слова, отримані з численних запитів в одній і тій самій області, не можна розглядати як загальну основу, оскільки деякі терміни можуть бути найзначущими порівняно з іншими. Це зумовлює потребу в контекстно-залежній інтерпретації релевантних термінів.

Одним із підходів до оброблення невизначених даних є інтегрування елементів нечіткої логіки в онтологію [18]. Метою використання теорії нечітких множин є опис невизначених понять за допомогою узагальненої множинної концепції. Наприклад, семантичне значення твердження «Правові

відносини є однією з найважливіших форм соціального існування» може мати ступінь істинності, який дорівнює 0,6. Наразі нечіткі множини та онтології спільно використовуються для зменшення невизначеності у різних сферах. Дослідження щодо застосування нечіткої логіки в онтологіях, мають на меті створення розширених онтологій, які ефективніше розв'язують проблеми невизначеності в логікових міркуваннях.

Розглянемо логіковий підхід до роботи з невизначеністю, який є корисним для всіх онтологічних застосувань, де користувач взаємодіє з інформацією, яка містить нечіткі та розмиті описи понять, таких як «дорогий товар» або «висока швидкість». Для розв'язання цих завдань пропонуємо підхід на основі теорії нечітких множин, який передбачає розширення онтологічного редактора для прямого оброблення невизначеностей у процесі побудови онтології з багатшими предметними концептами.

Розглянемо непорожню множину об'єктів U , яку називають універсумом. Нечітку множину визначимо як функцію $[0,1]$ на U , $f:U \rightarrow [0,1]$, де для елемента $x \in U$, $f(x)$ вказує ступінь належності елемента x множині, яка розглядається [19, 20].

Першим завданням є призначення нечітких значень для елементів онтології. Баланс між інтерпретованістю та точністю створює суперечливу мету, в якій на практиці одна з властивостей зазвичай переважає. Залежно від пріоритетів, нечітке моделювання можна поділити на два основні підходи:

1) *лінгвістичне нечітке моделювання* — спрямоване на забезпечення хороших інтерпретованих моделей;

2) *точне нечітке моделювання* — орієнтоване на високу точність нечіткої моделі.

Однак, необхідно об'єктивно підходити до обох ймовірностей, як визначення точного значення або визначення мови.

У разі «визначення точного значення», під час розроблення онтології визначається функція $f:(\text{Концепт} \cup \text{Екземпляр}) \times \text{Властивість} \rightarrow \text{Значення властивості} \times [0,1]$ з урахуванням того що $f(o, p)$ — це значення, яке властивість p концепта або екземпляра o приймає з відповідним ступенем. Наприклад, в гіпотетичній онтології про продукти, $f(\text{товар}, \text{ціна}) = (\text{дорогий}, 0,8)$ означає, що для властивості «ціна», екземпляр «товар», матиме значення «дорогий» зі ступенем 0,8 [21, 22].

Очевидно, що можуть існувати ситуації, в яких немає «значення властивості», необхідного для цієї «властивості». Наприклад, «Товар має побічний ефект зі значенням 0,9» не може бути коректно виражено з допомогою лише формалізму. У цій ситуації, необхідне зіставлення пари (концепт/екземпляр, властивість) на $[0,1]$, тобто $f':(\text{Концепт} \cup \text{Екземпляр}) \times \text{Властивість} \rightarrow [0,1]$, тоді описане матиме вид $f'(\text{товар}, \text{побічний ефект}) = 0,9$.

Для спрощення позначення, визначимо функцію $g: (\text{Концепт} \cup \text{Екземпляр}) \times (\text{Властивість} \cup \text{Значення властивості}) \rightarrow [0,1]$. Тоді, «Товар дорогий за ціною зі значенням 0,8» буде $g(\text{товар}, \text{дорогий}) = 0,8$. Використовуючи таку функцію g , з'являється шанс вибрати цінність значення з нескінченною точністю, тобто точністю кращою за інтерпретованість.

Табл. 2. Зіставлення нечіткого значення з міткою

Мітка	Значення
мало	0.2
достатньо	0.4
помірно	0.6
цілком	0.7
дуже	0.8
повністю	1

У разі «визначення мови» ідея полягає в тому, щоб використовувати мітки із заданої множини для позначення значень. Наприклад, нехай $L = \{\text{мало, достатньо, помірно, цілком, дуже, повністю}\}$ — це множина лінгвістичних оцінок, яка не охоплює всіх можливих значень, але може бути інтуїтивно адаптована. У такому разі значення $g(o, p)$ автоматично присвоюються відповідно до таблиці 2.

Підсумовуючи, ми пропонуємо два способи надання значення в парі (концепт/екземпляр, властивість): через точне значення $v \in [0, 1]$ або через вибір мітки з наперед визначеної множини L . Отож, через функцію g можна встановити нове співвідношення в онтології. Ще одним підходом є визначення об'єкта (концепта або екземпляра) як нечіткого значення, що можна здійснити за допомогою функції $h: \text{Концепт} \cup \text{Екземпляр} \rightarrow [0, 1]$.

Нечітке значення визначається за допомогою однієї з функцій g або h числом у діапазоні $[0, 1]$, це дає змогу застосункам на основі нечіткої онтології використовувати стандартизовані та перевірені інструменти. Наприклад, для комбінування двох або більше нечітких значень може застосовуватися оператор накопичення. Найпростішим прикладом таких операторів є t -норма і t -конорма, що відповідають семантиці операторів OR та AND, серед найвідоміших геделівська норма і конорма, які діють як мінімаксні оператори.

Виходячи з наведеного, обчислимо істинність виразу «Товар дорогий та має побічний ефект». Якщо відомо, що $f(\text{товар, дорогий}) = 0,8$ і $f(\text{товар, побічний ефект}) = 0,9$, то $[f(\text{товар, дорогий}) \wedge f(\text{товар, побічний ефект})] = \min\{0,8, 0,9\} = 0,8$ [19, 20].

Таким чином, нечітку онтологію можна визначити як онтологію, доповнену нечіткими значеннями, які задаються двома функціями:

- $g: (\text{Концепт} \cup \text{Екземпляр}) \times (\text{Властивість} \cup \text{Значення властивості}) \rightarrow [0, 1]$ та
- $h: \text{Концепт} \cup \text{Екземпляр} \rightarrow [0, 1]$.

Після розроблення нечіткої онтології є хибним припускати, що вона буде ефективною, і будь-яке нечітке значення чітко визначено, а онтологія придатна для застосування у будь-якій ситуації. Важливим кроком є розроблення механізму зміни нечітких значень для їх найкращого налаштування на конкретне середовище, ситуацію або для їх узагальнення.

Припустимо, що поточним нечітким значення є f . Як результат запиту його має бути оновлено до f_{new} . Найпростіший спосіб — задати $f: = f_{new}$. Закономірно припустити, що після того як якийсь запит нечіткої властивості досягне стабільного значення буде марним змінювати його на f_{new} , що веде до втрати всієї історії надбаних знань. Рішення цієї проблеми може полягати у зменшенні важливості f_{new} за будь якого змінення:

$$f := f + \frac{f_{new} - f}{Q+1} \quad (7)$$

де Q — кількість оновлень, виконаних для заданого значення. Зрозуміло, що значення Q має зберігатися в онтології для будь-якого визначення нечіткого значення [22].

Тепер завдання полягає в тому, аби обчислити нове нечітке значення f_{new} . Малоімовірно знайти у документі точне визначення нечіткого значення, але можна знайти мовний класифікатор. Наприклад, ми не знайшли «Відносини – важлива форма буття зі значенням 0,8». Але, сенс цієї тези полягає в тому, що «Відносини — дуже важлива форма буття». Для цього розглянемо підхід, базований на основі модифікаторів концептів.

Модифікатори концептів мають вплив на зміну нечіткого «значення властивості». Визначимо низку таких лінгвістичних обмежень таких як «дуже», «більш-менш», «ледве», тоді модифікатор концепта — система одного або більше обмежень, таких як «дуже небагато» або «дуже-дуже небагато». Будь-якому (лінгвістичному) модифікатору концепта необхідно зіставити (числовий) член модифікатора.

Член модифікатора — це значення $\beta > 0$, яке використовують як показник для зміни значення функції приналежності f як f^β .

Відповідно до їх впливу на нечітке значення, обмеження можна розділити на дві групи: концентрування та розтягування. Вплив модифікатора концепта полягає у зниженні ступеня належності значення. Тобто, у такому разі він має бути $\beta > 1$. Наприклад, обмеженню «дуже» присвоєно $\beta = 2$. Тоді, якщо ми знаємо, що $g(\text{відносини, важлива форма}) = 0,8$, то «відносини дуже важлива форма» зі значенням $0,8^2 = 0,64$. Навпаки, обмеження типу розширення може підвищувати значення належності, тобто $\beta \in (0,1)$. Наприклад, якщо «досить» присвоєно $\beta = 0,4$ і $g(\text{відносини, важлива форма}) = 0,8$, то «відносини досить важлива форма» зі значенням $0,8^{0,4} = 0,91$, інтуїтивно зрозуміло, що якщо щось є «важливим», то воно певною мірою є «досить важливим». Зауважимо, що такий підхід відрізняється від підходу Заде, де «досить» та інші модифікатори обробляються складнішим чином. Перевага методу полягає в тому, що він забезпечує універсальний і простий спосіб керування модифікаторами концептів, навіть якщо для поглиблення таким способом семантики необхідно звертатися до ланцюгів модифікатора [19, 21, 23].

Отож, модифікатори концептів використовують для визначення нового нечіткого члена з урахуванням вже наявних. Наприклад, якщо відомо нечітке значення властивості «важливий», можна вивести нечітке значення властивості «дуже важливий» простим зведенням в ступінь 2 значення «важливий». Водночас можна говорити про протилежну ситуацію, якщо з онтології відомо значення властивість «важливий» і, як результат запиту, що якийсь об'єкт є «дуже важливий». Тобто з «дуже важливий» ми маємо вивести нову властивість «важливий» для цього об'єкта. Якщо в онтології $g(o, \text{важливий}) = 0,7$, необхідно збільшити це значення, наприклад $g(o, \text{важливий}) = 0,7^{0,5}$. Ефект «дуже» підвищує значення якості, не зменшуючи його величину. Таким чином, можна сказати, що у зазви-

чай виконується виведення: «Важливий \rightarrow Дуже важливий», тоді як у цій ситуації: «Важливий та Дуже важливий \rightarrow Важливий».

Цей випадок поширюється й на інші модифікатори концептів. Таким чином, те, що вважають модифікатором концентрування, стає розтягуванням і навпаки.

Необхідно вирішити таку проблему: визначити який модифікатор розглядається (і яке його значення) і визначити метод обчислення значення β для ланцюгів модифікаторів концептів.

Виберемо множину обмежень $H = \{\text{дуже, набагато більше, більше, більше чи менше, середньо, ледве}\}$. Це лише один із можливих варіантів. Множину H може бути змінено залежно від потреб, за умови, що вона задовольняє такі дві властивості:

- H впорядкована множина, тобто «дуже $< \dots <$ злегка» і значення β фіксовано лише для найменших і найбільших елементів, як 0,5 і 2 відповідно.
- дві підмножини обмежень концентрування {більше чи менше, середньо, ледве} або обмежень розтягування {дуже, набагато більше, більше} мають однакову потужність (однакову кількість елементів).

Ці дві умови необхідні для обчислення належності модифікатора до послідовності обмежень. Єдина різниця полягає у заданні інверсії обмеження, пов'язаної з використанням модифікаторів концептів [21].

Відкритим питанням лишається те, що множина H модифікаторів концептів, що не вичерпує усіх нюансів природної мови. Багато елементів може бути додано в H , а це вимагатиме нових алгоритмів для їхнього оброблення; оскільки не всі наявні модифікатори концептів може бути повністю задано або точно поділено на дві підгрупи однакової потужності.

Розглянемо два можливі приклади використання нечітких онтологій. Перший приклад базується на нечіткому значенні пари, яка пов'язує випадок та властивість, а другий — на використанні концептів з нечітким значенням для вирішення проблеми навантаження.

Розширення запитів. Під час виконання запиту до документа часто застосовують метод розширення набору значень концептів, що вже містяться в запиті, з додаванням інших значень, які можна отримати з онтології. Очевидно, що до запиту можна долучати значення для поточного концепту, а також для його батьківських та нащадкових концептів, після чого проводиться пошук у документі. Використання нечітких онтологій дає змогу гнучкіше розширювати запити.

Для прийняття рішення про те, які значення додавати до розширеного запиту, необхідно визначити, як лінгвістичні мітки співвідносяться з числовими значеннями. Це вирішується так само, як і в попередньому описі, проте мітки, що належать множині L , використовуються в запитах і перетворюються на числові значення відповідно до таблиці 1.

Якщо c — концепт (поняття), p — власність і l — мітка, тоді $\mu(c, p, l)$ — значення даної мітки l для власності p і концепта c . Тепер розглянемо всі випадки i концепту c , які буде долучено в розширений запит, якщо i тільки якщо:

$$|\mu(c, p, l) - g(i, p)| \leq e \quad (8)$$

де e з $[0.1]$ — рівень толерантності.

Очевидно, що кількість екземплярів, які буде долучено до розширеного запиту, залежить від значення e , причому чим більший параметр e , тим більше екземплярів буде додано. Для граничних випадків $e = 0$ буде долучено лише ті екземпляри, які точно відповідають запиту і входять до нього, а за $e = 1$ до запиту буде долучено всі екземпляри.

Це найпростіший випадок, коли в запиті є лише одна властивість. Якщо ж запит містить дві або більше властивостей, необхідно узагальнити рівняння (8). Припустимо, що в запиті є n властивостей, які відносяться до тих самих концептів, що й екземпляр c , тоді

$$\frac{\sum_{j=1}^n |\mu(c, p, l) - g(i, p)|}{n} \leq e \quad (9)$$

Це означає, що середнє значення відстані між властивостями в запиті та властивостями екземпляра має бути меншим або рівним заданому рівню допустимості e .

Перевантаження поняття. Як зазначалося раніше, можливість використання нечітких значень, пов'язаних з концептом, допомагає вирішити проблему перевантаження поняття в онтології.

Припустимо, що поняття c зустрічається в різних фрагментах онтології, і мета полягає в тому, щоб вказати, який з цих фрагментів є найзначущим у контексті певного домену. На першому етапі кожному концепту, який присутньому у кількох місцях, призначається рівне нечітке значення, таке, що їх сума дорівнює 1.

Для будь-якого концепту c_i множину його локальних умов L_i , що охоплює батьківські та нащадкові елементи, обчислюють в межах онтології. Потім під час аналізу в документах шукають всі елементи $l \in L_i$ з призначеними вагам w_i^j , що відображають їх значимість у документі. Припустимо, що в певному документі, l_j входжень для елемента l . Тоді для будь-якого концепту c_i і для будь-якого документа d обчислюють такі функції:

$$\mu_d(c_i) = \sum_j l_j w_i^j \quad (10)$$

Сума за всіма n документами μ_d позначена як μ :

$$\mu(c_i) = \sum_{d=1}^n \mu_d$$

Потім нове значення приналежності для концепту c_i документа d обчислюють за формулою:

$$h_{new}(c_i) = \frac{\mu_d(c_i)}{\mu(c_i)} \quad (11)$$

Ці значення використовують для оновлення функції $h(c_i)$ згідно з рівнянням (7), і їх застосовують з урахуванням релевантності, так що значення з найзначущих документів обробляються першочергово та мають більший вплив, ніж значення з менш значущих.

РОЗРОБЛЕННЯ МОДЕЛІ ОНТОЛОГІЇ ДЛЯ СФЕРИ ПРАВА

Онтологічна парадигма підтримує розроблення систем онтологічних моделей та забезпечує інтеграцію з лінгвістичними компонентами під час створення лінгвістичних застосунків. Одним із ключових завдань є інтегрування онтологічних та лінгвістичних моделей для репрезентування знань. Онтологічна модель визначає структуру знань, тоді як лінгвістична модель фокусується на функціонуванні мовних одиниць у текстах. Іншими словами, лінгвістична модель описує лексичні, семантичні та синтаксичні структури концептуальних одиниць, а також онтологічні елементи. Це особливо важливо для стратегії висхідного моделювання, однак і для низхідного підходу кінцева онтологія повинна містити терміни, які використовують у правових актах. Тому правові онтології можна віднести радше до лінгвістичних ніж до формальних онтологій.

Що стосується правової інформації, зокрема під час розроблення галузевих онтологій, основний текстовий масив збігається із законодавчою базою відповідної галузі, що передбачає наявність чітко визначеного переліку правових документів, доступних для громадян. Незважаючи на різноманіття правових документів та можливі недоліки юридичної техніки, тексти правових документів повинні мати характеристики, які полегшують формалізацію ПрО.

До таких властивостей можна віднести:

- нейтральність (відсутність експресивності);
- безособовий стиль;
- зв'язність;
- послідовність;
- точність висловлення;
- наявність великої кількості стандартних виразів;
- виражену структурованість тексту.

У багатьох випадках для вирішення прикладних завдань онтологічного аналізу в правовій сфері доцільно використовувати загальну базову юридичну онтологію, яку розробляють без прив'язки до конкретного завдання і яку може бути легко розширено. На сьогодні відсутня єдина загальноприйнята правова онтологія, хоча досвід закордонних дослідників свідчить, що створення такої онтології є не лише можливим, а й необхідним. Така онтологія може слугувати основою для розроблення спеціалізованих онтологій для вирішення завдань у різних галузях права.

Не видається доцільним виконувати всю роботу з формалізування правової сфери без використання результатів інших онтологічних проєктів. Правова онтологія має бути частиною загальної архітектури систем правових знань, які забезпечують обмінювання інформацією між різними системами правових знань (це формат обмінювання правовими знаннями — LKIF) [24]. LKIF виконує дві основні функції: забезпечує переклад між правовими базами знань, поданих в різних форматах і формах; та подає формалізм знань, що є частиною загальної архітектури розроблення правових систем знань. Проте використання LKIF веде до класичного компромісу між гнучкістю та виразністю. Крім того, формат LKIF має відповідати сучасним стандартам Semantic Web для інтегрування правової інформації

через веб, комбінуючи OWL DL і SWRL, що дає змогу створити гібридне рішення для вирішення завдань.

Таким чином, актуальним завданням є моделювання юридичної онтології верхнього рівня на основі базової юридичної онтології LKIF для системи права з використанням україномовних концепцій.

Реалізування онтологічної моделі є неможливим без вибору відповідних інструментів моделювання та визначення стратегії розроблення. Вибір інструменту моделювання не викликає значних труднощів: як основу вибрано «батьківську» онтологію LKIF [24]. Онтологія LKIF не лише забезпечує теоретичне розуміння правової галузі, але й надає практичні інструменти для полегшення процесів набуття знань, обмінювання ними та їх подання, зокрема, для формалізування частин наявного законодавства.

Онтологія LKIF містить «фундаментальні поняття права». Вибір застосовного словника залежить від потреб потенційних користувачів. Визначено три основні категорії користувачів: громадяни (непрофесіонали), юристи та правознавці. Хоча юристи використовують правовий словник точніше і ретельніше, ніж непрофесіонали, більшість термінів мають загальне тлумачення, і ці терміни можна вважати більш-менш схожими.

Водночас, багато основних термінів мають специфічне юридично-технічне значення, наприклад, «відповідальність» чи «правовий факт».

Як стратегію моделювання обрано систему METHONTOLOGY [4]. METHONTOLOGY, розроблена Гомез-Перез [1] та колегами, є підходом до створення та супроводу онтологій. В рамках цієї методології реалізовано принципи Тома Грубера [3], а також створено програмний комплекс для специфікації онтологій — ODE (Ontology Design Environment).

Цей підхід визначає основні етапи «життєвого циклу» створення онтології: керування проєктом, розроблення онтології та підтримка її розроблення.

Процеси керування проєктом охоплюють планування, контроль і забезпечення якості. Планування передбачає визначення завдань, організування їх виконання, а також оцінювання необхідного часу та ресурсів для реалізації. Контроль гарантує, що завдання виконуються відповідно до плану. Забезпечення якості необхідне для підтвердження, що компоненти та кінцевий продукт відповідають заданим вимогам.

Процес розроблення онтології охоплює кілька етапів: специфікацію, концептуалізацію, формалізацію та реалізацію. На рисунку 2 подано фрагмент таксономії понять онтології для об'єкта права.

Згідно з обговорюваною методологією, спочатку створюється глосарій термінів, який мстить всі концепти, їхні екземпляри, атрибути, дії тощо, важливі для ПрО, а також їх опис у природномовній формі. Приклад цього глосарію подано в таблиці 3.

Після того, як глосарій термінів набуває суттєвого обсягу, будується дерево класифікації концептів. Це дає змогу ідентифікувати основні таксономії ПрО, і кожна таксономія стає частиною онтології відповідно до запропонованої методології.

Після побудови вищеописаних подань для кожного дерева класифікації концептів створюються такі таблиці:

1) *Словник концептів* (див. таблицю 4), що містить всі концепти ПрО, їхні екземпляри, атрибути, відношення, які слідують з концептів, а також (за необхідності) синоніми та акроніми.

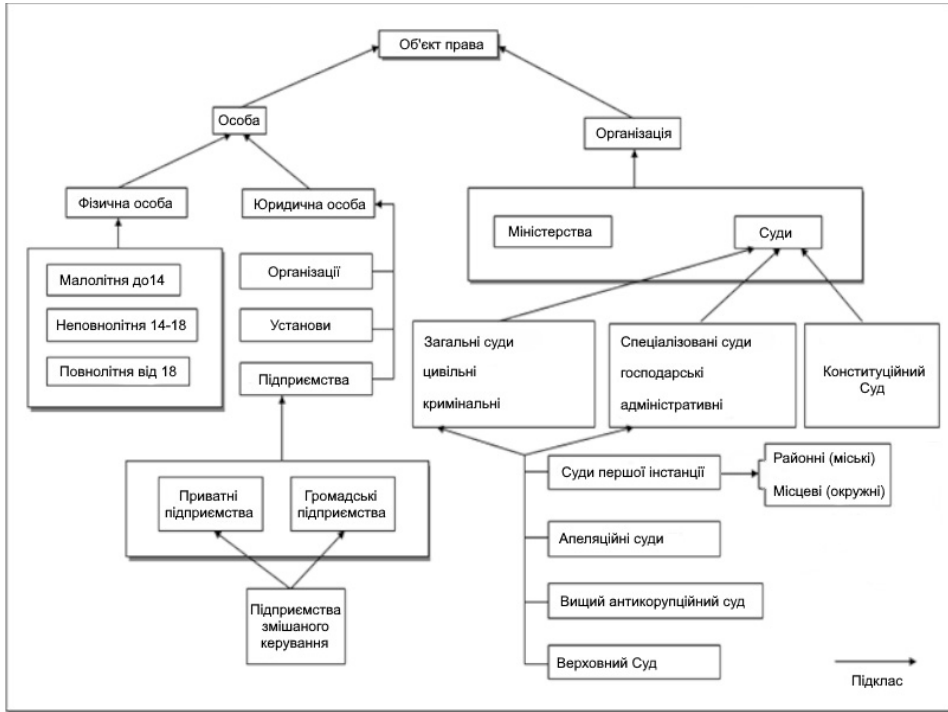


Рис. 2. Приклад фрагменту таксономії понять онтології для об'єкта права

Табл. 3. Уривок з глосарію термінів онтології для об'єкта права

Ім'я	Синоніми	Акроніми	Опис	Тип
вік повноліття	-	-	Вік повноліття з 18	постійна
суд	Судовий трибунал	-	Хоча «суд» може бути зрозумілий як фізичне місце або як суддя, вважаємо, що суд — судовий суд	концепт
дата народження	-	-	Дата, коли людина народилася	ставлення
відповідач (людина, судовий процес)	-	-	Судовий процес відповідача	атрибут екземпляра

Табл. 4. Уривок зі словника концептів онтології правового об'єкта

Ім'я концепту	Примірники	Атрибути класу	Атрибути екземпляра	Відношення
Суд	Конституційний Суд, Верховний Суд, Апеляційний Суд	-	кількість членів, місце, територіальна юрисдикція	розглядає
Підприємства	-	вид контролю	ім'я (назва)	-
Судовий процес	-	-	-	має відповідача, має позивача, розглянуто
Особа	-	-	-	відповідач, позивач
Фізична особа	-	-	вік, дата народження, дата смерті, перше прізвище, ім'я, національність, друге прізвище	мати має матір батько має батька



Рис. 3. Фрагмент діаграми спеціальних бінарних відношень правового суб'єкта онтології

2) Таблиця бінарних відношень для кожного відношення, вихідний концепт якого входить до класифікаційного дерева. Уривок зі слота бінарних відношень подано в таблиці 5. Для кожного відношення фіксують його ім'я, імена концепту-джерела та цільового концепту, інверсне відношення тощо. Приклад фрагмента діаграми спеціальних бінарних відношень наведено на рисунку 3.

3) Таблиця атрибутів екземпляра для кожного екземпляра зі словника концептів (див. таблицю 6). Основні характеристики містять: ім'я атрибута, тип значення, одиницю виміру, точність, діапазон зміни, значення за промовчуванням, атрибути, які можуть бути виведені з використанням даного атрибута, а також формулу чи правило для виведення атрибута.

4) Таблиця атрибутів класу для кожного класу зі словника концептів із подібними характеристиками (див. таблицю 7).

5) Таблиця логікових аксіом, у якій фіксують визначення концептів через логікові вирази, що є завжди дійсними (див. таблицю 8). Кожна аксіома містить її ім'я, опис, концепт, до якого вона стосується, атрибути, що їх використовують в аксіомі, логіковий вираз тощо.

Табл. 5. Уривок зі столу бінарного відношення онтології правового суб'єкта

Ім'я відношення	Початковий концепт	Початкова кількість елементів	Цільовий концепт	Зворотнє відношення
відповідач	особа	N	судовий процес	має відповідача
позивач	особа	N	судовий процес	має позивача
розглядає	суд	N	судовий процес	розглянуто
має відповідача	судовий процес	N	особа	відповідач
має позивача	судовий процес	N	особа	позивач
розглянуто	судовий процес	N	суд	розглядає

Табл. 6. Уривок з таблиці атрибутів екземпляра онтології правового об'єкта

Ім'я атрибута екземпляра	Ім'я концепту	Тип значення	Діапазон значення	Потужність
кількість членів	суд	ціле	1..	(1,1)
місце	суд	рядок	-	(1,1)
територіальна юрисдикція	суд	рядок	-	(1,1)

Табл. 7. Уривок з таблиці атрибутів класу онтології правового об'єкта

Ім'я атрибута класу	Певний концепт	Тип значення	Потужність	Значення
вид контролю	приватне підприємство	[приватне, акціонерне]	(1, 2)	приватне
вид контролю	акціонерне підприємство	[приватне, акціонерне]	(1, 2)	акціонерне

Табл. 8. Уривок з столу логікових аксіом онтології судового процесу

Ім'я аксіоми	Несумісність позивача відповідача
Опис	Людина не може бути позивачем і відповідачем у тому самому судовому процесі
Вираз	не (існує (X, Y) (особа (X) і судовий процес (Y) і [позивач (X, Y) і [відповідач (X, Y))])
Концепти посилання	особа судовий процес
Посилання	позивач відповідач
Змінні	X Y

6) *Таблиця констант* (див. таблицю 9), де для кожної константи зазначають її ім'я, опис, тип значення, значення, одиницю виміру, атрибуту, що можуть бути виведені з використанням цієї константи, тощо.

7) *Таблиця формул* для кожної формули з таблиці атрибутів екземпляра (див. таблицю 10). Кожна таблиця цього типу містить ім'я формули, атрибут, що виводиться за допомогою цієї формули, її опис, точність, обмеження на використання тощо.

8) *Дерева класифікування атрибутів*, що графічно зображають відповідні атрибути та константи, які використовуються для виведення значення кореневого атрибута та формули, яку застосовують для цього. Приклад фрагмента дерев подано на рисунку 4. Ці дерева використовують для перевірки повноти, аби усі подані в формулі атрибути мали описи.

Таблиця 9. Уривок з таблиці констант онтології правового об'єкта

Ім'я	Тип значення	Значення	Одиниця виміру
вік повноліття	кількісне числове	18	роки

Таблиця 10. Значення для формули «Ліберальної поведінки»

Ім'я формули	Ліберальна поведінка
Атрибут, що виводиться	Варіанти поведінки
Формула	Варіанти поведінки = Дозволено \cup Прямо не заборонено
Опис	"Дозволено все те, що прямо не заборонено", суб'єкту надається можливість вибирати будь-який варіант поведінки, крім тих, що прямо і строго сформульовані як заборони
Основний атрибут екземпляра	Дозволено Прямо не заборонено
Основний атрибут клас	-
Константи	-
Точність	-
Обмеження	Варіанти поведінки, які суворо заборонено законом



Рис. 4. Фрагмент дерева класифікування атрибутів за формулою

Таблиця 11. Витяг з таблиці примірників онтології правового об'єкта

Ім'я екземпляра	Ім'я концепту	Атрибут	Значення
Апеляційний суд	суд	місце	Київ
		територіальна юрисдикція	Україна
Верховний Суд	суд	територіальна юрисдикція	Україна
Конституційний Суд	суд	кількість членів	12
		територіальна юрисдикція	Україна

9) *Таблиця екземплярів* для кожного входу в словник концептів (див. таблицю 11). Вона містить ім'я відповідного концепту до якого стосується, ім'я екземпляра, його атрибуту та значення.

Поєднання онтологічного моделювання та нечіткої логіки дає змогу створювати інтелектуальні правові системи, здатні працювати з реальними викликами права: неоднозначністю, складністю контексту та необхідністю гнучкості. Такий підхід може бути корисним у судовій аналітиці, автоматизуванні юридичних консультацій або підтримці прийняття рішень.

ВИСНОВКИ

Онтологічне моделювання правових знань із використанням елементів нечіткої логіки дає змогу створювати адаптивніші та інтелектуальніші правові інформаційні системи. Розроблення базової онтології для інтелектуальної системи права є важливим завданням керування знаннями, правовою інформацією та її автоматизованого аналізу. Поставлена проблема потребує системного підходу, що має забезпечити правильне інтерпретування та організування знань у цій складній області.

Розроблення базової моделі правової онтології є початковим етапом для розвитку подальших досліджень у цій сфері і слугує основою для вдосконалення методів оброблення юридичних текстів. Автори удосконалили метод кількісного оцінювання точності моделей, що дає змогу ефективніше відображати складні правові відносини.

Традиційно для аналізу правових текстів застосовують методи оброблення документів, базовані зокрема на ключових фразах. Однак це підходить лише для простих випадків, тоді як у сфері права кожна фраза може мати кілька значень, і багато різних фраз можуть мати спільні значення. Запропонований метод, що враховує граматичну структуру речень і побудову онтологічної моделі документів, є точнішим і надає можливість здійснювати глибший аналіз. Порівняння документів за рівнем їх семантичної подібності та відмінностей між ними є надійнішим і ефективнішим, ніж використання традиційних методів класифікування на основі ключових фраз.

Крім того, представлений підхід уведення нечіткої логіки безпосередньо в онтологічну модель дає змогу коректніше відображати неточну та нечітку інформацію, що є важливим для правових систем, де багато понять мають невизначені або варіативні значення. Це дає змогу проводити нечіткі міркування, що особливо важливо для автоматизованих систем аналізу правових текстів.

Зокрема, запропоновано метод автоматичного оновлення понять через використання модифікаторів, що забезпечує постійне оновлення і уточнення знань в системі. Це є важливим для підтримки актуальності правової інформації в умовах швидких змін в законодавстві.

Наразі продовжуються роботи з розширення отриманої україномовної версії правової онтології та її використання для розв'язання практичних завдань правових систем, заснованих на знаннях. Перспективи подальших досліджень вказують на можливість інтегрування отриманих результатів у практичні правові системи для автоматизування юридичних процесів та забезпечення ефективнішої роботи правників і суддів.

REFERENCES

1. A. Gomez-Perez, M. Fernandez-Lopez, O. Corcho, *Ontological Engineering: With Examples from the Areas of Knowledge Management, E-commerce and the Semantic Web*. Springer, 2004.
2. A. Denicola, M. Missikoff, R. Navigli, A software engineering approach to ontology building. *Information Systems*, Vol. 34, no. 2, 2009, pp. 258-275 doi: 10.1016/j.is.2008.07.002. <https://doi.org/10.1016/j.is.2008.07.002>
3. Gruber T.R. A translation approach to portable ontology specifications, *Knowledge Acquisition*. 1993. <https://doi.org/10.1006/knac.1993.1008>
4. N. Noy, D. McGuinness, *Ontology Development 101. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05*, 2001, p. 217-228.
5. IDEF5 Method Report. *Knowledge Based Systems, Inc. for Information Integration for Concurrent Engineering*. 1994.
6. M.Fernandez-Lopez, Overview and Analysis of methodologies for building ontologies. *Knowledge Engineering Review (KER)*. Vol. 17(2), 2002. <https://doi.org/10.1017/S026988902000462>
7. DILIGENT Ontology Engineering. URL: <http://www.aifb.uni-karlsruhe.de/WBS/cte/ontologyengineering/diligent.htm>.
8. Neon Project. URL: <http://www.neon-project.org>.
9. N. Noy, D. McGuinness, *Ontology Development 101: A Guide to Creating Your First Ontology*. URL: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf
10. Z. Wu та M. Palmer. «Verb semantics and lexical selection. — в Proc. 12nd Anna. Meeting Assoc. Comput. Linguist. Las Cruces. NM. Jun. 27–30. 1994, p. 133–138. <https://doi.org/10.3115/981732.981751>
11. J.C. Bezdek, R.J. Hathaway, Some Notes on Alternating Optimization for Clustering and Visualizing Fuzzy c-Means and k-Means. *Proceedings of the IEEE*, 90(11), 2002, p. 1964–1972.
12. R.Xu, D. Wunsch, "Survey of Clustering Algorithms." *IEEE Transactions on Neural Networks*, 16(3), 2005, p. 645–678. <https://doi.org/10.1109/TNN.2005.845141>
13. M. A. Musen, The Protégé Project: A Look Back and a Look Forward. *AI Matters*, 1(4), 2015, p. 4–12. <https://doi.org/10.1145/2757001.2757003>
14. H. Liu, *MontyLingua: An End-to-End Natural Language Processor with Common Sense*. MIT Media Lab., 2004.
15. A. Gladun, J.Rogushina, R. Martínez-Béjar, UKR at EmoSPeech-IberLEF2024: Using Fine-tuning with BERT and MFCC Features for Emotion Detection, In: *IberLEF, Iberian Languages Evaluation Forum*, September 2024, Valladolid, Spain, CEUR, Vol. 3756, 2024, 6 p., URL: http://ceur-ws.org/Vol-3756/EmoSPeech2024_paper9.pdf.
16. E.H. Mamdani. Application of Fuzzy Logic to Approximate Reasoning Using Linguistic Synthesis. *Proc. 6th Int. Conf. Multi-Valued Logic*. Logan. UT, 1976, p. 196–202.
17. E.H. Mamdani, Application of fuzzy algorithm for control of simple dynamic plant. *Proc. Inst. Elect. Eng.*, vol. 121. no. 12. 1974, p. 1585–1588. <https://doi.org/10.1049/piee.1974.0328>

18. D. Kalibatiënė, J. Miliauskaitė, A. Slotkienė, Ontology and Fuzzy Theory Application in Information Systems: A Bibliometric Analysis, *J. Informatica*, Vol. 35, no. 3, 2024, p. 557–576. <https://doi.org/10.15388/24-INFOR557>
19. R. Saatchi, Fuzzy Logic Concepts, Developments and Implementation, *J. Information*, 15(10), 656, 2204. <https://doi.org/10.3390/info15100656>
20. M. Malyszko, Fuzzy Logic in Selection of Maritime Search and Rescue Units, *J. Appl. Sci.*, 12(1), 21, 2022. <https://doi.org/10.3390/app12010021>
21. N. Markiz, A. Jrade, Integrating a fuzzy-logic decision support system with bridge information modelling and cost estimation at conceptual design stage of concrete box-girder bridges. *Int. J. Sustain. Built Environ.*, no 3, 2014, p. 135–152. <https://doi.org/10.1016/j.ijse.2014.08.002>
22. B. Cardone, F. Di Martino, Fuzzy Rule-Based GIS Framework to Partition an Urban System Based on Characteristics of Urban Greenery in Relation to the Urban Context, *J. Appl. Sci.*, 10(24), 8781, 2020. <https://doi.org/10.3390/app10248781>
23. A.P. Plerou, E. Vlamou, V. Papadopoulos, Fuzzy Genetic Algorithms: Fuzzy Logic Controllers and Genetics Algorithms. *Glob. J. Res. Anal.*, no. 5, 2016, pp. 497–500.
24. LKIF Ontology. A core ontology of basic legal concepts. URL: <http://www.estrellaproject.org/lkifcore>.

Received: 20.09.2024

Khala K.O.,

Researcher of the Department of Complex Research
of Information Technologies and Systems
<https://orcid.org/0000-0002-9477-970X>, e-mail: khala@irtc.org.ua

*Gladun A.Ya.*¹, PhD (Engineering),
Senior Researcher of the Department of Complex Research
of Information Technologies and Systems
<https://orcid.org/0000-0002-4133-8169>, e-mail: glanat@yahoo.com
International Research and Training Center
for Information Technologies and Systems
of the National Academy of Sciences
of Ukraine and the Ministry of Education and Science of Ukraine,
40, Acad. Glushkov av., 03187, Kyiv, Ukraine

EXPANDING THE CAPABILITIES OF ONTOLOGICAL MODELING OF LEGAL KNOWLEDGE USING ELEMENTS OF FUZZY LOGIC

Introduction. *Ontological analysis is a significant area in the field of intelligent information technologies, particularly in the development of legal systems where there is a continuous need for efficient management and exchange of legal knowledge. Due to the complexity of legal systems, the application of semantic technologies allows for formalizing legal concepts, simplifying the development of ontological models for representing legal knowledge, and integrating heterogeneous legal information systems. Additionally, incorporating fuzzy logic is essential for handling uncertainty and incompleteness in legal information.*

The purpose of the paper is to develop a legal ontology model capable of efficiently processing ambiguous legal terms and concepts while automating the classification and analysis of legal documents. The primary objective is to create a flexible system for formalizing legal knowledge that accounts for the specifics of legal acts, enhances the law enforcement process, and supports informed decision-making.

Methods. *The study employs semantic ontological modeling methods to create legal ontologies and fuzzy logic methods for processing vague and incomplete data. Modern tools and ontology development languages, such as Protege and OWL (Web Ontology Language), are used alongside machine learning techniques for classifying and analyzing legal texts. The approach also explores integrating fuzzy logic elements for evaluating document similarity and representing complex legal concepts.*

Results. A legal ontology model was developed to automate the classification and analysis of legal terms, concepts, and their relationships. The proposed methodology enables the system to extract information from various legal sources and analyze legal documents while addressing ambiguous data. Testing demonstrated improved classification accuracy and increased efficiency in retrieving legal norms from large volumes of unstructured data.

Conclusions. The proposed legal ontology model, incorporating elements of fuzzy logic, significantly enhances the representation and processing of legal knowledge. The methodology includes grammar analysis and the construction of document ontological models, allowing for more precise comparisons of document similarities and differences. The semantic approach proved more effective than the *k*-means clustering method for key phrase classification. Integrating fuzzy sets into the ontology model facilitates the description of imprecise information and supports reasoning with varying levels of completeness. Ongoing work aims to expand the Ukrainian-language version of the legal ontology to address practical challenges in knowledge-based legal systems. The obtained results serve as a foundation for further advancements in intelligent information systems within the legal domain.

Keywords: fuzzy reasoning, fuzzy logic, text analysis, model, decision-making, clustering method, legal knowledge textual content, data processing, knowledge representation, OWL, intelligent information systems.