

УДК 681.3

**С.М. ІВАНОВ**, кандидат технічних наук, доцент,  
**В.В. КАРАСЮК**, кандидат технічних наук, доцент,  
**О.С. ЛУГОВИЙ**, інженер,  
**О.Ю. СОКОЛОВ**, доктор технічних наук, професор

## ОНТОЛОГІЧНІ МОДЕЛІ В КОРПОРАТИВНОМУ ЮРИДИЧНОМУ ІНФОРМАЦІЙНОМУ ПРОСТОРИ

*Анотація.* Щодо створення інформаційно-навчальних систем правових знань.

**Корпоративні інформаційні системи, що базуються на знаннях.** Розвиток інформатизації суспільства неможливий без створення єдиного розподіленого інформаційного середовища, в якому спеціалісти певної галузі знань мають можливість накопичувати дані, знання та типові рішення щодо проблем цієї галузі. Метою функціонування такого середовища є можливість узагальнювати інформацію, що є важливою складовою наукового прогресу в обраній галузі. Для математичного або технічного наукового спрямування таке корпоративне середовище може розглядатися як взаємопов'язана база наукових термінів, визначень, лем, теорем та навіть теорій або технічних рішень і статистичних даних, які використовуються в науці та практиці.

Особливості юридичної інформації полягають в тому, що як для науки гуманітарної більшість знань має описовий характер, для визначень часто використовують синонімічні терміни, тому формалізація саме таких знань є досить нетривіальною проблемою, пов'язаною з необхідністю відслідковувати взаємозв'язки різних гілок знань в даній галузі. Більш того, важливим елементом юридичних знань є тлумачення, тобто додаткова текстова інформація, яка нарівні з визначеннями посідає самостійне важливе місце і яку треба також зберігати.

Важливою також є процедура використання спеціалістами різного рівня – від студента юридичної спеціальності до професора такого інформаційного середовища, де система має автоматично налаштовуватись на рівень подання інформації відповідно до вимог користувача.

Таким чином, процес автоматизації та структуризації інформації для створення корпоративного юридичного простору є задачею створення інтелектуальної комп'ютерної системи, що базується на орієнтованому на знання підході та відповідає всім вимогам до процесу створення систем штучного інтелекту.

В цілому задача корпоративної системи – накопичення не різномірної інформації, а структурованих, формалізованих знань, закономірностей та принципів. Основна мета системи – надати ефективний доступ до накопиченої інформації та забезпечити повторне використання знань. Знання, накопичені в корпоративних системах, стають доступними для більшості користувачів.

Знаннями вважається вся відповідно організована доступна інформація – документи, монографії, закони, словники, підручники, результати практичної роботи тощо.

Оскільки така система має зберігати інформацію у вигляді знань, то при її створенні в першу чергу треба вибрати найбільш відповідну модель знань. В таблиці 1 наведено порівняльні характеристики відомих класичних моделей та їх можливості відносно зазначених вище задач.

Таблиця 1. Моделі знань

Модель подання знань	Характеристика	Переваги та хиби для створення єдиної інформаційної системи
Продукційна	Описує причинно-наслідкові ієрархічно організовані взаємозв'язки процесів мислення, які дозволяють створювати складні структури знань та використовувати механізм прямого або зворотного логічного висновку	<u>Переваги:</u> суворий логічний порядок взаємопов'язаних визначень. <u>Хиби:</u> труднощі зберігання синонімічної інформації, неможливо зберігати текстову інформацію як елемент моделі, неможливість організувати міжмережну взаємодію між об'єктами
Фреймова	Описує ієрархічно організовані взаємозв'язки процесів мислення, використовує механізм логічного висновку на основі процедури наслідування, дозволяє створювати міжмережну взаємодію між об'єктами	<u>Переваги:</u> можливість створення довільної складної структури об'єктів інформаційної системи, можливість використання міжмережних зв'язків між поняттями. <u>Хиби:</u> не завжди потрібний складний механізм наслідування, оскільки у більшості випадків поняття предметної галузі є одноранговими.
Логічна	Описує логічні взаємозв'язки процесів у формі логічних виразів теорії висловлювань або предикатів, використовує механізм дедукції для доказу теорем.	<u>Переваги:</u> формування предметної галузі у вигляді аксіом, логічних висловлювань та строгої теорії математичної логіки для доказу теорем. <u>Хиби:</u> не дозволяє створювати складні структури даних
Семантичні мережі	Описує предметну галузь за допомогою взаємопов'язаних понять та типів зв'язку між ними, використовує механізм узагальнення та декомпозиції для організації висновків.	<u>Переваги:</u> допускає створення синонімічних понять, але неможливо зберігати текстову інформацію як елемент моделі. <u>Хиби:</u> Не дозволяє динамічно змінювати структуру мережі

**Реалізація фреймової моделі на основі онтологічної системи.** Оскільки знання орієнтована система формування інформаційного простору мають спиратись на існуючі моделі, найбільш відповідною для використання у галузі правознавства є фреймова модель знань. Сучасним розвитком реалізації такої моделі є онтологічні системи. На сьогодні відомо декілька визначень онтології.

**Онтологія** – це знання, що формально представлені на базі концептуалізації. Концептуалізація припускає опис безлічі об'єктів і понять, знань про них і зв'язків між ними. Таким чином, онтологією є експліцитна специфікація концептуалізації. Формально онтологія складається з термінів, організованих в таксономію, їх визначень і атрибутів, а також пов'язаних з ними аксіом і правил висновків.

Часто набір пропозицій, що складають онтологію, має форму логічної теорії першого порядку, де терміни словника є іменами унарних і бінарних предикатів, названих відповідно концептами і відносинами. У простому випадку онтологія описує тільки ієрархію концептів, пов'язаних відносинами категоризації. У складніших випадках до неї додаються відповідні аксіоми для виразу інших відносин між концептами і для того, щоб обмежити їх передбачувану інтерпретацію.

Таким чином, онтологія є базою знань, що описує факти, які передбачаються завжди істинними в рамках певного співтовариства на основі загальноприйнятого сенсу використовуваного словника.

Якщо говорити про формальний опис, то більшість джерел пропонують формулювання, аналогічне з [1, с.285], де онтологія представлена трійкою:

$$O = \langle X, R, \Phi \rangle,$$

де  $X$  – скінченна множина концептів (понять, термінів);  $R$  – скінченна множина відношень між елементами  $X$ ;  $\Phi$  – скінченна множина функцій інтерпретації, що задані на концептах і/або відношеннях онтології  $O$ .

Складність онтології залежить від вмісту кожної з множин. В [3] виділяють п’ять рівнів:

- глосарій,
- проста таксономія,
- тезаурус (таксономія з термінами),
- понятійна система з довільним набором відношень,
- повністю аксіоматизована теорія.

Для розв’язання поставленої задачі був вибраний четвертий варіант, в якому множина  $\Phi$  пуста, а  $X$  та  $R$  мають довільне наповнення. Причиною відмови від більш складного рівня зображення онтології стала необхідність автоматизованого (автоматичного) наповнення онтології з текстів. Також довелося відмовитися від систем, що базуються на ієрархії класів з успадкованими наборами атрибутів і механізму екземплярів (таких як Protegé), оскільки виділити означені аспекти у предметної області виявилось проблематичним навіть у ручному режимі.

В цілому система представлена мережею пов’язаних понять з можливістю збільшення не тільки кількості використаних понять, але й кількості зв’язків між ними.

Оскільки інформаційна система, що розробляється, є унікальною, використання стандартних рішень або систем є доволі складною задачею, оскільки такі системи містять досить обмежений арсенал засобів побудови онтологій. Тому запропонована власна структура зберігання онтології (база даних) та програмне середовище для роботи з нею.

На Рис. 1 показано структуру основних елементів онтологічної моделі.

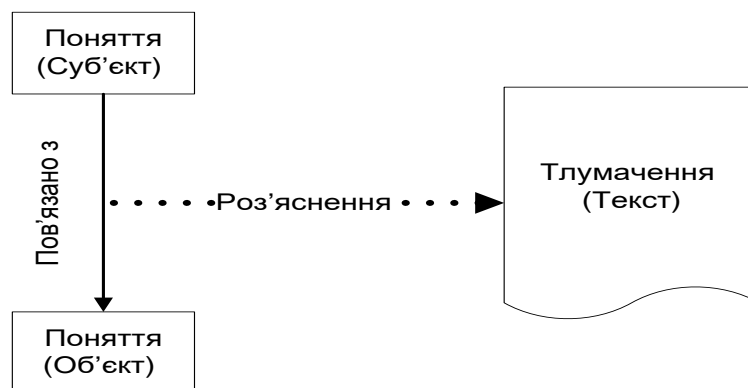


Рис.1. Основний тип зв’язку понять

Одними з найпоширеніших сучасних корпоративних інформаційних систем є сховища даних, в яких, незважаючи на використання аналітичних функцій, міститься інформація гомогенної структури, тобто дані. Корпоративна пам'ять системи, що пропонується, має містити не тільки відомі факти, але й знання у вигляді онтологічних структур. Ця корпоративна пам'ять має фіксувати інформацію з різних джерел та робити її доступною фахівцям при виконанні різних задач.

Структура бази даних містить (див. Рис. 2):

- дані про можливі написання понять та зв'язків між ними;
- зв'язки між поняттями;
- тексти, що відносяться до понять та зв'язків між ними, у вигляді набору речень;
- дані про джерела текстів та їх структуру.

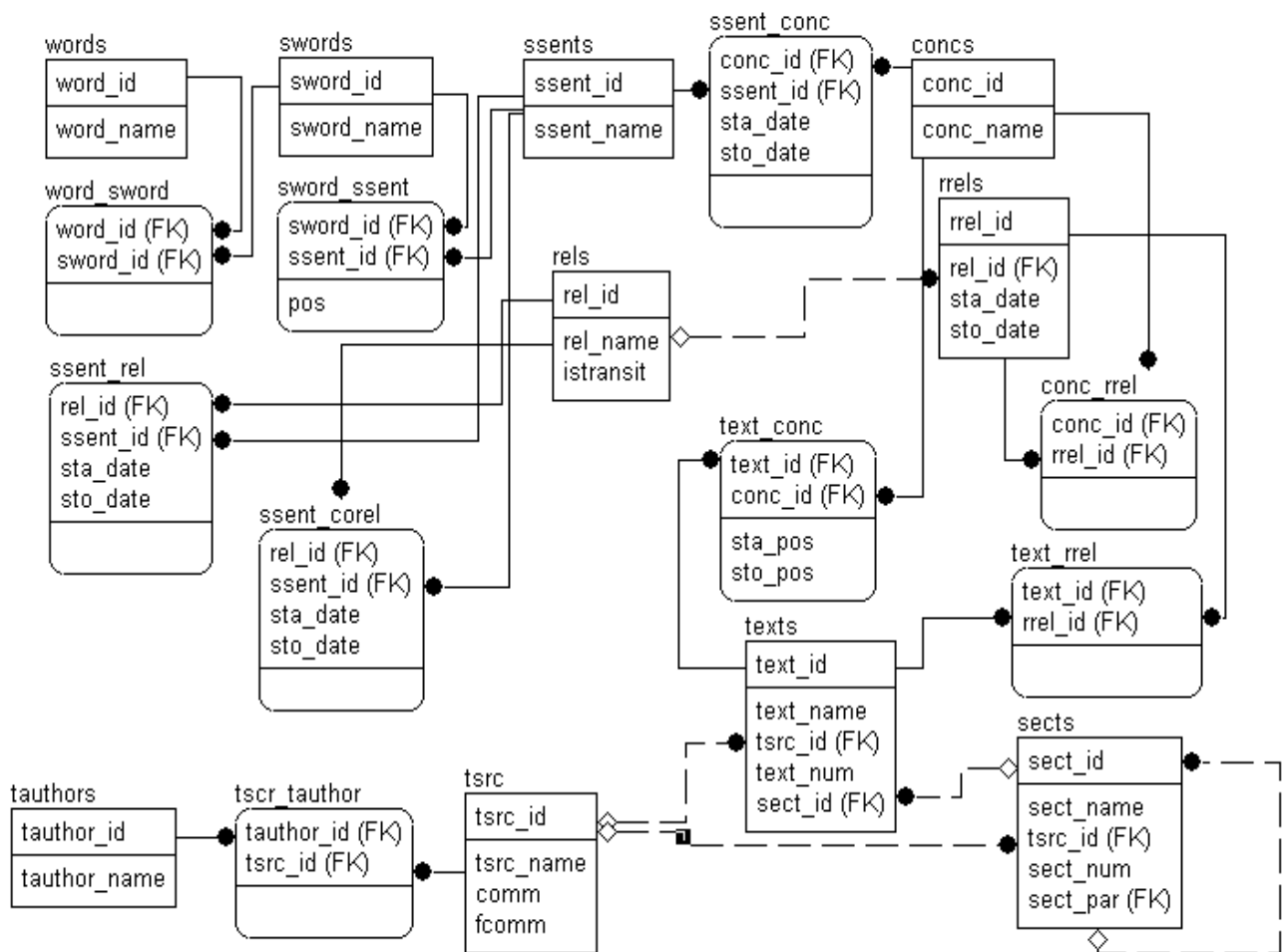


Рис. 2. ER-діаграма бази даних онтології

Різні форми запису понять та зв'язки у тексті зберігаються у восьми таблицях. Усі слова, що були використані для запису понять та зв'язків, записані в таблиці “words”. Слова об'єднуються в групи синонімів, що зберігаються в таблиці “swords”. Кожне з понять (таблиця “concs”) та зв'язок (таблиця “rels”) зображується у вигляді множини словосполучень (таблиця “sents”), які являють собою упорядкований набір груп синонімів. Зв'язки між поняттями записуються як у прямому вигляді, так і в оборотному. Вони зберігаються в таблицях “rrels” та “conc\_rrel”. Перша зберігає тип зв'язку та дати його дії, друга – безпосередньо набір понять, пов'язаних між собою.

Таблиці “text\_conc” та “text\_rel” зберігають зіставлені з поняттями та зв’язками між поняттями речення відповідно. До того ж у випадку з поняттями зберігається його місце в реченні (у вигляді початкового та кінцевого номерів символів). Тексти представлені у вигляді набору речень. Речення зберігаються в таблиці “texts”, що також зберігає номер джерела й розділу, котрому належить речення, та позицію речення в розділі. Номери розділів, їх ієрархія, номер в батьківському розділі та номер джерела зберігаються в таблиці “sects”. Атрибути джерел текстів зберігаються в таблиці “tsrc”, котра зв’язана з таблицею авторів джерел “tauthors”.

Таким чином, хоча запропоноване зображення зберігання понять та зв’язків між ними порівняно з простим послівним записом сповільнює роботу з базою в певних випадках, проте забезпечує значну гнучкість як зберігання, так і пошуку даних.

Також слід відзначити, що запропонована структура забезпечує повне зберігання джерел онтології з накопиченням зв’язності між ними. Серед передбачених функціональних задач системи треба відзначити:

- 1) запис та корегування зображення понять та зв’язків у тексті;
- 2) запис та редагування безпосередніх зв’язків між поняттями;
- 3) автоматизований пошук в тексті понять з визначеної онтології та виявлення зв’язків між ними;
- 4) видача всіх відомих зв’язків вказаного поняття з посиланням на їх джерела.

Для роботи з онтологією запропоновано спеціальне середовище, деякі з екранних форм якого наведено на Рис. 3.

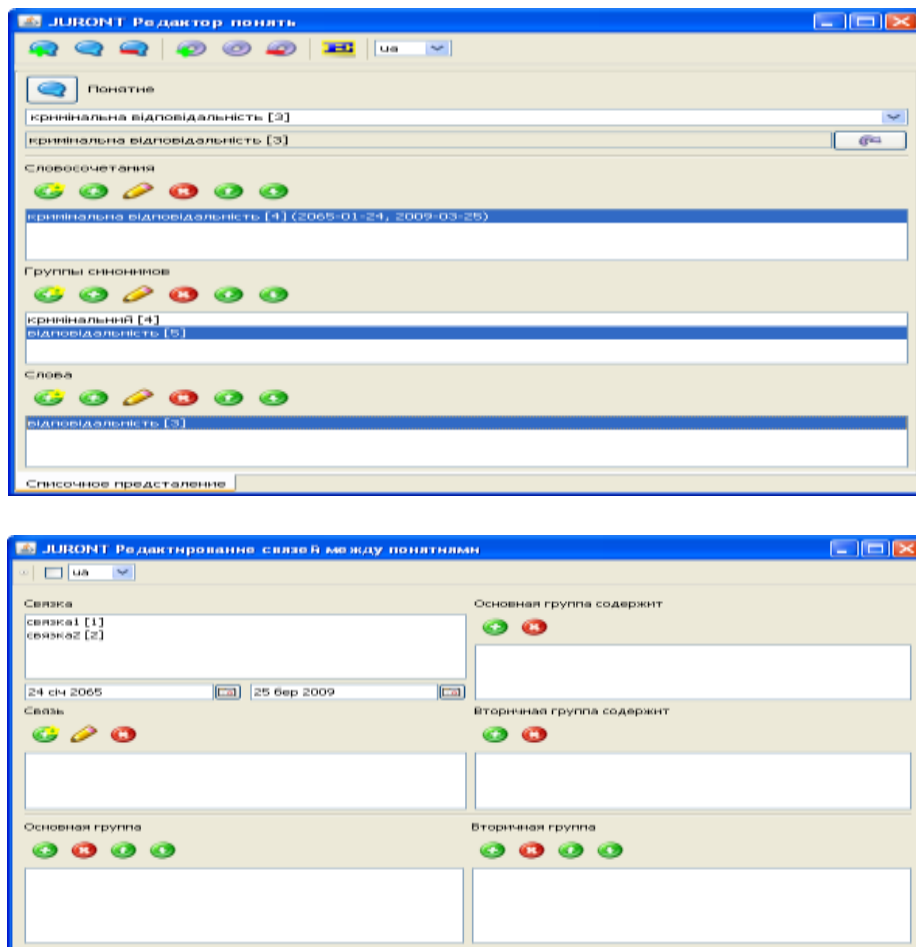


Рис. 3. Екранні форми редагування понять та зав’язків між ними

Оскільки система, що будується, буде використовуватися в інтегрованому середовищі, пропонується створення веб-портала для роботи з нею. На Рис. 4 наведено засоби організації використання системи в мережі.

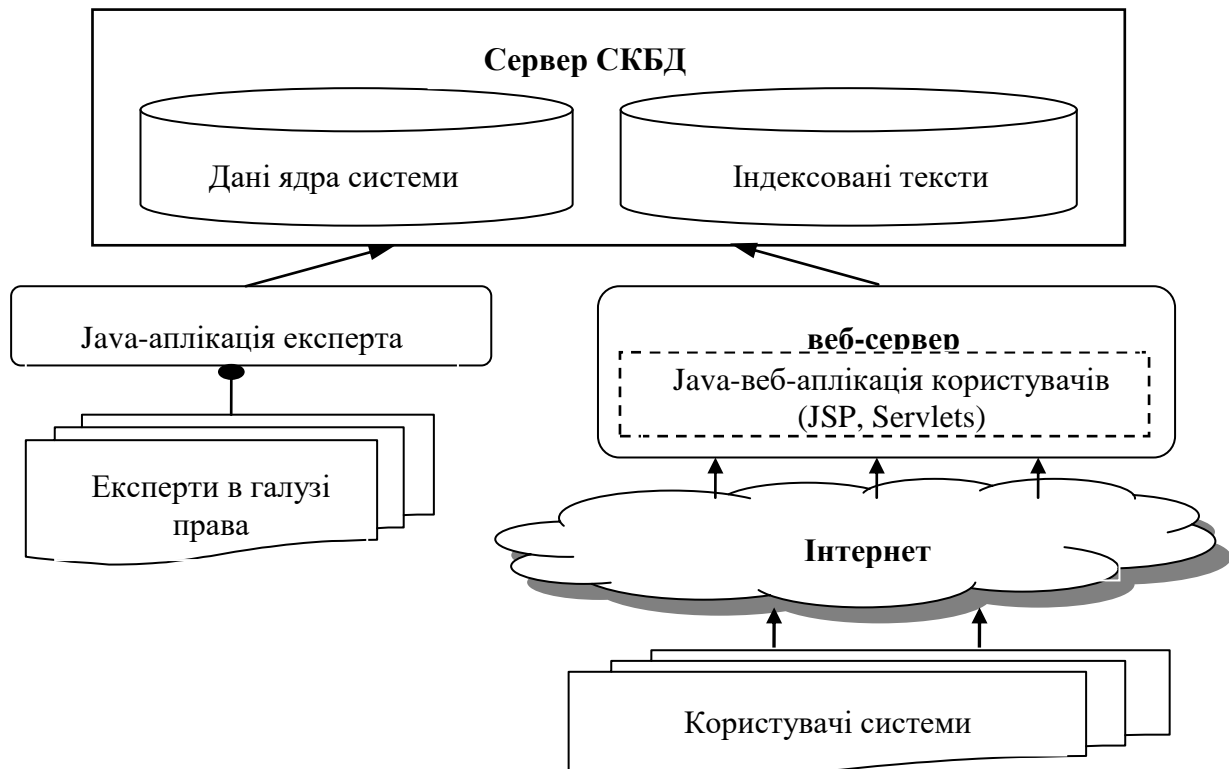


Рис. 4. Мережне використання онтологічної системи

**Режими роботи з онтологічною системою.** Відповідно до цілей створення єдиного інформаційного середовища юридичної діяльності передбачаються наступні режими роботи з системою. Представимо та охарактеризуємо їх.

- Режим створення онтологічної системи. У режимі фахівці-когнітологи розробляють ядро онтологічної системи та наповнюють середовище первинними знаннями.
- Режим експлуатації. Передбачає одночасне використання системи багатьма користувачами, де знання накопичуються в єдиній базі даних.
- Режим індивідуального доповнення, розвитку та інтеграції. Передбачає роботу в локальній версії програми зі створенням власної бази знань. Також можлива подальша інтеграція моделей, створених окремими користувачами, в єдину систему з перевіркою можливих протиріч у тлумаченні.
- Навчання та оцінювання знань. Режим дозволяє створювати систему оцінювання на базі ланцюгів взаємодії понять та формування онтології “питання – відповідь”.
- Тренінги та формування умінь. Призначений для навчання користувачів прийняттю рішень в умовах неповної інформації, спеціально частково прихованої, тобто створення онтології “ситуація – дія”.
- Онтологічний аналіз. В основі онтологічного аналізу лежить опис предметної області в термінах сутностей, їх взаємозв’язку, а також перетворення сутностей в процесі прийняття рішень.

Основною операцією в режимі експлуатації є запити до онтологічної системи. Оскільки система може мати неоднозначності та протиріччя, запропоновано розробити спеціальну мову запитів, орієнтовану на нечіткі формулювання та нечіткі результати [4].

З цією метою розроблено мову нечітких запитів як розширення мови запитів до баз даних SQL. На Рис. 5 наведено структурну схему підсистеми обробки нечітких запитів.

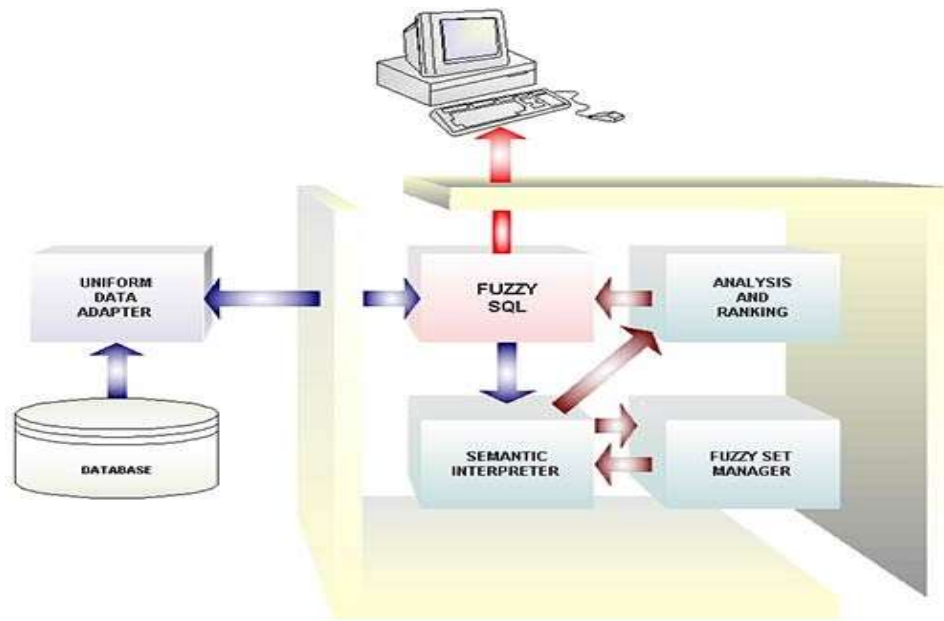


Рис. 5. Підсистема нечітких запитів до онтологічної системи

Структура запиту наступна:

`FUZZY SELECT * | expression [AS output_name] [...]`

`FROM from_item [...]` WHERE condition,

результат запиту – множина альтернатив онтологічної системи, релевантних нечіткому умовному вислову.

**Перспективними вважаються наступні напрями:**

- Порівняння онтологій. Це дуже важливий інструмент для оцінки повноти, суперечності онтологій, створених різними користувачами, а також для порівняння онтологій різних напрямів юридичної діяльності або різних країн
- Спілкування на природній мові. Важлива частина створення інтелектуальних систем, яка дозволить залучати до роботи непрофесійних користувачів, або використовувати систему для автоматичного наповнення баз знань з довільних текстів
- Використання у юридичній клініці. За допомогою пропонованої системи можливо створення інтернет-портала для надання консультаційних послуг без участі фахівця.

### Використана література

1. Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – СПб.: Питер, 2000. – 383 с.
2. Константинова Н.С. Онтологии как системы хранения знаний / Н.С. Константинова, О.А. Митрофанова // Всероссийский конкурсный отбор обзорно-аналитических статей по приоритетному направлению “Информационно-телекоммуникационные системы”, 2008. – 54 с.
3. Добров Б.В. Лингвистическая онтология по естественным наукам и технологиям для приложений в сфере информационного поиска / Б.В. Добров, Н.В. Лукашевич. – Режим доступу: [//www.fccl.ksu.ru/issue\\_spec/docs/oent-kgu.doc](http://www.fccl.ksu.ru/issue_spec/docs/oent-kgu.doc)
4. The Fuzzy SQL Tutorial: Scianta Intelligence, 2005. – Р. 9. – Режим доступу: [//www.scianta.com/pdfs/FS-TU-001%20FuzzySQL%20Tutorial.pdf](http://www.scianta.com/pdfs/FS-TU-001%20FuzzySQL%20Tutorial.pdf)