

# Онтологічний інжиніринг інформаційно-освітнього середовища

А.П. Гетьман<sup>1</sup>, С.М. Іванов<sup>1</sup>, В.В. Карасюк<sup>1</sup>

*Анотація* – The paper is devoted to development of ontology application in intellectual technologies and systems. The ontology engineering problems are discussed. The report deals with the principles of semantic analysis in the ontological representation of law information.

*Ключові слова* – Семантична мережа, Онтологія, Графова модель, Web-інтерфейс.

## I. ВСТУП

Необмежені перспективи створення баз і сховищ знань надають технології Інтернету. Одне з напрямків акумулювання знань для забезпечення доступу до них широкого кола користувачів пов'язано з розвитком відомих систем типу Wikipedia. Але відкриті доступні системи не мають необхідних механізмів управління своїм контентом, особливо при використанні слабоструктурованої інформації. У Центрі інформаційних технологій Національного університету «Юридична академія України ім. Я. Мудрого» ведуться роботи зі створення системи навчання, яка орієнтована на роботу з правовою інформацією і базується на принципах штучного інтелекту [1].

## II. ТЕОРЕТИЧНІ ЗАСАДИ ДЛЯ СТВОРЕННЯ БАЗИ ЗНАТЬ ПРАВОВОЇ ІНФОРМАЦІЇ

Сьогодні на передньому плані розробок у сфері створення баз знань знаходяться принципи формалізації інформації у вигляді семантичної мережі, архітектура якої передбачає наявність у будь-якої інформації, що знаходиться в мережі, пов'язаного з цією інформацією точного змісту, який не можна було б переплутати навіть у разі збігу фраз або слів, зустрінutih в різних контекстах. Фактично це означає, що будь-яка інформація пов'язана з деяким невіддільним від неї контекстом. В основі моделей більшості систем, заснованих на знаннях, лежать онтології предметних областей. Використання онтології особливо необхідно в додатках-агентах, що здійснюють пошук і об'єднання інформації з різних джерел і з різних середовищ, в яких той же термін може означати різні речі. По кожній з аналізованих тем формується набір речень - цитат, що відносяться до відповідних термінів, які представляють тематичне наповнення тексту. Знання асоціативних зв'язків дозволяє виявити залежності між об'єктами, які покривають певну предметну область. У галузі правознавства є певні особливості представлення інформації: - великий обсяг, наприклад, чинна нормативна база України об'єднує

близько 300 тисяч документів; - ця інформація в основному представлена у вигляді неструктурованої текстової інформації, яку важко формалізувати; - правова інформація повинна бути достовірною, строго відповідати чинному законодавству та правозастосовним документам; - повинна бути своєчасною, інакше вона втратить своє практичне значення; - має часові обмеження своєї легальності; - розташована в окремих базах даних, сховищах, локальних мережах, для кожного її виду застосовується свій формат. Ідея використання онтологічних схем для представлення системи знань в навчально-контролюючих системах вже давно опрацьовується фахівцями. Як приклад успішної розробки такого типу вкажемо на систему "Бігор" (База і генератор освітніх ресурсів на основі технології контенту розділяємих одиниць), розроблену в Московському державному технічному університеті ім. М. Я. Баумана [2].

## III. СЕМАНТИЧНИЙ АНАЛІЗ В ОНТОЛОГІЧНОМУ УЯВЛЕННІ

Крім власне зберігання інформації онтологічна система дозволяє виконувати семантичний аналіз даних. Онтологія в загальному випадку являє собою граф з розміченими вершинами і дугами. Формально онтологія представляє собою пару  $\langle T, L \rangle$ , де  $T$  - множина вершин графа,  $T = \{t_1, \dots, t_w\}$ , що представляють поняття онтології,  $L$  - множина дуг графа, що представляють зв'язки між термінами онтології,  $L = \{l_1, \dots, l_v\}$ .

Кожна вершина  $t_i \in T$  характеризується своїм ім'ям і типом,  $t_i = \{\text{termname}, \text{type}\}$ . Termname - унікальне ім'я поняття онтології, що пов'язано з вершиною. Type - позначає тип вершини.  $\text{Type} = \{\text{term} \mid \text{terminal} \mid \text{value}\}$ . Term - означає термін предметної області. Terminal - позначає термінальну вершину, при цьому, термінальні вершини характеризуються типом значення. Terminal = {valuetype}. Valuetype - тип значення термінальної вершини.  $\text{Valuetype} = \{\text{integer} \mid \text{string} \mid \text{boolean} \mid \text{float} \mid \text{date-time}\}$ . Тип значення: цілий (integer), рядковий (string), логічний (boolean), число з плаваючою точкою (float) і дата-час (date-time). Value - конкретне значення, що належить певному типу.

Кожна дуга  $l_j \in L$  характеризується своїм ім'ям, множинністю початкової вершини і множинністю кінцевої вершини.  $l_j = \{\text{name}, \text{multiplicitybegin}, \text{multiplicityend}, \text{optionality}\}$ . Name - ім'я зв'язку, який представляє дуга. Multiplicitybegin - множинність

<sup>1</sup> Національний університет «Юридична академія України імені Ярослава Мудрого», вул. Пушкінська, 77, Харків, 61024, УКРАЇНА, E-mail: karasiuk@yahoo.com

початкової вершини. Multiplicitybegin = {concrete | one | set}. Concrete - позначає конкретну вершину. One - позначає одиночну вершину. Set - позначає множину вершин. Multiplicityend = {concrete | one | set | variants | sequence}. Аналогічно Concrete - позначає конкретну кінцеву вершину. One - позначає одиночну вершину. Set - позначає множину вершин. Variants - позначає множину варіантів. Sequence - позначає послідовність. Optionality = {true, false}, - факультативність, позначає, чи є даний зв'язок обов'язковим чи може бути відсутнім.

Одною з головних переваг онтологічного представлення в порівнянні з базами даних є можливість побудови ієрархії класів. У зв'язку з цим актуальним є питання про порівняння різнотипних об'єктів, тобто визначення їх семантичної подоби - функції, принципово не існуючої в базах даних.

Вважається, що два об'єкти  $e_1$  і  $f_1$  є семантично подібними, якщо існує два семантичних шляхи  $e_1, P_1, e_2, P_2, e_3, P_3, \dots, e_{n-1}, P_{n-1}, e_n$  і  $f_1, Q_1, f_2, Q_2, f_3, \dots, f_{n-1}, Q_{n-1}, f_n$ , що семантично з'єднують  $e_1$  з  $e_n$  і  $f_1$  з  $f_n$ , відповідно, і таких, що для кожної пари властивостей  $P_i$  і  $Q_i$ ,  $1 \leq i < n$ , виконується одна з умов  $P_i = Q_i$ , або  $P_i \subseteq Q_i$ , або  $Q_i \subseteq P_i$ .

Подібність об'єктів обчислюється на 3 рівнях: рівні даних, онтологічному рівні і контекстному рівні.

#### IV. ПРОГРАМНА РЕАЛІЗАЦІЯ ЗНАННЯОРІЄНТОВАНОЇ ІНФОРМАЦІЙНОЇ СИСТЕМИ

Розробка запропонованого підходу виконана з використанням сучасних технологій об'єктно-орієнтованого візуального програмування, в середовищі Eclipse 3.4 на мові Java з підтримкою JDK версії 1.6.

Структура бази даних містить: дані про можливі записи понять і зв'язків між ними; зв'язки між поняттями, тексти, які відносяться до понять і зв'язків між ними у вигляді набору речень; дані про джерела текстів та їхню структуру.

В структурі програмного комплексу реалізовано три аплікації: - аплікація для наповнення онтологічної структури та текстового контенту (аплікація експерта); - для перегляду текстів з можливістю навігації по ним на основі онтологічного ядра, це безпосередня навігація по онтології для вивчення принципових взаємозв'язків області знань, що вивчається (аплікація користувача); - допоміжна аплікація експерта для початкової підготовки текстової інформації (приведення різноманітних форматів до спеціального, визначеного при розробці системи).

Програмний комплекс передбачає роботу з двома основними сутностями: онтологією (вміщає поняття та зв'язки між ними) та текстами. Для зберігання даних було обрано реляційні бази даних, що передбачає створення системи взаємопов'язаних таблиць. В узагальненому вигляді структура бази даних складається з таких частин: - поняття і зв'язки; - зв'язки між групами понять; - тексти-джерела; - словесне зображення понять і зв'язок; - індекси вживання понять та зв'язків в тексті (Рис. 1).

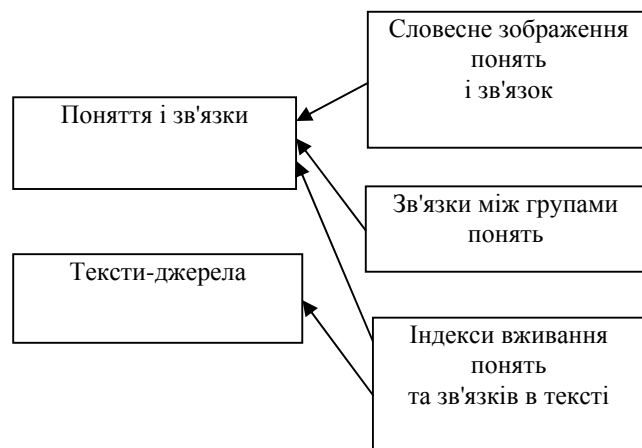


Рис. 1. Загальна структура бази даних

Розроблений програмний комплекс JURONT (юридична онтологія) знаходиться у дослідній експлуатації в рамках локальної мережі юридичної академії, одночасно продовжується процес його вдосконалення.

### III. ВИСНОВОК

У роботі запропоноване рішення задачі формування загальної бази знань навчання спільнотою користувачів в галузі правознавства. В якості моделі представлення інформації обрано її опис у вигляді онтології. Для порівняння побудованих різними користувачами онтологій використовуються графі онтологій. Надалі планується впровадження онтологічної системи навчання з використанням принципів самоорганізації як однієї із складових частин порталу спрямованого (індивідуального) навчання студентів університету. Перспективні дослідження передбачається виконати в напрямку подання нечітких зв'язків між поняттями в базі знань. Також передбачається дослідити вплив принципів самоорганізації на якість створюваної багатьма користувачами онтології в предметній області правознавства.

#### СПИСОК ПОСИЛАНЬ

- [1] Tatsyi, V. Семантична мережа знань в правознавстві = Semantic network of knowledge in science of law / V. Tatsyi, A. Getman, S. Ivanov, V. Karasiuk, O. Lugoviy, O. Sokolov // Автоматика, управління і інформаційні технології: Праці IASTED Міжнародної конференції = Automation, Control, and Information Technology (ACIT 2010): Proceedings of the IASTED International Conference on Automation, Control, and Information Technology, June 15 – 18 2010, Novosibirsk, Russia / The International Association of Science and Technology for Development. – Anaheim, USA, Calgary, Canada, Zurich, Switzerland: ACTA Press 2010. p. 218 – 222.
- [2] Норенков И.П. Интеллектуальные технологии на основе онтологий // Информационные технологии, 2010, № 1, с. 17-23.