

БАЗИ ДАНИХ, БАЗИ ЗНАНЬ ТА ІНЖЕНЕРІЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

УДК 004.853, 004.55

Рогошина Ю.В.

Інститут програмних систем НАН України

# СЕМАНТИЧНИЙ ПОШУК СКЛАДНИХ ІНФОРМАЦІЙНИХ ОБ'ЄКТІВ 3 ВИКОРИСТАННЯМ ОНТОЛОГІЧНИХ ЗНАНЬ

*Розглядаються проблеми, пов'язані з пошуком інформації в Web інтелектуальними застосуваннями. Онтологічний аналіз використовується як основа для подання знань у семантичному пошуку. Запропонована онтологічна модель взаємодії між відкритим інформаційним середовищем, інтелектуальною інформаційною системою та її користувачами. Розроблено метод здобуття знань про складні інформаційні об'єкти, структура яких теж формалізується за допомогою онтологій, який проаналізовано на прикладі задачі аналізу компетенцій.*

*Рассматриваются проблемы, связанные с поиском информации в Web интеллектуальными приложениями. Онтологический анализ используется как основа для представления знаний в семантическом поиске.*

*Предложена онтологическая модель взаимодействия между открытой информационной средой, интеллектуальной информационной системой и ее пользователями. Разработан метод получения знаний о сложных информационных объектах, структура которых также формализуется с помощью онтологий, которые проанализированы на примере задачи анализа компетенций.*

*The problems deal with the Web information retrieval by intelligent applications are analysed. Ontological analysis is used as a basis for knowledge representation in the semantic search. Ontological model of the interaction between the open information environment, intelligent information system and its users is proposed. The method of acquisition of knowledge about the complex information objects which structure is also formalized by ontologies is developed and analyzed on example of the competence analysis problem.*

**Ключові слова:** семантичний пошук, онтологія, інформаційний об'єкт, знання.

## Вступ

Значна частина сучасних інформаційних систем тією чи іншою мірою інтелектуальні й орієнтовані на використання і перетворення знань щодо цікавлячої користувача предметної області. Більшість з таких застосовних програм динамічні, тобто орієнтовані на функціонування у відкритому інформаційному середовищі, зокрема, на пошук і використання ресурсів Web. Одним з визначальних факторів ефективності інтелектуальних інформаційних систем, що орієнтовані на роботу в розподіленому середовищі Web, є їх здатність до пошуку актуальних та пертинентних відомостей, що потрібні для вирішення задач користувачів. При цьому об'єктом пошуку можуть бути не тільки дані, але й програми, сервіси, знання та інші складно структуровані інформаційні сутності.

Таким чином, проблема пошуку в Web виявляється складовою частиною найрізноманітніших інформаційних систем, а різні прогнози в області інформаційних технологій вказують на загальну тенденцію: у найближчі роки обсяг інформації і потреба в ній будуть збільшуватися експоненційно. Тому при інформаційному пошуку виникає потреба в переході від обробки великих обсягів даних до обробки знань – значно більш компактних, але зі значно складнішою структурою.

Швидке зростання обсягу інформаційних ресурсів (IP), до якої можна отримати доступ через Web, а також ускладнення її структури зумовлюють усе більшу потребу в розвитку автоматизації та інтелектуалізації засобів знаходження таких відомостей, в побудові моделей та методів семантичного пошуку, який дозволяє використовувати знання щодо

користувачів, інформаційних ресурсів та тих об'єктів, що мають стати результатом пошуку, здобуті як з зовнішніх джерел, так і з досвіду виконання пошукового процесу. Персоніфікація семантичного пошуку, що базується на використанні знань про конкретного користувача, його сферу інтересів, поточні інформаційні потреби, здатність до сприйняття інформації та досвід, дозволяє інформаційним системам є одним з важливих напрямків наукових досліджень в цій сфері.

У семантичному пошуку широко застосовуються технології проекту Semantic Web, розробленого Т.Бернес-Лі, та онтологічний аналіз, якому присвячені дослідження Т.Грубера, Н.Гуаріно, Д.Сова. Семантичний пошук є комплексною науковою задачею, він базується на досягненнях в області штучного інтелекту, зокрема – загальної теорії представлення та обробки знань, розпізнавання образів та логічного виведення; методах математичної статистики; соціопсихології, математичної лінгвістики тощо. Існує велика кількість інформаційно-пошукових систем (ІПС), що базуються на цих дослідженнях, в яких реалізовано різні аспекти семантичного пошуку, найбільш відомі серед яких – це Nukia, WolframAlfa, Cognition, Sensebot. Цій проблемі приділяли значну увагу багато вчених, чії дослідження були стосувалися інформаційного пошуку та його вдосконаленню на основі знань: Р.Басза-Ятес, С.Брін, К.Маннінг, Л.Пейдж, А.С.Нариньяні, Б. В. Добров, Н. В. Лукашевич, А. Ф. Тузовський тощо. Серед українських дослідників проблемам, пов'язаним з семантичним пошуком та аналізом знань, приділяли увагу П.І.Андон, А.В.Анісімов, В.П.Гладун, Ю.Р.Валькман, Д.В.Ланде.

Але ці дослідження не вирішують всі проблеми, що виникають в процесі семантичного пошуку у відкритому інформаційному просторі Web. Це зумовлює потребу в формалізованому представленні відомостей щодо інформаційних об'єктів та інтересів користувачів та в методах їх співставлення, засоби їх автоматизованого здобуття, а також в алгоритмах залучення до пошукових процедур знань, специфічних до обраної предметної області (ПрО). Крім того, щоб результати пошуку викликали довіру користувачів та відповідали їх реальним потребам, необхідно надати їм гнучкі засоби керування процесом семантичного пошуку та пояснення його результатів. Саме це визначає

актуальність побудови теоретично обґрунтованої онтологічної моделі семантичного пошуку, яка формалізує відношення між інформаційним середовищем сучасного Web та користувачами інформаційно-пошукових систем, їх інформаційними потребами та проблемами, які вони прагне вирішити; джерел поповнення цієї моделі та методів використання отриманих знань для персоніфікації семантичного пошуку, і, таким чином може бути основою для інтеграції семантичного пошуку з різними типами інтелектуальних систем, що працюють у відкритому середовищі Web, спрямовані на обробку знань та застосовують інформацію з відкритих джерел.

### Постановка задачі

Створення онтологічної моделі, яка формалізує інформаційні відношення між відкритим інформаційним середовищем, інтелектуальною інформаційною системою (ІС) та її користувачами, є актуальною науково-прикладною проблемою, вирішення якої потребує розробки представлення знань про предметну область пошуку, його об'єкти та суб'єкти, побудови методів як здобуття цих знань з різних інформаційних джерел, так і методів використання цих знань для підвищення ефективності роботи ІС.

Це визначає необхідність розробки деталізованої онтологічної моделі інформаційних потреб користувачів ІС, а саме – тих інформаційних об'єктів (ІО), що обробляються в ІС, їх структуру та властивості, та розробити методи та засоби використання та здобуття інформації щодо цих ІО, які дозволяють побудувати систему семантичного пошуку, орієнтовану на персоніфіковану обробку інформації та запропонувати шляхи інтеграції підсистеми семантичного пошуку з різними типами інтелектуальних застосовних ІС.

### Семантичний пошук та його характерні особливості

Інформація в Web має тенденцію до структуризації, тобто широко використовуються бази даних і програми роботи з ними. При цьому найбільше природно шукати інформацію безпосередньо за місцем збереження – у базах даних і інших джерелах структурованої інформації (наприклад, XML-документах і сховищах). Виконання пошуку на самих Web-сайтах дозволяє власникам ресурсів

контролювати правильність роботи пошуку. Для кожного типу інформаційних джерел та кожного типу інформаційних об'єктів, які шукає користувач, потрібно мати свій клас пошукових компонентів з уніфікованим інтерфейсом взаємодії.

Коли у пошуку інформації застосовують будь-яким чином знання та засоби керування знаннями, то кажуть про *семантичний* або *інтелектуальний* пошук. Для того, щоб ефективно оперувати цими термінами, надалі потрібно дати більш чітке визначення цих термінів.

*Система семантичного пошуку* (ССП) – це інформаційна система, що забезпечує пошук та розпізнавання ІО різних типів з використанням знань для співставлення запиту з наявними інформаційними ресурсами на семантичному рівні. ССП можна розглядати як певну інтелектуальну надбудову над традиційними інформаційно-пошуковими системами (ІПС) – як загального призначення, так і спеціалізованими. Можна казати, що ССП є інтелектуалізацією ІПС. ССП дозволяють обробляти більш складні ІО та використовувати знання щодо відповідної предметної області.

Процес пошуку можна зробити більш інтелектуальним за допомогою структурування даних, отриманих з Web та здійсненню пошуку в «глибинному» Web [1]. Подальша інтелектуалізація пошуку пов'язана також із застосуванням онтологічного подання знань, яке забезпечує їх багаторазове використання [2]. З розвитком інфраструктури Semantic Web семантичні метадані стали більш доступними. Розробка стандартів семантичної розмітки, мов опису ІР (RDF), онтологічних мов, сервісів, онтологічних баз, систем пошуку в семантичних даних (Swoogle, SWSE, WatsOn), точок SPARQL-доступу, систем логічного виведення, обробки правил, тощо, сприяє подальшому розвитку області інформаційного пошуку в напрямку використання семантики [3].

У найбільш узагальненому розумінні *пошук інформації* – це зіставлення представлення користувача про потрібних йому знаннях з контентом доступних ІР і побудови на основі цього зіставлення ІО, значення властивостей якого витягаються з цих ІР. Найбільш перспективні напрямки розвитку інформаційного пошуку пов'язані зараз з його персоніфікацією та інтелектуалізацією. При семантичному пошуку в цьому зіставленні використовуються *знання* щодо різних суб'єктів пошуку – користувачів, ресурсів, результатів

раніше виконаних пошукових процедур, а також знання щодо Про пошуку. Якщо для формалізації знань про суб'єктів цього процесу використовуються онтології, то можна говорити про *онтологічну модель* пошуку.

Надалі під *семантичним пошуком* будемо розуміти процес пошуку інформації, який задовольняє інформаційну потребу користувача, що виникає у нього в процесі вирішення певної проблеми, в якому застосовуються – явно або неявно для користувача – знання щодо різних суб'єктів та об'єктів пошукової процедури та методи аналізу цих знань [4]. Ці знання можуть стосуватися як самого користувача та його інформаційних потреб (персоніфікація пошуку), так і ІР, серед яких здійснюється пошукова процедура.

Семантичний пошук – це метод інформаційного пошуку, у якому релевантність документа запиту визначається семантично (за близькістю змісту), а не синтаксично (приміром, за частотою використання ключових слів у документі). Його можна розглядати як надбудову над традиційним інформаційним пошуком, в якому для підвищення пертинентності пошуку (тобто для більш ефективного задоволення інформаційних потреб користувача). ССП – це програмне забезпечення, що виконує семантичний пошук – незалежно або як семантична надбудова над іншими ІПС. При цьому взаємодія ССП з іншим програмним забезпеченням та базами знань може бути відкрита або прихована від кінцевого користувача. Результатом роботи ССП може бути як здобуття інформації, неявно присутньої в певному ІР (як текстовому, так і мультимедійному), так і надання користувачеві відомостей про наявні ІР у певному порядку та певній формі, що відповідають персональним потребам саме цього користувача.

Для семантичного пошуку у відкритому середовищі характерно, що потрібні для пошуку знання також динамічно здобуваються з цього відкритого середовища, а не закладаються в систему в процесі її створення. Семантичний пошук може здійснюватися в розподіленому інформаційному середовищі, використовуючи додаткові знання щодо цього середовища (приміром, пошук в InternetofThings або в WebofThings, в середовищі GRID) або щодо його особливостей певного класу інформаційних об'єктів (приміром, пошук Web-сервісів або онтологій). Те, які саме знання

використовуються в ССП, як вони представлені і як вони обробляються, залежить як від специфіки розроблювальної ППС, так і від концепції, обраної її розроблювачами, але в загальному випадку результати пошуку є функцією від запиту користувача, індексу ППС та знань, що містяться в базі знань (БЗ) ППС:  $I_s = \{i_j, j = \overline{1, n}\} = f(z, DB_{unc}, KB_{unc})$ .

Якщо мова йде про семантичний пошук у Web, то слід враховувати, що при цьому в Web можуть знаходитися не тільки інформаційні об'єкти, серед яких здійснюється пошук, але і зовнішні бази знань, що використовуються для пошуку. Тому при створенні таких систем варто враховувати, що такі зовнішні БЗ можуть змінювати контент, структуру і доступність незалежно від розроблювачів ППС. Тому результати пошуку в Web є функцією ще й від вмісту зовнішніх БЗ:

$$I_{web\_s} = \{i_j, j = \overline{1, n}\} = f(z, DB_{unc}, KB_{unc}, \{KB_{web\_k}, k = \overline{1, m}\})$$

### Семантичний пошук інформаційних об'єктів

Виконання семантичному пошуку, на відміну від звичайного, дозволяє користувачеві вказувати, що саме має бути предметом пошуку. При цьому ССП може знаходити не якийсь конкретний інформаційний ресурс (ІР) – документ чи його фрагмент, а відомості про ІО певного класу, тобто користувач може (явно чи неявно) вказати клас шуканого об'єкта. Це може бути досить простий і розповсюджений клас, наприклад, «людина» чи «мультимедійний об'єкт», або клас,

специфічний для певної ПрО, наприклад, «наукова публікація». Користувач може явним чином вказувати тип інформаційного об'єкта, який він прагне знайти, використовуючи для цього відповідні стандарти та таксономії, приміром, для здійснення пошуку в репозиторіях онтологій або Web-сервісів [5], серед RDF-описів та XML-структур [6]. Крім того, семантичний пошук дозволяє інтегрувати відомості щодо ІО, здобуті з різних приступних ІР.

*Інформаційний об'єкт* – це інформаційна модель об'єкту предметної області (ПрО), яка визначає структуру, атрибути, обмеження цілісності і, можливо, поведження цього об'єкта. В семантичному пошуку під ІО будемо розуміти ту інформацію, яку користувач отримує в результаті виконання процесу пошуку. Важливо, що сам об'єкт ПрО може бути як матеріальним (наприклад, людина, організація, місце), так і віртуальним, що не має конкретного матеріального еквівалента (наприклад, книга, програма), хоча і зв'язаного з тим чи іншим матеріальною носієм інформації. Крім того, об'єкт ПрО може являти собою абстрактне поняття, у принципі не пов'язане з конкретним носієм інформації, яке відображається лише у свідомості людини (наприклад, «час», «мораль»).

Семантичний пошук можна розглядати як окремий випадок проблеми розпізнавання, якщо результатом пошуку має стати інформаційний об'єкт зі складною структурою, знання про яку використовуються в пошукових процедурах. При цьому задача розпізнавання поділяється на наступні етапи (рис.1):

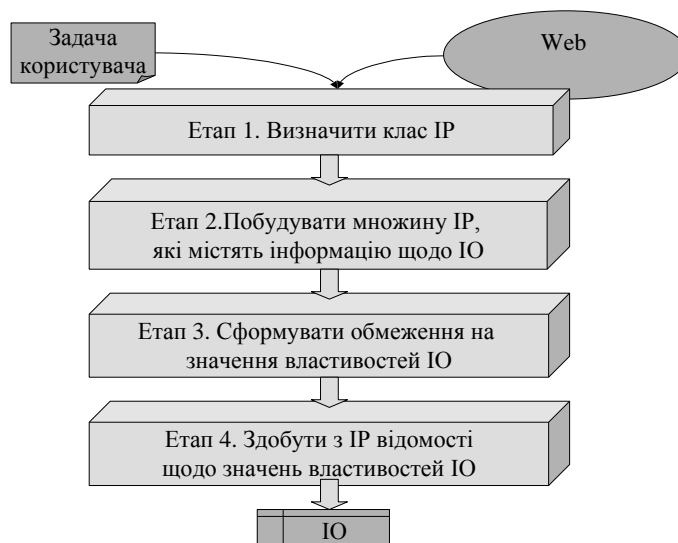


Рис.1. Етапи розпізнавання ІО в Web

При цьому можуть застосовуватися методи лінгвістичного аналізу, розпізнавання мови і зображень (для аналізу мультимедійних ІР), а також засоби менеджменту знань, спрямовані на автоматизоване використання знань про структуру шуканих ІО, представлених у інтероперабельній формі, що допускають автоматизовану обробку.

Онтологію можна використовувати як основу для представлення структури ІО (класи), а ІР – для створення екземплярів ІО. Це дозволяє інтегрувати інформацію з різних ІР.

При цьому потрібно вирішити наступні питання:

1) знайти онтологію, клас якої відображає структуру ІО, знання про які необхідні користувачу;

2) знайти множину ІР, в яких містяться відомості щодо значень властивостей цих ІО;

3) здобуття знань про ІО з ІР;

4) представлення здобутих знань у формі, зрозумілій і зручній для користувачу.

Слід зазначити, що в багатьох випадках цей процес є ітеративним, і при повторному рішенні проблеми інформацію про ІО необхідно обновляти, витягаючи необхідні відомості з тих ІР, що доступні користувачу (наприклад, через Web чи корпоративну мережу).

Складну структуру ІО користувач може описати у вигляді класу відповідної онтології. Приклади ІО – організація, навчальний заклад, людина-експерт, Web-сервіс, аналітична модель бізнесу-процесу.

### **Класифікація ІО та пов'язаних з ними пошукових ситуацій**

Прирішенні задачі семантичного пошуку, пов'язаного з розпізнаванням набору складних ІО, виникає ряд різних видів проблем, для опису яких необхідно визначити використовувати терміни, зокрема, визначити, які відомості є результатом такого пошуку, які – його умовами.

У найбільш простому варіанті інформаційного пошуку на вході пошукова система отримує набір ключових слів, а на виході дає набір посилань на документи.

Значно складніше вирішити пошукову задачу, для якої вхідними даними служить опис складної проблеми, у якій описана взаємодія множини складно структурованих ІО, а на виході необхідно отримати посилання на екземпляри ІО, що задовольняють поставленим умовам.

*Онтологія*  $IOO_{IO}$  – онтологічна структура, що містить клас  $IO$   $t_{IO} \in T_{IO}$  і його підкласи, які описують різні підмножини ІО, а також класи  $T_{Prop}$ , необхідні для опису властивостей різних ІО:  $O_{IO} = \langle T_{IO} \cup T_{Prop}, R, A \rangle$ . Для того, щоб користувач міг охарактеризувати цікавлячий його ІО, йому потрібно послатися на клас довільної формально описаної онтології. Таким чином,  $IO$  – це клас онтології ІО, що має набір характеристик, які описують його структуру і можливі зв'язки з іншими класами й екземплярами класів.

*Екземпляр ІО* – екземпляр якого-небудь підкласу ІО відповідної онтології, який можна однозначно ідентифікувати і який має власне ім'я.

*Ситуація* – непорожня множина екземплярів ІО і ІО одного чи різних класів, така, що для будь-якого ІО з цієї множини існує зв'язок хоча б з одним іншим ІО з цієї множини.

Якщо в ситуації використовуються ІО, описані за допомогою різних онтологій, то необхідно (явно чи за допомогою засобів автоматизованого зіставлення онтологій) встановити зв'язки між цими онтологіями (чи хоча б між тими ІО і класами, що описують властивості тих ІО, що фігурують у ситуації).

*Схема ситуації* – ситуація, у якій не використовуються екземпляри ІО.

*Неприпустима схема ситуації* – схема ситуації, всі умови якої не можуть бути виконані при жодному наборі екземплярів ІО.

Схема є *неприпустимою*, якщо в ній містяться умови суперечні умови:  $f_0(a_1, \dots, a_n), f_1(a_1, \dots, a_n), \dots, f_m(a_1, \dots, a_n)$ ,  $a_i \in t_i \subseteq T_{IO}$ ,

і з деякої їхньої підмножини можна вивести заперечення однієї з цих умов, тобто з  $f_1(a_1, \dots, a_n), \dots, f_m(a_1, \dots, a_n), a_i \in t_i \subseteq T_{IO}$  логічно виводиться  $\neg f_0(a_1, \dots, a_n)$ .

*Унікальна ситуація* – ситуація, всі умови якої можуть бути виконані лише при єдиному наборі екземплярів ІО. Приклад такої ситуації – пошук книги, посилання на яку міститься у відеофільмі.

*Конкретизована ситуація* – ситуація, в описі якої, крім класів ІО, описано хоча б один конкретний екземпляр ІО. Приклад такої ситуації – знайти всі організації, в яких працювати особи, які проживали в одному будинку з особою X, що має ідентифікаційний номер Y.

*Персональна ситуація* – ситуація, в якій використовується екземпляр класу “користувач” онтології семантичного пошуку, який характеризує ту особу, що задає цю ситуацію. Цей варіант задачі досить поширений, коли користувач намагається знайти якісь відомості саме для себе – приміром, посилання на власні публікації, можливість свого працевлаштування в певній організації, рейтинг своєї спеціальності тощо. кожна персональна ситуація є конкретизованою через використання конкретного екземпляра класу “користувач”. Використання персональних ситуацій дозволяє розробляти типові запити, в яких певна частина інформації не вводиться користувачем вручну, а імпортується з його профілю. Приміром, можна задавати замість запиту “знайти в публікаціях автора А всі посилання на публікації автору В” задавати значно простіший для користувача запит “знайти публікаціях автора А всі посилання на мої публікації”, для якого перелік “мої публікації” має будуватися автоматично та оновлюватися за рахунок пошуку у Web-ресурсах.

Ситуація *задовольняє схемі*, якщо для всіх ІО її екземплярів ІО виконуються умови, що входять до складу схеми.

Пошук *нездійснений*, якщо його умовою є неприпустима схема ситуації.

Пошук *виконуваний*, якщо його умови можуть бути виконані (навіть якщо не виявлена така комбінація ІО, що задовольняє цим умовам).

Пошук *тривіальний*, якщо його результатом є унікальна ситуація.

В інших термінах можна вважати схему ситуації пошуковим запитом, а множину ситуацій – його результатом.

Деякі з таких пошукових задач, пов’язаних з розпізнаванням ситуацій, дозволяють вирішувати запити до семантичних Wiki-ресурсів шляхом виконання відповідного SPARQL-запиту. На жаль, лише відносно невелика частина ресурсів Web має семантичну Wiki-розмітку. Крім того, ІО, які можна використовувати в таких запитах, обмежені специфікою цих Wiki-ресурсів та тими тегами, що в них використовуються – “людина”, “географічний об’єкт”, “дата” тощо (ці теги безпосередньо пов’язані з класами відповідної онтології). Більше інформації доступно у вигляді несемантизованих Wiki-ресурсів, але в цьому випадку для опису ІО можна

використовувати лише обмежену множину тегів розмітки, таких як “стаття”.

Таким чином, можна певним чином розглядати семантичний пошук (як складову частину рішення ІС, призначеної для розв’язання якоїсь проблеми) як задачу, *зворотну* пошуку Web-сервісів. При пошуку Web-сервісів ми маємо структуру вхідних і вихідних параметрів і набір зв’язаних з ними умов, а також опис цього перетворення, а намагаємося знайти відповідний їм сервіс. Передбачається, що конкретні значення вхідних параметрів відомі користувачу, що шукає Web-сервіс. Наприклад, користувач хоче знайти сервіс, який за температурою повітря обчислює час закипання чайника. При цьому вхідні дані – температура повітря – йому відомі.

При семантичному пошуку ми також маємо структуру вхідних і вихідних параметрів і набір зв’язаних з ними умов, а намагаємося знайти конкретні значення вхідних і вихідних параметрів. При цьому передбачається, що сам сервіс (тобто деякий алгоритм обробки даних щодо ситуації, що приводить до рішення задачі, яка стоїть перед користувачем,) уже доступний користувачу. Наприклад, користувач має деяку ІС, здатну організувати подачу та рецензування наукових статей через Web-середовище. Для цього необхідно підбирати експертів для оцінки наукових статей, для чого до складу ІС якої входить система семантичного пошуку і зіставлення компетенцій експертів і тематики статей. Такий пошук має забезпечувати здобуття з ресурсів Web інформації щодо компетенцій потенційних експертів саме в ПрО конкретної статті, яка задається через структуру відповідної онтології. При цьому відомості про структуру шуканих даних і обмеження задаються як вхідні параметри пошуку, одержувана при пошуку інформація про екземпляри – конкретних людей експертах – це вхідні параметри Web-сервісу, а результати роботи ІС можна розглядати як його вихідні параметри.

Розпізнавання ситуацій, в яких використовується кілька ІО одного або різних класів зі складною структурою, є у загальному випадку надто комплексною задачею для пошукових систем, навіть для тих, що здійснюють пошук з урахуванням семантики. Але для деяких окремих випадків вже зараз існують досить високо розвинуті методи та засоби такого пошуку, що працюють з урахуванням специфіки окремих видів ІО. Такі

засоби розробляються для тих ІО, які найбільш активно використовуються та для яких автоматизація їх поєднання для певних ситуацій є необхідною умовою їх ефективного застосування. Найбільш відомі приклади такого пошуку – це композиція семантичних Web-сервісів, коли для рішення задачі користувача необхідно знайти набір сервісів, що реалізують різні підзадачі, та визначити порядок їх виконання; та задача комплексного пошуку мультимедійних даних, коли за інформацією, що здобувається з одного мультимедійного ІО, потрібно знайти інший ІО, що відповідає певним умовам (приміром, знайти всі інші фільми режисера вказаного фільму).

Більшість сучасних інформаційно-пошукових систем орієнтована на пошук однотипних інформаційних об'єктів та не придатна для врахування специфіки властивостей їх підкласів. Якщо ж користувачу потрібно знайти кілька різних інформаційних об'єктів, що знаходяться один з одним у різних відношеннях, то йому потрібно виконувати ці порівняння самостійно. Виникає потреба у розробці формального апарату для семантичного пошуку, результатом якого є певна ситуація, що поєднує кілька інформаційних об'єктів. Певні досягнення у цьому напрямку можна відмітити зараз лише для окремих випадків – інформаційних об'єктів з відносно чітко формалізованою структурою та семантикою, таких, як Web-сервіси та мультимедійні дані.

Наприклад, онтологія OWL-S забезпечує механізм для моделювання бізнес-процесів, але відрізняється від нього виразністю термінів, уявлень, семантики, підтримки пошуку і виконання, обробки помилок. Опис OWL-S для сервісу складається з профілю сервісу, моделі сервісу та обґрунтування сервісу, тобто пояснення того, що виконує цей сервіс, як він працює, як можна дістати до нього доступ. Саме ця структура може використовуватися як основа для інтелектуального аналізу відомостей щодо сервісів та застосування методів Data Mining з метою автоматизованої компоновки сервісів відповідно до задач, сформульованих користувачами.

Досліджуючи такі приклади, можна намагатися узагальнити їх для більш широкого класу інформаційних об'єктів.

### **Семантичний пошук компетенцій як інформаційна підтримка науково-дослідного та освітнього процесу**

Модель семантичного пошуку досить легко адаптувати для рішення різних прикладних задач, у тому числі і до тих, що пов'язані з пошуком складних ситуацій, в умовах яких використовуються ІО різних класів. Розглянемо це на прикладі проблеми зіставлення компетенцій [7], що є складовою частиною таких поширених на практиці задач, як пошук роботодавцем придатних виконавців робіт; вибір експертів в нових Про; оцінка перспективи успішного виконання проекту; порівняння кваліфікації фахівців, що мають різні спеціальності (зокрема, що відповідають нормативам різних країн); вибір навчального закладу, що пропонує необхідний йому набір дисциплін; оцінка можливості переходу студента з одного навчального закладу до іншого. Виконання таких задач потребує використання знань про запропоновані роботи, навчальні заклади та навички потенційний виконавців робіт.

Ці задачі є окремими випадками пошуку ситуації. Приміром, для вибору абітурієнтом навчального закладу потрібно порівнювати екземпляри класу “навчальний заклад” з екземплярами класу “дисципліна”, тобто застосовуються ІО двох різних класів. Обидва ці класи містять властивості класу “компетенція”: для першого класу це властивість “надає в процесі навчання”, для другого – властивість “включає”. Саме за цими параметрами й проводиться порівняння. Через те, що результати порівняння можуть не співпадати повністю, то потрібно враховувати вагу різних дисциплін для абітурієнта. Використання класу “компетенція” дозволяє порівнювати дисципліни та навчальні заклади не на рівні назв, а на рівні семантики того, що в них викладається.

Слід розрізняти *компетенцію* та *компетентність*. *Компетенція* – це поняття, загалом не пов'язане з конкретною особою; *компетентність* – це відношення між особою і компетенцією, яке означає, що певна людина володіє цією компетенцією. Відповідно, до поняття компетенції належать її найменування, категорія та опис, а свідчення компетенції має підтверджувати наявність відношення *компетентності*. *Компетентність* – це рівень досягнень (досвіду, знань, навичок) особи в певній Про. Компетентність може бути визначена на основі

аналізу діяльності фахівця, рівня й широти обізнаності з досягненнями науки й техніки, розуміння досліджуваних проблем, можливих шляхів їхнього розвитку. Для кількісної оцінки рівня компетентності використовується *коефіцієнт компетентності*, з урахуванням якого зважуються висновки експерта.

Останнім часом, завдяки комплексним дослідженням з психології, менеджменту та теорії навчання, набули поширення такі теорії керування персоналом, як *керування компетенціями* (competence management) і *керування талантами* (talent management).

У загальному значенні *компетенція* – здатність успішно діяти на основі наявних знань і практичного досвіду під час розв’язання задач. Елементи компетенції – знання й навички, життєвий досвід, фізичний потенціал, здібності, риси характеру, інтелект тощо – об’єднані в різні конфігурації для розв’язання людиною конкретних проблем. У керуванні персоналом під компетенцією, зазвичай, розуміють формальний опис вимог до особистісних, професійних та інших якостей співробітника. На сьогодні поняття «компетенція» є нечітким і дуже залежить від специфіки конкретної Про. Поняття компетентності передбачає сукупність фізичних, інтелектуальних і психологічних (соціальних) якостей і властивостей людини, необхідних їй для самостійного або спільного з іншими ефективного виходу з різних професійних і життєвих ситуацій. Уперше вивчення компетенцій для прогнозування рівня ефективності виконання роботи було запропоновано американським психологом Д. МакКлеландом [8], який доводив, що традиційні тестування здібностей і знань, а також наявність яких-небудь дипломів у потенційних виконавців робіт не забезпечують ні ефективного виконання таких робіт, ні успіхів у професійній діяльності. Подібні висновки змусили Д. МакКлеланда розпочати пошук таких характеристик – *компетенцій*, що дають змогу віднайти конкретних виконавців, здатних до виконання певних видів робіт. Для цього він вивчав характеристики успішних виконавців робіт і порівнював їх з характеристиками менш успішних виконавців тієї ж роботи. Зараз є багато робіт з автоматичного визначення

на основі документів [9]. Усі вони використовують текстовий пошук для знаходження в документі свідчень щодо компетенції його автора, якими є слова й словосполучення. Свідченням володіння компетенцією є згадування слів запиту в текстах документів, автором яких був користувач.

Задача оцінки можливості переходу студента з одного навчального закладу в інше є прикладом персональної ситуації: в порівнянні приймають участь дисципліни різних навчальних закладів з наведеного користувачем переліку або ті, що задовольняють певній умові (приміром, ті, що знаходиться у певному місті). В цій ситуації застосовуються ІО різних класів – “дисципліна”, “навчальний заклад”, “компетенція” та “користувач”. Слід відмітити, що кожна персоніфікована ситуація є конкретизованою: в ній застосовується конкретний екземпляр, пов’язаний з користувачем системи. Вони порівнюються з тими дисциплінами, які вже вивчив користувач, що надає запит. При цьому йому не потрібно вводити перелік того, що він вивчав, – ці дані можуть здобуватися з його персонального профілю в цій системі. Порівняння теж виконується за допомогою властивостей класу “дисципліна”, що містять екземпляри класу “компетенція”. Таким чином, в такому порівнянні теж враховується семантика оброблюваних ІО. Рішення всіх цих задач може базуватися на спільній онтологічній моделі, в якій встановлюються основні поняття, пов’язані з компетенціями, їх структура та відношення між екземплярами класів, що відповідають цим поняттям. Вона містить такі основні класи, “компетенція”, “дисципліна”, “особа”, “учбовий заклад” тощо та їх підкласи. Крім того, така модель має містити класи, що відповідають інформаційним об’єктам, що є результатом пошуку (рис.2). Приміром, у задачі пошуку виконавців робіт ІО відноситься до класу “особа”, а у задачі оцінки навчальних закладів – до класу “організація”, хоча може характеризуватися однаковими наборами властивостей з класу “компетенція”. Можливість вибору типу та структури шуканого ІО значно підвищує пертинентність пошуку та робить його результати значно більш придатними для користувачів.



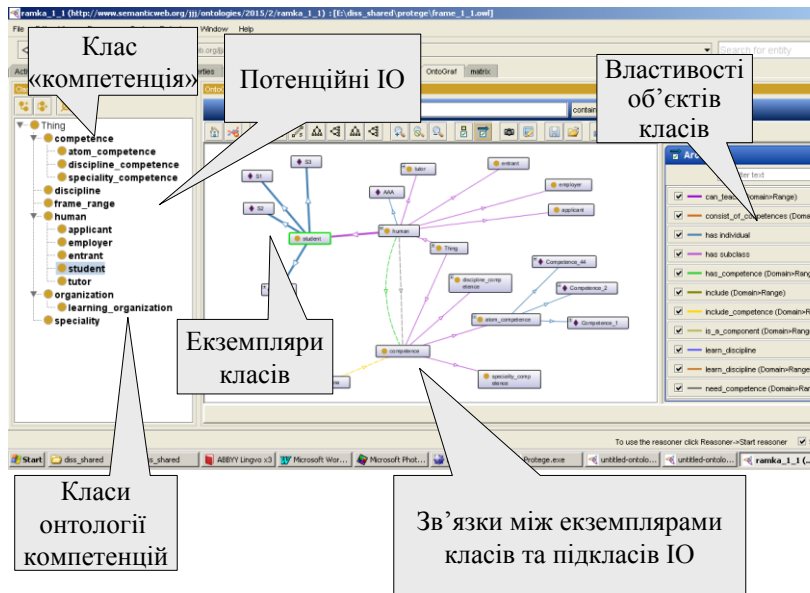


Рис.2. Онтологічна модель задачі аналізу компетенцій

Така онтологічна модель семантичного пошуку надає прикладним системам актуальні знання щодо аналізу компетенцій, які дозволяють виконувати логічне виведення та підтримувати складні інформаційні запити. Важливою рисою цієї моделі є можливість імпортувати персональні особливості користувачів зі складними і стабільними інформаційними потребами, характерними для науково-дослідної діяльності й освітньої сфери, враховуючи їх освіту, досвід та компетентність у різних сферах діяльності, до інших інтелектуальних застосувань, що базуються на онтологіях.

Користувач описує свою інформаційну потребу, указуючи клас онтології, до якого відноситься шуканий ІО, і умови, що накладаються на його. Таким ІО може бути організація, людина, спеціальність і т.д. умови можуть бути описані як множина екземплярів класу «атомарна компетенція». Таким чином, запити зводяться до зіставлення значень властивостей одного класу для екземплярів різних класів. Наприклад, зіставлення властивостей «включає компетенції», що відноситься до класу «атомарна компетенція», для екземпляра класу «фахівець» і для екземпляра класу «спеціальність».

При цьому явно вказується семантика інформаційної потреби, і тому можна здійснювати пошук, що диференціює різні відношення між шуканим ІО і набором компетенцій. Наприклад, з тим самим екземпляром класу «людина» деякі різні підмножини атомарних компетенцій можуть бути зв'язані відносинами «володіє», «має

сертифікат», «може викладати», «має досвід використання». Це дозволяє точніше задовольняти інформаційну потребу користувача, знаходячи ІО, що відповідають його вимогам.

Для того, щоб виконувати співставлення ІО різних типів – приміром, спеціальностей, умінь і компетенцій людей і організацій, потрібно знайти їх спільні параметри, тобто властивості екземплярів цих класів, що відносяться до одного класу. Аналіз досліджень в цій сфері вказує на доцільність використання для цього значень класу «атомарна компетенція», що є підкласом класу «компетенція». Таким чином, екземпляр кожного з наведених класів характеризує з точки зору даної ПрО набір еталонних атомарних компетенцій [10].

Екземпляр вважається атомарним, якщо жоден інший екземпляр цього класу не є його підмножиною, тобто екземпляри класу «атомарна компетенція» не перетинаються.

Таке визначення забезпечує узагальнений механізм для побудови атомарних компетенцій в обраній ПрО за множиною компетенцій, які характеризують цю ПрО та можуть бути здобути з відповідних нормативних документів – описів спеціальностей, дисциплін тощо. Наприклад, якщо дві компетенції А та В перетинаються, те по ними будується 3 потенційно атомарні компетенції А1, В1 і В, такі, що  $A \cap B = C$ ,  $A_1 \cup C = A$ ,  $B_1 \cup C = B$

Тому доцільно поповнити онтологічну модель ПрО додатковим службовим класом. Клас «Атомарна компетенція» – підклас класу «компетенція», такий, що для  $\forall a \in$  «Атомарна компетенція» існує хоча б один елемент класу

$\forall b \in$  "Компетенція", такий, що  $a \subseteq b$ , але для жодного елемента з класу "атомарна компетенція" не існує іншого елемента цього класу  $c \in$  "Атомарна компетенція", такого, що  $c \subseteq a, a \not\subseteq c$ . Клас "Атомарна компетенція" має властивість "входити до складу" класу "дисципліна" і властивість "входити в" клас "компетенція". Найбільш важливі проблеми, що виникають у процесі рішення такої задачі, зв'язані з формуванням множини атомарних компетенцій, що вимагає значних інтелектуальних зусиль експертів і може бути автоматизовано тільки частково, і з поповненням бази знань відомостями про екземпляри інформаційних об'єктів, що вимагають і постійного відновлення великого обсягу даних.

З урахуванням того, що сучасні інтелектуальні застосування, як правило, функціонують у відкритому інформаційному просторі, як джерела знань про екземпляри інформаційних об'єктів доцільно використовувати інформаційні ресурси Web:

- різні державні і міжнародні стандарти і нормативи, зв'язані з описом структури і рівня
- офіційні сайти й організаційні онтології навчальних закладів, що містять інформацію про спеціальності, що викладаються, і дисципліна, які входять до їх складу;
- зовнішні оцінки різних навчальних закладів і організацій, що дозволяють оцінити якість одержуваних у них компетенцій;
- персональні відомості про людей, що витягаються із соціальних мереж, Wiki-ресурсів і різних природно-мовних документів,

доступних через Web (наукових публікацій, технічних звітів, методичних матеріалів і т.д.).

### Структура онтології компетенцій.

Онтологія компетенцій визначає семантичні властивості і відношення головних інформаційних об'єктів, що стосуються кваліфікації осіб, навчальних організацій та потреб роботодавців.

Ми пропонуємо використовувати в якості основного елемента компетенцію  $c \in C$ . Компетенції підрозділяються на атомарні  $C_{atomic}$  і складні  $C_{complex}$ ,  $C = C_{atomic} \cup C_{complex}$ .

$c \in C_{complex}$ , якщо  $\exists c', c' \neq c, c' \in C, c' \subseteq c$ .

Атомарні і комплексні компетенції є класами онтології компетенцій. Обидва є підкласами класу "компетенція". Інші важливі класи цієї онтології – Дисципліна; Спеціальність; Людина; Організація.

Все підкласи класу мають деякі загальні властивості. Наприклад, всі підкласи "особи" мають властивості "ім'я", "рік народження", "адреса" тощо. Ці класи робляться більш конкретними за допомогою підкласів і різних семантичних властивостей об'єкта.

Наприклад, клас "людина" має підкласи "студент", "роботодавець", "викладач", "дослідник", "аспірант" тощо. Такі підкласи відрізняються деякими властивостями: "студент" має властивості "місце навчання", "спеціальність", "рік", а "аспірант" має властивості "Науковий керівник", "тема науково-дослідницької роботи" (рис.3).

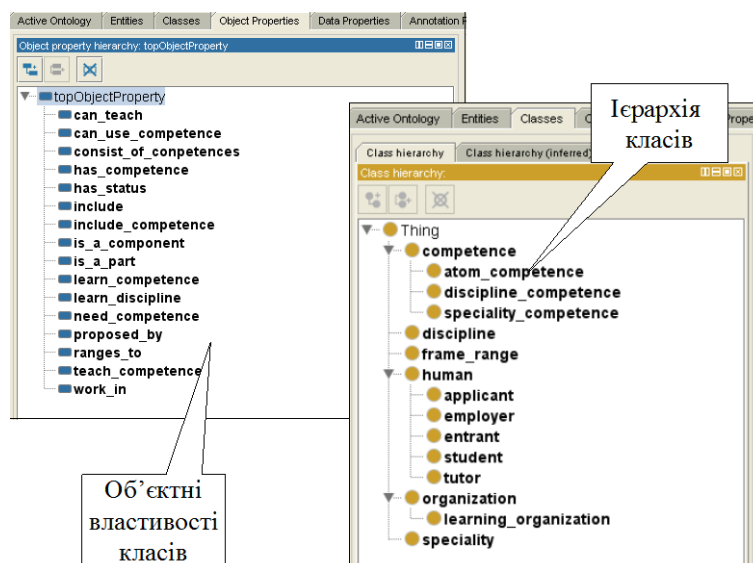


Рис.3. Класи і об'єктні властивості в онтології компетенцій

Важливою характеристикою запропонованого підходу є той факт, що всі основні класи мають семантичні властивості об'єкта зі значенням з класу "компетенція", які визначають їх семантичні аспекти, що стосуються аналізу компетенції.

Приміром, 8 рівнів європейського кваліфікаційного стандарту EQF можна формалізувати як підкласи класу "кваліфікація" зі значенням властивості даних "рівень". Кожен екземпляр класу «кваліфікація», який має властивість даних "кваліфікаційна система", що дорівнює "EQF", обов'язкове має унікальне значення властивості даних "рівень" від 1 до 8 і три непусти множини об'єктних властивостей "Знання", "Навички" і "Комунікації" зі значеннями з класу "компетентності".

Найпростіша модель кваліфікацій на основі онтології компетенцій може бути формально представлена як трійка

$$q \in Q = \left\langle \begin{array}{l} Iq, Lq, Compet = \\ = Kn \cup Sk \cup Com \cup \dots \cup Compet_p \end{array} \right\rangle, \text{ де } p = \overline{0, r}$$

- $iq_j \in Iq, j = \overline{1, n}$  – ідентифікатор системи кваліфікацій;

- $Lq = \bigcup_{j=1}^n \{lq_{i_1}, \dots, lq_{i_{s_j}}\}$ , де  $lq_{i_{s_j}}$  –

кількість різних рівнів в системі класифікацій;

- $Kn$  – сукупність атомарних компетенцій, які характеризують знання відповідної кваліфікації;

- $Sk$  – набір атомарних компетенцій, що характеризує навички відповідної кваліфікації;

- $Com$  – набір атомарних компетенцій, які характеризують комунікації відповідної кваліфікації;

- $Compet_p$  – набір атомарних компетенцій, що характеризує р-й набір відповідної кваліфікації (ми не конкретизуємо критерії побудови цих множин, які мають справу з особливостями різних національних і міжнародних кваліфікаційних систем).

Різні набори  $Compet_p$  можуть бути використані в різних кваліфікаційних системах, але ми вважаємо дві кваліфікації  $A \in L$  та  $B \in L$  еквівалентними, якщо їх набори компетенцій однакові  $A \in L \equiv B \in L \Leftrightarrow Compet_A \equiv Compet_B$ .

Якщо деякі ідентичні кваліфікації мають різні набори компетенцій в різних

кваліфікаційних систем, то ми повинні уточнити множин атомарних компетенцій або правила переходу від складних компетенцій до набору атомарних одиниць.

Спеціальності і дисципліни моделюються аналогічним чином.

Модель спеціальностей на основі компетентності онтологія може бути формально представлена трійкою

$$s \in Sp = \left\langle \begin{array}{l} Is, Ls, Compet = \\ = Compet_1 \cup \dots \cup Compet_m \end{array} \right\rangle, \text{ де}$$

- $is_j \in Is, j = \overline{1, n}$  – ідентифікатор системи класифікації спеціальностей;

- $Ls = \bigcup_{j=1}^n \{ls_{i_1}, \dots, ls_{i_{s_j}}\}$ , де  $ls_{i_{s_j}}$  –

кількість рівнів в системі класифікації спеціальностей  $is_j$ ;

- $Compet$  – набір атомарних компетенцій, які характеризують відповідні компетенції спеціальностей.

Формальна модель дисциплін на основі компетенції онтології може бути формально представлена трійкою

$$d \in Disc = \left\langle \begin{array}{l} Id, Ld, Compet \\ = Compet_1 \cup \dots \cup Compet_m \end{array} \right\rangle, \text{ де}$$

- $id_j \in Id, j = \overline{1, n}$  – ідентифікатор системи кваліфікації;

- $Ld = \bigcup_{j=1}^n \{l_{i_1}, \dots, l_{i_{s_j}}\}$ , де  $l_{i_{s_j}}$  – кількість

рівнів в класифікаційній системі дисциплін  $id_j$ ;

- $Compet$  – це набір атомарних компетенцій, які характеризують компетенції дисциплін.

Аналогічним чином інші поняття, що стосуються компетенцій, можуть бути змодельовані на основі класів пропонованої онтології компетенцій.

Наприклад, формальна модель "роботодавець", який є підкласом класу "особа", можна уявити як елемент  $Emp$ ,  $e \in Emp = \langle Ie, Name_e, Adress_e, Type_e, Country_e, CE = \{Compet_1, \dots, Compet_x\} \rangle$ , де

- $ie_j \in Ie, j = \overline{1, n}$  – ідентифікатор роботодавця;

- $ne_j \in Name_e, j = \overline{1, n}$  – найменування роботодавця;

- $ae_j \in \text{address}_e, j = \overline{1, n}$  - адреса роботодавця;
- $CE$  – набір множин компетентності, де  $Compet_y$  є множиною компетентностей, що роботодавець вимагає від  $y$ -го співробітника.

Відношення між цими поняттями візуалізовано на рис.4.

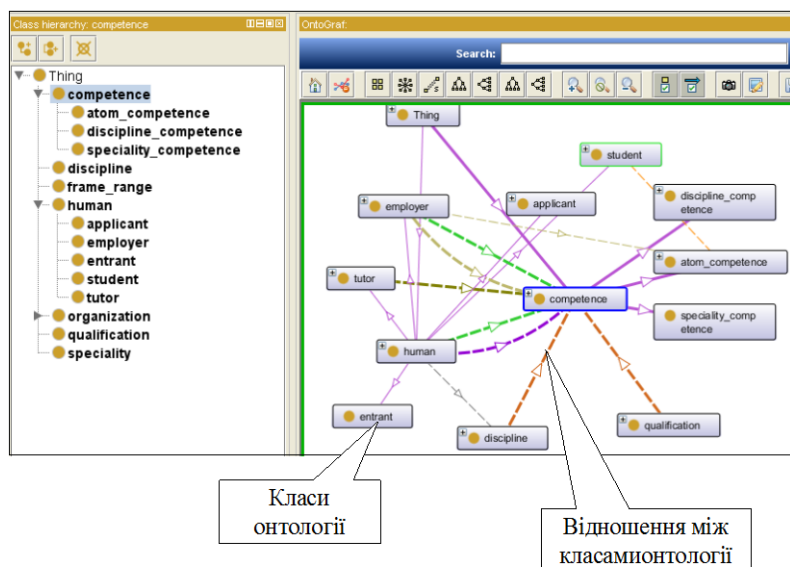


Рис.4. Відносини понять компетентності онтологій

### Висновки

Використання формалізованої онтологічної моделі для подання знань як щодо структури складних інформаційних об'єктів, та і щодо предметної області, що цікавить користувача, та тієї проблеми, для рішення якої використовується ІС, дозволяє більш ефективно здійснювати пошук інформації у відкритому середовищі та відбирати найбільш пертинентні відомості. враховувати індивідуальні інформаційні потреби користувачів та пояснювати йому результати, отримані в процесі семантичного пошуку в Web.

### Список використаних джерел

1. Nobles R. The Future Of Search Engine Optimizing: Theme Engines. the next generation of search engines has arrived. [Електронний ресурс] – Режим доступу: <http://www.searchengineworkshops.com/articles/se-optimization-future.html>. Lawrence S. Context in the Web Search. – <http://citeser.nj.nec.com/lawrence00context.html/>.
2. Россеева О.И., Загорюлько Ю.А. Организация эффективного поиска на основе онтологий. [Электронный ресурс] – Режим

доступа: [http://www.dialog-21.ru/archive\\_article.asp?param=7029](http://www.dialog-21.ru/archive_article.asp?param=7029).

3. Gladun A., Rogushina J., Andrushevich A., Kurbatski A. User-Oriented Recognition of Intelligent Information Objects in Distributed Dynamic Informational Web-Space // Proc. of the 12th International Conference on Pattern Recognition and Information Processing (PRIP-2014) Minsk, Belarus, 2014. – P. 70–75.

4. Рогушина Ю. В. Семантичний пошук у Web на основі онтологій: розробка моделей, засобів і методів / Ю. В. Рогушина. – Мелітополь: МДУПУ ім. Богдана Хмельницького, 2015. – 291 с.

5. Гладун А.Я., Рогушина Ю.В. Формирование та применение онтологий предметных областей для поиска Web-сервисов на семантическом уровне // Труды Всероссийской конференции с международным участием "Знания-Онтологии-Теории", Т.2, 2007. – С.176–185.

6. De Campos L. M., Fernández-Luna J. M., Huete J. F., Vicente-Lopez E. Using personalization to improve XML retrieval // Knowledge and Data Engineering, IEEE Transactions on, 26(5), 2014. – P.1280–1292.

7. Harzallah B. Knowledge management for competence management // Proc. of International Conference on Knowledge Management and Knowledge Technologies I-KNOW'05, 2005. – P.449–460.

8. Армстронг М. Практика управления человеческими ресурсами. – 8-е изд. / Пер. с англ. под ред. С. К. Мордовина. – СПб.: Питер, 2008. – 832 с.

9. Macdonald C. Voting for candidates: adapting data fusion techniques for an expert

search task / C. Macdonald, I. Ounis // CIKM 2006: Proc. of the 15th ACM International Conference on Information and Knowledge Management. ACM, New York, 2006. – P. 387–396.

10. Рогушина Ю.В., Гладун А.Я., Осадчий В.В., Прийма С.М. Онтологічний аналіз у Web. Монографія. – Мелітополь: МДУПУ ім. Богдана Хмельницького, 2015. – 407 с.

#### Відомості про автора:



**Рогушина Юлія Віталіївна** – кандидат фізико-математичних наук, старший науковий співробітник Інституту програмних систем Національної Академії Наук України. Наукові інтереси: онтологічний аналіз на базі технологій Semantic Web, інтелектуальний інформаційний пошук, методи індуктивного вилучення знання, дослідження інтелектуальних інформаційних систем і поведінку програмних агентів.

**E-mail:** [ladamandraka2010@gmail.com](mailto:ladamandraka2010@gmail.com)