

## КЛАСИФІКАЦІЯ ЗАСОБІВ ТА МЕТОДІВ СЕМАНТИЧНОГО ПОШУКУ В WEB

Розглянуто проблеми, пов'язані з удосконаленням пошуку інформації у відкритому середовищі, обґрунтована потреба в його семантизації. Проаналізовано сучасний стан та перспективи розвитку систем семантичного пошуку, орієнтованих на обробку інформаційних ресурсів Web, розглянуто критерії класифікації таких систем. В цьому аналізі значна увага приділяється використанню у семантичному пошуку онтологій, що містять знання щодо предметної області пошуку та користувача, для якого виконується пошук.

На основі аналізу властивостей існуючих систем семантичного пошуку з точки зору цих критеріїв виділені області подальшого вдосконалення цих систем, запропоновано їх реалізацію у системі семантичного пошуку "МАПС".

Ключові слова: семантичний пошук, онтологія, Semantic Web, персоніфікація пошуку.

### Вступ

Сьогодні Web забезпечує доступ до значної частки інформаційних ресурсів людства. Обсяг такої інформації постійно збільшується, а її структура стає все складнішою та більш гетерогенною.

Пошук у Web – це ключова технологія, тому що використання *інформаційно-пошукових систем* (ІПС) є основним способом отримати доступ до вмісту його ресурсів і забезпечити їх ефективно використання. Це зумовлює потребу в розвитку засобів інформаційного пошуку, що враховують специфіку відкритого інформаційного середовища та здатні забезпечити специфічні інформаційні потреби окремих користувачів, позбавивши їх від рутинного аналізу тих *інформаційних ресурсів* (ІР), які не стосуються їх поточних інтересів.

Традиційно *інформаційні потреби* (ІП) користувачів формалізуються за допомогою *пошукового запиту* (ПЗ), на основі якого будується *пошуковий образ запиту* (ПОЗ). Такий ПОЗ співставляється з *пошуковими образами документів* (ПОД), доступ до контенту яких має система, що здійснює інформаційний пошук. Першими ПОД були дані бібліотечних, що містили назви книг та імена їх авторів. Класичними моделями інформаційного пошуку вважають булеві, імовірнісні, векторні та дескрипторні моделі, на основі яких виконується співставлення ПОД та ПОЗ. Яку саме інформацію про документи відображають ПОЗ та ПОД і як саме вони співставляються

ся, залежить від конкретної реалізації пошукової системи.

Зазвичай пошуковий запит складається лише з кількох слів, але нині цього стає недостатньо, тому що кожному такому запиту релевантна надто велика кількість ІР, і виникає потреба в застосуванні додаткової інформації щодо інформаційної потреби користувача як для самого пошуку, так і для ранжування його результатів.

Інформаційні потреби залежать від персональних властивостей користувача, тобто за тим самим запитом різні користувачі прагнуть б отримати різну інформацію (приміром, запит "прогноз погоди" є релевантним різноманітним синоптичним сайтам, але у більшості випадків користувача цікавить погода у тому місті, де він знаходиться). Можна ідентифікувати певні ІП як одноразові (у більшості випадків – довідкова інформація або однозначно ідентифікований інформаційний об'єкт) або постійні (приміром, пов'язані з професійною діяльністю або хобі, коли користувач зацікавлений у постійному отриманні нових відомостей з певної тематики), але більшість з них мають певну скінчену довжину у часі і перестають бути актуальними через певний час – користувач знаходить відомості, що йому потрібні, або змушений діяти без цих відомостей. Крім того, ІП безпосередньо пов'язані з тією поточною задачею, яку вирішує користувач, тобто пертинентність відповіді може

залежати від часу доби, дня тижня та інших додаткових ознак (приміром, у робочі дні та у вихідні або свята користувача цікавлять різні проблеми). Такі відомості щодо своїх ІІ може безпосередньо вводити сам користувач, але це потребує багато зусиль та часу. Таким чином, виникає питання щодо того, яку саме інформацію доцільно враховувати в інформаційному пошуку, звідки її отримувати та як саме формалізувати.

Основна тенденція розвитку семантичних пошукових засобів, які використовуються в найрізноманітніших сферах, пов'язана з переходом від виявлення документів, що містять певні ключові слова, до пошуку знань, необхідних для виконання поставленого перед користувачем завдання. Використання пошуку в Web, який в основному базується на комбінації пошуку тексту за ключовими словами з обчисленням ранжування важливості документів у залежності від структури посилань у мережі, має багато обмежень, і тому існує велика кількість науково-дослідних проєктів, спрямованих у бік більш інтелектуальних форм отримання інформації, тобто *семантичного пошуку* [1, 2], під яким надалі будемо розуміти такий пошук інформації, що задовольняє ІІ користувача в процесі розв'язання певної проблеми, в процесі якого застосовуються (наочно або приховано від користувача) знання щодо різних суб'єктів і об'єктів пошукової процедури й методи аналізу цих знань [3]. Ці знання можуть стосуватися як користувача та його інформаційних потреб (персоніфікація пошуку), так і ІР, серед яких здійснюється пошукова процедура.

Такий пошук, як правило, також ставить ціль забезпечити виконання більш складних пошукових запитів, оцінки яких потребують міркувань через Web. Інше поширене використання поняття семантичного пошуку в Web є пошук у великих наборах даних у Semantic Web, який у майбутньому замінить нинішній Web. Це використання тісно пов'язане з першим, тому що семантичне анотування Web-ресурсів та здобуття знань з Web-ресурсів, пов'язані з побудовою бази

знань, яка може бути реалізована з використанням технології Semantic Web. Таким чином, семантичний пошук у Web можна аналізувати як розвиток існуючих пошукових засобів на основі нових семантичних технологій, орієнтованих на Web-застосування.

Ще один напрямок досліджень, тісно пов'язаний з семантичним пошуком, – це пошук у Web, де ІІ формулюються як *природномовні вирази*. Існує багато підходів до перетворення таких запитів у формалізовані структуровані запити, які можуть виконуватися в різноманітних системах семантичного пошуку. Відповідями на ці запити можуть бути як звичайні Web-ресурси, так і структурована або природномовна інформація (приміром, фрагменти природномовних ІР або структурована інформація з Вікіпедії).

Часто поняття семантичного пошуку пов'язують також з деякими іншими семантичними концепціями. Наприклад, *фасетний* пошук дозволяє вивчати результати відповідно з колекцією попередньо визначених категорій – фасет. У тісному зв'язку з семантичним пошуком знаходиться й кластерний пошук, де такі фасети можуть бути не тільки заздалегідь визначеними. Ще одним прикладом є надання рекомендацій щодо пов'язаних пошукових запитів, таких як продовження і корекції пошукових запитів, які добре відомі також у не семантичних ІІС. Також до проблеми семантичного пошуку стосується визначення ступеню повнотекстової подоби, де як запит використовуються не окремі ключові слова, а блоки тексту, починаючи від окремих фрази і до цілого документу. Досить близькою задачею є *онтологічний пошук подібності* (наприклад, [4]), що базується на подібності між онтологічними сутностями.

## Постановка задачі

Ефективне застосування інформаційних технологій, орієнтованих на роботу з інформаційними ресурсами відкритого середовища Web, потребує аналізу вимог до засобів пошуку інформації та розробки критеріїв їх класифікації та оцінки. Це дозволить користувачам обирати такі систе-

ми семантичного пошуку, що більш ефективно задовольняють їх специфічні інформаційні потреби та вирішують різні типи проблем.

### Специфіка пошуку в Web

Пошукові механізми, орієнтовані на пошук у Web, мають враховувати специфіку цього середовища:

- *Web є розподіленим.* Однією з рушійних чинників поширення в Web є відсутність централізації. Однак, оскільки Web є продуктом діяльності багатьох людей, відсутність централізованого контролю створює багато проблем для дослідження цієї інформації. По-перше, різні спільноти використовують різні словники, що призводить до проблем синонімії (коли два різних слова мають однакове значення) і багатозначності (коли те ж слово використовується в різних значеннях). По-друге, відсутність редагування або контролю якості означає, що надійність кожної сторінки потрібно перевіряти. Інтелектуальний Web-агент не може припускати, що вся інформація, яку він збирає, є коректною та не суперечною. Через те, що не може бути жодних глобальних обмежень цілісності Web, інформація з різних джерел може конфліктувати. Деякі з цих конфліктів можуть бути викликані тим, що різні спільноти – політичні, національні, релігійні, професійні тощо – можуть мати фундаментальні розходження в думках з певних питань.

- *Web є динамічним.* Web змінюється в неймовірно швидкому темпі, набагато швидше, ніж користувач або навіть спеціалізований програмний агент можуть відслідковувати усі зміни. Додаються нові сторінки, а зміст існуючих сторінок змінюється. Деякі сторінки досить статичні або міняються на регулярній основі, а інші змінюють контент з непередбачуваними інтервалами. Крім того, значущість таких змін може бути різною: від корекції орфографічних та пунктуаційних помилок, що практично не впливає на зміст документа, до таких, які можуть повністю змінити зміст документа або видалити великі обсяги інформації. Тому потрібно враховувати, що будь-яка знайдена в Web інформація

може бути застарілою. Динамічність інформації у Web створює додаткову проблему для формалізації її семантики.

- *Web має великий обсяг* і постійно збільшується. Для масштабування за розміром постійно зростаючої мережі виникає необхідність в обмеженні виразності мови запитів або використанні спрощених алгоритмів виведення.

- *Web – це відкрите середовище.* Жоден Web-агент не може вважати, що він зібрав всі доступні знання (навіть найбільші ПС індексують близько 25 % наявних сторінок).

Запити до Web-ресурсів можна поділити на наступних класи відповідно до намірів користувачів [5]: *навігаційні* – побачити певний сайт; *інформаційні* – отримати певну інформацію; *транзакційні* – виконати якісь дії у Web. В роботі [6] проаналізовано різні визначення пошуку в Web, в яких ПС, що орієнтовані на обробку ресурсів Web, розглядаються як комп'ютерні програми, що призначені для пошуку даних у мережі, та розглянуто ПС різних поколінь. ПС *першого* покоління (AltaVista, Excite, Webcrawler тощо) знаходило інформацію, що містилася безпосередньо на Web-сторінках, а ранжування результатів враховувало тільки контент знайдених сторінок. ПС *другого* покоління для визначення релевантності IP використовують відомості про структуру самої мережі Web, аналізуючи посилання на відповідну сторінку й дані, що передаються за http-запитом, а також індекс популярності й репутацію IP. Вперше аналіз посилань між сторінками як один з основних факторів ранжування був використаний у Google (механізм PageRank), що й стало визначальним чинником популярності цієї ПС. Для ПС *третього* покоління характерна інтеграція відомостей, здобутих з різних джерел, на основі семантичного аналізу, що дає змогу враховувати в процесі пошуку та фільтрації як персональні інформаційні потреби користувачів, так і знання щодо Про пошуку [7]. Прикладами таких ПС є NAKIA, Google Squared, SenseBot та Wolfram Alpha. Ці ПС спеціалізуються на пошуку в середовищі семантично струк-

турованої інформації та її метаописах у форматах OWL і RDF. З розвитком інфраструктури Semantic Web семантичні метадані стають більш доступними. Розробка стандартів семантичної розмітки, мов опису IP (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 сфери інформаційного пошуку в напрямі використання семантики [8].

### Семантизація пошукових систем

Різні підходи до пошуку інформації у Web, включаючи частково структуровані бази даних, машинне навчання та обробку природної мови, застосовувалися до проблеми аналізу та розуміння Web-сторінок у форматі HTML та XML. Проте відсутність семантики в цих засобах представлення інформації і в інструментах їх обробки значно обмежували якість методів. Тому виникає потреба у представленні знань, що пов'язані з тематикою пошуку.

Для того, щоб забезпечити значення для даних, знання мають бути представлені в певній формалізованій формі. На проблемі подання знань спеціалізується така галузь, як штучний інтелект. Ціллю подання знань є надання структур, які дозволяють ефективно зберігати інформацію, модифікувати її, здійснювати над нею логічне виведення. В процесі досліджень у цій сфері розвивалися численні мови подання знання з різними властивостями (від ранніх мов, таких як KL-ONE [9] і KRL [10] до сучасних мов опису онтологій OWL [11]).

Для подання знань можна використовувати семантичні мережі, які представляють знання у вигляді набору вузлів, поєднаних поміченими зв'язками та фрейми. Розширені семантичні мережі та фреймові системи, як правило, включають у себе поняття абстракції, яка представлена за допомогою відношень *is-a* та *instance-of*. Відношення *is-a* вказує, що один клас є підкласом іншого, а відношення *instance-of* вказує, що поняття є

членом класу. Ці відношення мають зв'язки з теорією множин: *is-a* – як відношення підмножини, *instance-of* – як відношення екземпляра множини. Множина відношень *is-a* визначає частковий порядок на класах, який часто називають таксономією або ієрархічною класифікацією. Таксономія може бути використана для узагальнення поняття до більш абстрактного класу або спеціалізації класу для своїх більш специфічних понять. Таксономії допомагають користувачам у пошуку необхідної інформації в Web (приміром, як це робиться в Yahoo і Open Directory).

Багато дослідників надають перевагу онтологіям для подання знань для інформаційного пошуку [12]. Якщо система представлення знань визначає, як представляти поняття, то онтологія визначає, які саме поняття визначені в такій системі й як саме вони взаємопов'язані. Такі формалізовані визначення дозволяють автоматизувати набагато глибші міркування, але такі визначення набагато важче побудувати.

Безпосередньо пов'язано з представлення знань у Web використання Resource Description Framework (RDF). Модель даних RDF [13] – це семантична мережа без успадкування: воно складається з вузлів, які пов'язані поміченими дугами, де вузли представляють собою Web-ресурсів, а дуги – атрибути цих ресурсів. RDF може бути вбудований у довільні Web-документів з використанням XML-синтаксису.

Щоб забезпечити створення контрольованих, сумісно доступних, розширюваних словників (наприклад, онтологій), створено специфікацію RDF Schema, яка визначає ряд властивостей зі специфічною семантикою. RDF Schema визначає властивості, які еквівалентні *instance-of* та *is-a*, які часто використовуються в поданні знань, а також засоби для опису властивостей, домену властивостей і діапазону тощо.

Хоча RDF є поліпшенням у порівнянні HTML і XML, цього недостатньо для відображення семантики, тому що він забезпечує дуже невеликий набір семантичних примітивів і має відносно слабкі меха-

нізми керування еволюцією схеми. Тому виникає потреба в мовах опису онтологій, таких як OWL.

*Семантичний пошук* – це метод інформаційного пошуку, в якому релевантність документа запиту визначається семантично (за близькістю змісту), а не синтаксично (приміром, за частотою використання ключових слів у документі) [14]. Його можна розглядати як надбудову над традиційним інформаційним пошуком, необхідну для підвищення пертинентності пошуку. Під *пертинентністю* пошуку будемо розуміти те, наскільки знайдена інформація задовольняє ІІ користувача (на відміну від релевантності, яка визначає відповідність результатів пошуку наданому користувачем запиту).

*Система семантичного пошуку* (ССП) – це програмне забезпечення, що виконує семантичний пошук або незалежно, або як семантична надбудова над іншими ІІС. При цьому взаємодія ССП з іншим програмним забезпеченням і базами знань є наочною або прихованою від кінцевого користувача. Результатом роботи ССП може бути як здобуття інформації, яка приховано наявна 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 [15].

Для семантичного пошуку у відкритому середовищі характерним є те, що потрібні для пошуку знання також динамічно здобуваються з цього відкритого середовища, а не закладаються у систему в процесі її створення.

### Сучасні підходи до семантичного пошуку

У семантичному пошуку в Web виділяють:

1) підходи, що базуються на структурованих мовах запитів [16–22];

2) підходи, орієнтовані на некваліфікованих користувачів, які не потребують знайомства зі спеціалізованими мовами запитів.

У свою чергу, у другій групі можна виділити: підходи, де запити складаються із списків ключових слів [23–24]; підходи, де користувачі висловлюють запити природною мовою [25–29]. Більш детальний аналіз систем семантичного пошуку, класифікованих за цими параметрами, наведено в [30]. Розглянемо кілька прикладів цих підходів.

Одна з перших спроб створення семантичних запитів у Web – *SHOE* [31] – базується на структурованих мовах, та надає користувачеві:

1) інструмент для анотування Web-сторінок, що дозволяє користувачам додавати розмітку сторінки SHOE, обравши онтологію, класи та властивості з списку;

2) Web-краулер, який шукає Web-сторінки з розміткою SHOE і зберігає інформацію щодо них в індексній базі знань;

3) машину виведення, яка забезпечує розмітку нових сторінок за допомогою правил виведення (в основному, правил логіки Хорна);

4) інструменти для запитів, які дозволяють користувачам задавати структуровані запити з використанням онтологій.

Один з інструментів запитів дозволяє користувачам будувати граф, на якому вузли представляють постійні або змінні екземпляри, а дуги – відношення між ними. Щоб відповісти на запит, система знаходить підграфи, що відповідні до графа користувача. Пошуковий інструмент SHOE дозволяє користувачеві задавати запити, вибравши спочатку онтологію із списку, потім обрати класи та властивості з іншого списку. Після цього система будує запит, що з'єднує обрані елементи, здійснює цей запит до БЗ, і представляє результати його виконання у вигляді таблиці.

Більшість інших підходів цієї групи базуються на RDF. Наприклад, ІІС *Swoogle* [32] використовує краулер для виявлення, індексації та запитування документів у форматі RDF. *Swoogle* в основному забезпечує пошук документів та термінів Semantic Web (тобто, URI, класів і властивостей). Це дозволяє користувачам будувати запити, що містять умови щодо ме-

таданих на рівні документа (тобто, запити обробляють документи, що мають RDF як розширення файлу), а також дозволяє користувачам шукати документи Semantic Web, використовуючи RDF / XML як мову синтаксису. Знайдені документи ранжуються відповідно до алгоритму впорядкування, який враховує важливість документів у Semantic Web.

Система *Corese* [33] – це орієнтована на Semantic Web ІПС на онтологічній основі, яка знаходить Web-ресурси, що ановані в RDF(S), за допомогою мови запитів, що базується теж на RDF(S). *Corese* дозволяє виконувати приблизний пошук у Semantic Web. Для приблизного пошуку використовуються правила виведення і обчислення семантичної відстані між класами та властивостями в онтологічних ієрархіях. За допомогою цього апарату *Corese* знаходить Web-ресурси, анотації яких є уточненням (спеціалізацією) запиту, а також ресурси, анотації яких відносяться до понять і відношень, які ієрархічно досить близькі до запиту.

Інший підхід до пошуку за структурованими запитами наведено в [34]. Метою цього підходу є апроксимація запитів до наборів даних RDF за допомогою SPARQL [35]. Для цього запит SPARQL кодується як набір трійок обмежень зі змінними, а приблизна відповідь – це заміна змінних з даними, які можуть не задовольняти всім обмеженням. Запропонована стратегія поступово покращує достовірність відповідей й спиняється, коли отриманий результат задовольняє користувача.

Подальший розвиток підходів на основі структурованих мов розглядається далі. *ONTOSEARCH2* [36] надає механізм для пошуку онтологій у Semantic Web. Він зберігає копії онтологій у легкій для обробки дескриптивній логіці і дозволяє виконувати запити SPARQL як до структури, так і до екземплярів онтології.

Система *Coraal* [37] – це ІПС на основі знань, що спеціалізується у пошуку літератури з біомедицини, яка використовує евристики на обробки природної мови. Це дозволяє аналізувати тексти і будувати з них трійки RDF, які інтегруються з ная-

вними знаннями щодо ПрО пошуку, і користувач може надавати запити до всієї зібраної інформації за допомогою відповідної мови запитів.

*NAGA* – це семантична ІПС [38], яка пропонує мову запитів на основі графів для пошуку в базі знань (БЗ), яка також представлена у вигляді графу. БЗ створюється автоматично за допомогою інструменту для здобуття знань з Web-джерел на основі підходу, запропонованого в [39]. Вузли та ребра в графі знань представляють сутності й відношення між об'єктами, відповідно. Мова запитів *NAGA* розширює SPARQL таким чином, що дозволяє будувати складні запити у вигляді графів з регулярними виразами відносно відношень на мітках дуг. Відповідями на запит є підграфи графа знань, які відповідають графу запиту і впорядковуються за допомогою моделі зважених помічених графів.

### Семантичний пошук за ключовими словами

Розглянемо кілька прикладів підходів до семантичного пошуку на основі ключових слів. ССП *OntoSelect* [40] фокусується на питаннях, пов'язаних з пошуком онтологій. Ця система дозволяє користувачам шукати онтологію, вказавши її назву або тему, яка цікавить користувача. Щоб описати область своєї зацікавленості, користувач може вказати URL Web-документа, який містить інформацію щодо цієї теми. Потім за допомогою лінгвістичних та статистичних методів автоматично будується набір ключових слів, які використовуються для пошуку.

В роботі [41] описано підхід, який фокусується на покращенні результатів традиційного пошуку за ключовими словами за допомогою даних, отриманих від Semantic Web. Коли користувач надає запит, то терміни цього запиту (ключові слова) відображаються у вузлах Semantic Web. У випадку неоднозначного співставлення використовуються евристики (наприклад, з урахуванням профілю користувача) для зняття неоднозначності. Після того, як знайдені вузли, що відповідають пошуковим термінам, використовує евристики для

того, щоб визначити, яка саме частина графу Semantic Web навколо цих вузлів має бути надана користувачеві як результат пошуку (тобто, яким є поріг  $N$  для надання перших  $N$  трійок).

Пізніші підходи до семантичного пошуку за ключовими словами в основному орієнтовані на некваліфікованих користувачів. ССП *SemSearch* [42] забезпечує подібний до Google інтерфейс запитів, що не потребує від користувачів мати знання щодо онтологій або мов. Запити користувачів складаються з двох або більше ключових слів, семантичне значення яких враховується для переформулювання запиту відповідно до синтаксису формальної мови запитів. Ключовим словам присвоюються семантичні значення шляхом співставлення їх з набором класів, властивостей і екземплярів у репозиторії семантичних даних. Кожному ключовому слову може відповідати клас, властивість або екземпляр. Розглядаються різні комбінації семантичних співставлень ключових слів. Наприклад, кожному ключовому слову може відповідати клас, або першому ключовому слову відповідає клас, а другому – властивість і так далі. Всі комбінації співставлень враховуються в процесі переформулювання запиту, а за кожною комбінацією створюється вдосконалений формальний запит, який отримується із заздалегідь визначеного набору шаблонів запитів. Після переформулювання формальні запити точно оцінюються, і це забезпечує результати, які семантично пов'язані з усіма ключовими словами із запиту користувача.

У роботі [43], аналогічно до розглянутого вище підходу, запити за ключовими словами трансформуються в кон'юнктивні запити, які мають бути оцінені проти відносно базової БЗ. Але структура таких формальних запитів, які оцінюються, не відповідає якимось попередньо визначеним шаблонам. Формальні запити будуються шляхом використання техніки на основі графів, щоб знайти зв'язки між сутностями в запиті користувача. Трансформація запиту складається з трьох кроків:

- ключові слова запиту співставляються з елементами онтології;

- аналізуються відношення між цими елементами онтології, і на основі цього будуються підграфи БЗ, кожен з яких являє собою набір співвідношень, що пов'язують всі розглянуті елементи, а множина цих підграфів представляє всі можливі відношення між ключовими словами запиту користувача, які не можуть бути явно задані цим користувачем, тобто ці підграфи відповідають різним запитам, що можуть зацікавити користувача;

- ці підграфи трансформуються відповідно до правил мови подання запитів та генеруються формальні запити, які співставляються із вмістом БЗ.

*Falcons* [22] відноситься до ПС на основі ключових слів для Semantic Web. Ця система підтримує пошук концептів та об'єктів. Пошук концептів здійснюється шляхом пошуку класів і властивостей в онтології, обраній користувачем, які відповідають умовам запиту. Крім того, інші онтології можуть бути рекомендовані користувачу на основі комбінації методу TF-IDF і рейтингу популярності онтологій. Пошук об'єкта виконується аналогічним чином: крім пошуку об'єктів, які відповідають умовам запиту, система також рекомендує інші типи об'єктів, які, ймовірно, теж зацікавлять користувача.

*SWSE* [23] і *Sig.Ma* [44] – це інструменти, що дозволяють користувачам знаходити об'єкти RDF за ключовими словами. Зокрема, результат пошуку за ключовим словом в *SWSE* – це список об'єктів, що співставлені з цим ключовим словом з невеликим описом і ім'ям концепту, приміром, "Person", "Professor". Якщо користувач натискає на "Person", то результати фільтруються і користувачу показують тільки список об'єктів класу "Person". Інформація про об'єкт інтегрується з кількох джерел і представлена в однорідному вигляді.

Ядром *SWSE* є розподілена архітектура *YARS2* [45] для індексації та пошуку наборів даних RDF. *YARS2* збирає фрагменти інформації і агрегує їх або за рахунок використання URI об'єктів (у тому випадку, якщо унікальний ідентифікатор використовується в різних джерелах), або шляхом використання інших методів кон-

солідації об'єкта. Крім того, SWSE дозволяє досвідченим користувачам будувати складні запити у SPARQL.

Подібно до SWSE, Sig.Ma об'єднує результати з кількох джерел, забезпечуючи користувачеві, крім самих ресурсів, узагальнений погляд на інформації. Фаза зняття неоднозначності аналогічна SWSE, але в цьому випадку дії користувача використовуються для усунення нерелевантних джерел. Sig.Ma також дозволяє користувачам вказувати список не тільки об'єктів, але й інших властивостей. Ключові слова користувача переводяться в набір запитань, деякі з них звертаються до Yahoo Boss [46] для отримання Web-сторінок, водночас як інші звертаються до індексу даних Semantic Web Sindice [47], щоб зібрати об'єкти і властивості RDF. Нарешті, вся отримана інформація інтегрується, використовуючи деякі евристичні методи, що базуються на використанні URI і методів консолідації міток.

Новий підхід, спрямований на допомогу користувачеві в побудові семантичних запитів за запитами з ключових слів, представлено в системі *QUICK* [48]. Семантичний запит – це запит, який враховує онтологію відповідної Про. Коли користувач формулює запит з ключового слова, QUICK починає роботу з того, перетворює його в кілька семантичних запитів, кожен з яких отримується шляхом присвоєння якогось онтологічного поняття (властивість, об'єкт тощо з обраної онтології) для кожного з ключових слів. Потім користувач має обрати серед семантичних запитів, згенерованих системою, найбільш відповідний його інтересам. Якщо жоден семантичний запит, згенерований QUICK, не подобається користувачеві, то користувач самостійно може вести систему до генерації відповідного запиту, забезпечуючи додаткові характеристики (наприклад, вказуючи, що певне ключове слово має бути використане як властивість або як об'єкт).

Серед ІПС на основі ключових слів для Semantic Web важливо відмітити *YahooSearch-Monkey* [49], яка ставить ціллю поліпшення якості результатів пошуку ІПС Yahoo! Вона дозволяє власни-

кам сторінки вказувати, яку інформацію по Web-сторінці вони хочуть запропонувати для відображення на сторінці результатів пошуку Yahoo! Видавці можуть надати ці характеристики у вигляді мікроформатів або метаданих RDF, які будуть автоматично отримані в процесі сканування сторінок і нададуть ІПС велику кількість інформації.

### Семантичний пошук за природномовними запитами

Багато ССП виконують пошук на основі обробки природної мови. Приміром, у системі *ORAKEL* [27] запити спочатку переводяться в логічну форму, а потім переформулюються відповідно до цільової мови, тобто тієї мови, на якій базується БЗ. Переклад з логічної форми на цільову мовою описується декларативно за допомогою Prolog-програми. Загальний підхід не залежить від конкретної цільової мови, тому зміна мови онтології вимагає тільки декларативного опису трансформації в Prolog, але ніяких подальших змін в основній системі. Система орієнтована на користувача певного типу – інженера лексики, який визначає, як вирази природної мови відображаються на предикатів в БЗ, тобто, як дієслова, прикметники і відносні іменники тощо відображаються на відношення в онтології Про.

В роботі [29] представлена система семантичного пошуку, яка підтримує як пошук у Semantic Web через онтології, так і семантичний пошук у документах, що не відносяться до Semantic Web. Для першого виду пошуку, відповіді на запит природною мовою здобуваються шляхом застосування системи *PowerAqua* [50], яка працює таким чином: запит користувача перекладається з природної мови в структурований формат – “лінгвістичну трійку”, а терміни цієї трійки відображаються на семантично релевантні об'єкти онтології. Потім обираються ті онтологічні сутності, які найкраще представляють запит користувача.

ССП *PowerAqua* розширює систему *Aqua-Log*, запропоновану в [51], що працює з використанням тільки однієї онтології, на випадок кількох онтологій, та вирі-



шує проблему неповноти знань шляхом переходу до традиційного пошуку за ключовими словами, якщо немає онтології, що задовольняє запити. Другий вид пошуку в [29], а саме, семантичний пошук документів, що не відносяться до Semantic Web, здійснюється шляхом розширення системи, запропонованої в [52], за допомогою нового підходу до анотування документів, який містить такі етапи:

- здобуття текстового представлення семантичних сутностей;
- пошук цього текстового представлення у Web-документах;
- створення анотації, яка пов'язує семантичні сутності з кожним документом, що містить їх текстові представлення.

Сучасний підхід для побудови запитів SPARQL із запитів природною мовою представлено в [53]. Першим кроком у генерації запитів SPARQL є трансформація природномовних запитів у набір понять онтології (класів, екземплярів, властивостей і літералів), які базуються на призначенні правильного поняття онтології для кожного слова. Якщо система не може автоматично призначити правильне поняття онтології для слова, то викликається користувач для виконання цієї дії. Вибір користувачів використовується для навчання системи, щоб поліпшити її характеристики. Другим кроком є будівництво трійок онтологічних понять, які, нарешті, вставляються в розділи SELECT та WHERE для генерації запиту SPARQL. Результати оцінки отриманого SPARQL запиту показують користувачу в табличному та в графічному вигляді.

### Використання онтологій у семантичному пошуку

Підхід до семантичного пошуку в Web, представлений в [54], базується на структурованій мові запитів, яка дозволяє сформулювати складні пошукові запити на основі онтологій. Онтологічно збагачений Web разом з комплексним пошуком на основі онтологій здійснюється на базі існуючого Web і з використанням існуючих орієнтованих на Web ІПС. Замість того, щоб інтерпретувати на синтаксису

ключових слів чином, фрагменти даних існуючих Web-сторінок пов'язані з якоюсь онтологічною БЗ, а потім інтерпретуються відносно цієї БЗ. Web-контент пов'язаний з семантичним анотаціям, або, з іншої точки зору, Web співставляється з онтологічною БЗ, які потім забезпечує семантичний пошук у Web щодо цієї онтології. Вважається, що семантичні анотації та онтології, що лежать в їх основі, надаються явно. Обробка запитів у цьому підході до семантичного пошуку в Web поділяється на:

- офлайнове виведення для попередньої обробки онтологічних знань з використанням стандартних методів виведення на онтологіях, щоб перетворити семантичні анотації в так звані складні семантичні анотації, які публікуються як стандартні Web-сторінки, тобто вони можуть бути знайдені за допомогою стандартних пошукових систем для Web;

- онлайнове скорочення складних запитів на основі онтологій для пошуку в Web у послідовні або стандартні запити для пошуку в Web, відповіді на які можуть бути отримані за допомогою звичайного пошуку в Web, а потім використані для побудови відповіді на оригінальний пошуковий запит на основі онтології.

Цей спосіб обробки семантичних запитів є онтологічно коректним (і в багатьох випадках також повним). Ранжування результатів пошуку базується на ранжуванні об'єктів ObjectRank, яке узагальнює (і може бути зведено до) звичайне рейтингування PageRank для Web-сторінок. Таким чином, основні частини онтологічного пошуку в Web фактично зводяться до сучасних ІПС. Але важливою перевагою даного підходу є те, що цей підхід може бути негайно застосовуватися до всього існуючого Web, і це можна здійснити на основі існуючих Web-технологій пошуку. Такий напрям досліджень спрямований на додавання структури та семантики на основі онтологій до існуючих ІПС у Web шляхом об'єднання існуючих Web-сторінок і запитів з онтологічними знаннями.

Онтологічні знання і анотації, що лежать в основі такого семантичного по-

шуку, можна класифікувати відповідно до їх вмісту:

- загальні знання (наприклад, знання, які містяться у Вікіпедії) для загального пошуку в Web на основі онтології;
- спеціалізовані знання (наприклад, біомедичні знань) для ієрархічного пошуку в Web на основі онтологій.

Інтерфейс на основі загальної онтології для більш досвідчених користувачів базується на повній потужності структурованого мови запитів (для яких базова онтологія може бути візуалізована для підтримки формулювання запитів). Для менш досвідчених користувачів використовується заздалегідь визначені прості форми інтерфейси (наприклад, схожі на ті, які використовуються в Google для розширеного пошуку).

У роботі [55] досліджується варіант описаного вище підходу, який використовує індуктивні методи міркування на додаток до дедуктивних. Це додає системі здатність обробляти суперечність, шум і неповноту у даних.

Хоча існує багато підходів до семантичного пошуку в Web, і вже створено різноманітні системи на основі цих підходів, дослідження в цій області ще знаходяться на самому початку, і багато проблем відкриті для подальших наукових досліджень. Виконання Web-пошуку у вигляді повернення простих природномовних відповідей на прості запитання природною мовою все ще залишається неможливим, не кажучи вже про виконання Web-пошуку для запитів, відповідь на які залежить від певної Про.

Деякі з найбільш нагальних проблем пов'язані з тим, як автоматично трансформувати природномовні запити в формальні на основі онтологій, як автоматично додати семантичні анотації до Web-контенту та як автоматично здобувати знання з Web-контенту.

Іншим важливим питанням у дослідженнях семантичного пошуку в Web є створення та підтримка базових онтологій. Це може здійснюватися:

- вручну відповідними фахівцями (наприклад, як це робиться в семантичній

Вікіпедії, де різні спільноти можуть визначати свої власні онтології);

- автоматично, наприклад, шляхом знаходження в Web вже існуючих фрагментів онтологічних знань та анотацій (з існуючих онтологій або фрагментів онтологій та з існуючих анотацій Web-сторінок у мікроформатах або RDFa);
- напівавтоматично за допомогою комбінації перших двох підходів.

Очевидно, що чим більше ступінь автоматизації, тим більше і потенційний розмір онтологій, які можуть бути побудовані, і менші витрати і зусилля для їх створення і підтримки. Таким чином, для дуже великих масштабів Web є бажаним дуже високий ступінь автоматизації. З цим тісно пов'язаний такий важливий напрямок досліджень, як еволюція, оновлення і відображення онтологій, які є основою семантичного пошуку в Web.

Важливим є й питанням щодо того, як враховувати явну та неявну контекстну інформацію, щоб адаптувати результати пошуку до потреб користувачів. Наприклад, потреби і мотивація користувачів можуть бути визначені в термінах онтології на основі строгих або м'яких обмежень і умовних переваг.

### Критерії класифікації систем семантичного пошуку

У роботі [56] проаналізовано підходи до пошуку в Web текстових документів, не враховуючи пошук структурованих документів (XML, RDF, OWL тощо). Також не аналізуються ті підходи, які вимагають, щоб користувач формулював запити формальною мовою, наприклад, SQL або SPARQL, та рішення, де користувач задає не тільки ключові слова, а ще й, приміром, використовує запит за зразком. В цьому огляді аналізуються ПС, орієнтовані на пошук інформації у Web або на метапошук на основі інших ПС. При цьому аналізуються тільки ті рішення, які використовують семантичний пошук. Якщо традиційний пошук документів залежить головним чином від наявності ключових слів у документах, то семантичний пошук відрізняється від нього

тим, що використовує ще й певні знання щодо домену пошук. Знання домену може бути формалізована за допомогою онтології ПрО, класи якої описують поняття даної ПрО, екземпляри характеризують факти, а властивості – відношення між поняттями та фактами. Ресурсами називають як класи, так і екземпляри.

Класифікувати різні підходи до семантичного пошуку можна за такими параметрами: архітектура, з'єднаність (coupling), прозорість, контекст користувача, модифікація запитів, структура онтології і онтологічна технологія. Ці критерії не є повністю незалежними один від одного, але вони відображають важливі характеристики ССП. Слід відмітити, що існують й інші критерії для класифікації ССП, наприклад, продуктивність, масштабованість, розподіленість, адаптованість і рейтинг результатів, але вони менше пов'язані зі специфікою саме семантичного пошуку.

Семантичні ППС (так само, як і традиційні) поділяються за архітектурою на два типи:

1) автономні машини пошуку (з власними краулером та індексною БД);

2) метапошукові системи, яка використовує індексні БД інших ППС та інтегрує отримані від них результати.

*З'єднаність* – це критерій класифікації ППС, який стосується взаємодії між документами і онтологіями. Існують два типи зв'язку – тісний та слабкий.

*Тісний зв'язок* між документами і онтологіями існує тоді, коли метадані документів відносяться безпосередньо до понять конкретної онтології або навпаки (приміром, коли окремі документи розглядаються як екземпляри відповідної онтології). Такий підхід дозволяє легко вирішувати проблему омонімії через вибір відповідного поняття в онтології, але це підвищує вартість семантичного анотування документа. На практиці такий підхід застосовують не тільки для спеціалізованих інформаційних систем, які керують даними в обмеженій ПрО або для певної організації, але й у більш загальних областях застосування, таких як пошук у Web.

*Слабкий зв'язок* між документами і онтологіями встановлюється тоді, коли документи не прив'язані до певної онтології. У такому випадку існує проблема вибору відповідної онтології для даної ПрО. Таким чином, семантична потужність слабо пов'язаних систем обмежена: приміром, важко подолати проблему омонімії. Але цей підхід широко застосовують для пошуку в Web, якщо тільки дуже невелика частина документів має семантичні анотування. ППС зі слабким зв'язком можуть бути легко реалізовані у вигляді метапошукових систем.

Цей критерій має велике значення для вибору ССП через те, що слабкий зв'язок між документами і онтологіями значно обмежує точність семантичного пошуку, водночас як тісний зв'язок між ними дозволяє знаходити тільки анотовані документи, тобто обмежує його повноту.

*Прозорість* ППС відносно взаємодії користувача з особливостями семантичних систем можна поділити на наступні типи:

- *прозорі*: семантичні можливості системи невидимі для користувача, тобто семантична ППС виглядає так само, як і традиційна;

- *інтерактивні*: системи можуть запитувати користувача щодо пояснень до запиту або рекомендувати внести модифікації у запит;

- *гібридні*: системи поєднують властивості інтерактивних і прозорих систем, тобто у стандартному випадку вони виступають як прозорі системи, але для дуже специфічних завдань можуть використовувати інтерактивну взаємодію з користувачем.

Прозорість ППС є одночасно і недоліком, і перевагою: з одного боку, користувач позбавлений тривалих діалогів з системою, що робить ППС зручною у використанні, але з іншого боку, користувач не може впливати на семантичні рішення такої системи (наприклад, для вирішення омонімії), що потенційно призводить до зниження точності пошуку.

*Контекст користувача* дозволяє оцінити корисність документів для того,

щоб краще задовольняти інформаційні потреби конкретного користувача. Можна виділити наступні класи ППС відносно того, як саме вони можуть враховувати цей контекст:

- *навчання*: контекст користувача динамічно здобувається в процесі взаємодії користувача з ППС. На основі запиту та історії уточнень цього запиту система прогнозує, які саме результати прагне отримати цей користувач. Якщо терміни запиту завжди належать до того самого онтологічного контексту, то на основі цього система може вирішити омонімію, обравши термін з цієї ж онтології.

- *жорстке кодування* (hard-coded) – цей підхід поділяє запити на так звані категорії питань, які визначають інформаційну потребу користувача. Система забезпечує фіксовану кількість категорій питань, які використовуються для оцінки запиту. Типові категорії можна визначити, яку інформацію необхідно надати користувачеві, наприклад, "розташування" або "загальні ресурси для". Це дозволяє вирішити омонімію, відокремлюючи, приміром, назви географічних об'єктів від прізвищ осіб.

Зв'язок запиту користувача з категорією питань може здійснюватися явно самим користувачем, або неявно – на основі належності до групи користувачів, або шляхом аналізу самого запиту. Як правило, в рішеннях на основі онтологій класи інформаційних потреб мають відповідати певним онтологічним структурам, таких як типів властивостей. Контекст користувача забезпечує важливу інформацію про інформаційні потреби користувача. Ця інформація може використовуватися також для модифікації запиту.

Семантична *модифікація запитів* користувача – це добре відомий метод пошуку інформації. В області семантичного пошуку часто використовують інформацію з онтології. Він грає центральну роль у багатьох семантичних пошукових системах. Різні методи були розроблені для того, щоб збільшити повноту і точність пошуку. Наявність онтологій дозволяє відносно легко підвищити повноту запиту: онтологія стає для пошукової ма-

шини джерелом більш загальних, порівняно з ключовими словами запиту, термінів, – їх надкласів в онтології. Дещо важче поліпшити точність запиту. Це потребує вирішення проблеми омонімії та вибору певного підкласу для кожного з використаних у запиті термінів.

Модифікувати запит можна вручну, на основі онтологічного графу та шляхом переписування запиту. *Вручну* – це найпростіший спосіб змінити запит, залишивши модифікацію запиту користувачеві. Коли користувач вводить запит, система повертає йому не тільки знайдені ним документи, але й відповідну частину онтології. Використовуючи знання щодо ПрО, отримані з цього фрагмента онтології, користувач може сам переформулювати запит, додавши або видаливши певні терміни.

Щоб оптимізувати запити користувачів *на основі графу*, необхідно підтримувати тісний зв'язок між базою документів і онтологією. При цьому як онтологічні поняття, так і документи розглядаються як вузли графу. Терміни запиту використовуються для того, щоб знайти відповідні вузли графу. З цих вузлів алгоритм обходить граф, щоб визначити документи, семантично пов'язані з ними. Ця ціль може бути досягнута, наприклад, за допомогою алгоритму поширення збудження. Модифікація запиту на основі графу відрізняється від перезапису запитів тим, що не будується новий запит, який потім має оброблятися в пошуковій системі, а замість цього безпосередньо повертаються відповідні документи. Крім того, така обробка аналізує запит у цілому, а не розкладає його на окремі терміни.

*Перезапис запиту* базується на тому, що запит можна оптимізувати за допомогою семантичної ППС. Основні способи перезапису запиту – розширення, обрізка і заміщення термінів. У випадку *розширення* запит доповнюється термінами, які отримуються від онтологічного контексту термінів оригінального запиту. Залежно від структури онтології, можуть бути використані різні семантики.

*Обрізка* запиту видаляє терміни з запиту і має протилежний ефект. Також

вона може бути реалізована шляхом порівняння результатів обрізаного запиту з результатами початкового запиту. Використання розширення і обрізки запитів у тому випадку, коли запит складається з кон'юнкції (AND) термінів, викликає те, що запит стає більш специфічним з кожним додатковим членом, а коли запит складається з диз'юнкції (OR) термінів, то він стає більш загальним. Іншими словами, відносно інформаційної потреби користувача, довгі кон'юнкції у запиті дають високу точність, а довгі диз'юнкції забезпечують високу повноту пошуку.

Отже, обидва методи, розширення диз'юнктивних запитів і обрізка кон'юнктивних запитів, збільшують повноту пошуку, а обрізка диз'юнктивних запитів і розширення кон'юнктивних запитів збільшують точність пошуку. Багато систем забезпечують лише кон'юнктивні або диз'юнктивні запити, тобто терміни пошуку неявно і незмінно пов'язані тільки оператором AND або тільки оператором OR. Тому ідея диз'юнктивного розширення або обрізки не застосовується до систем, які забезпечують тільки кон'юнктивні запити.

*Заміщення* термінів запиту – це заміна термінів запиту іншими онтологічно пов'язаними з ними термінами. В цілому, терміни можуть бути замінені на синоніми, підкласи або надкласи з онтології для того, щоб збільшити точність або повноту пошуку відповідно. Заміщення розглядається окремо від розширення та обрізки з наступних причин: у порівнянні з набором результатів оригінального запиту диз'юнктивне розширення та кон'юнктивна є надмножиною результатів, а диз'юнктивна обрізка і кон'юнктивне розширення є підмножиною результатів. На відміну від цього, заміна може дати набір результатів, який лише частково перекриває початковий набір результатів.

Якщо користувач ітеративно виконує кілька запитів, щоб задовольняють якусь інформаційну потребу, заміщення може бути більш ефективним методом для того, щоб допомогти користувачеві відповідно до його нових знань щодо ПрО.

ССП на основі онтологій використовують таку *онтологічну структуру*, яка містить поняття та відношення між ними. Серед відношень можна виділити наступні:

- *анонімні* відношення, коли ПС ігнорує ім'я та семантику відношення, а враховує лише наявність взаємозв'язку між двома поняттями онтології, який означає тільки, що ці поняття мають той самий контекст;

- *стандартні* відношення, до яких можна віднести синонімію, меронімію, заперечення, “клас-підклас”, “клас-екземпляр”, використання яких збагачує семантичні можливості пошуку, але також вводить залежність від онтологічних структур;

- відношення, *специфічні для ПрО*, які визначають семантику зв'язку між термінами конкретної ПрО (приміром, для географічної ПрО відношенням “бути столицею” можуть бути пов'язані екземпляри класів “місто” та “країна”).

В процесі пошуку в ССП можуть застосовуватися будь-які комбінації відношень всіх трьох типів.

*Онтологічна технологія* пов'язана з тим, яка мова опису онтологій використовується (різні рішення використовують для подання онтологій різні мови – F-Logic, RDF, DAML (+OIL) та OWL) та з технологічними питаннями повторного використання та взаємодії онтологій.

Аналіз існуючих ССП на основі цих критеріїв показує, що в різних системах реалізовані різні підмножини семантичних можливостей. Це дозволяє виділити області подальшого розвитку ССП:

- аналіз засобів модифікації запитів, який забезпечив би кількісне порівняння методів модифікації запиту;

- розвиток метапошукових семантичних ПС: існуючі метапошукові системи змінюють запити користувачів і переспрямовують їх до традиційних ПС, але не можуть переспрямовувати пошук до інших ССП;

- аналіз визнання користувачів: у більшості існуючих ССП семантика є прозорою для користувача, але чим більше

інтерактивність цих ПС, тим потужніше може бути система. Проте важко оцінити, наскільки користувач готовий до певних зусиль з семантичної взаємодії, спрямованих на те, щоб поліпшити результати його пошуку;

- збільшення гнучкості пошуку – багато семантичних ПС орієнтовані на певну структуру онтології, тобто накладають обмеження на клас онтологій, які вони можуть використовувати, водночас як інші системи (орієнтовані на обробку «анонімних» властивостей) можуть впоратися з довільними онтологій, але забезпечують більш слабкі семантичні можливості;
- впорядкування знайдених документів на основі обраної онтології;
- інтеграція з системами керування контентом та документами (CMS /DMS);
- продуктивність та масштабованість – семантичні пошукові системи мають бути здатні конкурувати зі стандартними пошуковими ПС.

### Система семантичного пошуку “МАПС”

Вищенаведений аналіз показує, що при розробці засобів семантичного пошуку в Web необхідно звернути увагу на наступні питання:

- надати можливість використовувати зовнішні бази знань, які містять знання в сфері інформаційних потреб користувача;
- застосовувати інтероперабельні засоби подання знань, що підтримують розподілене та повторне використання відомостей щодо предметної області, що цікавить користувача, щодо інформаційних об’єктів, доступ до яких він прагне отримати, та щодо його особистих можливостей, пов’язаних із сприйняттям інформації;
- забезпечити користувачам можливість самостійно встановлювати, які саме знання використовуються в процесі пошуку (приміром, надати можливість обирати онтологію Про, а потім за цією

онтологією модифікувати та вдосконалювати пошукові запити);

- враховувати історію взаємодії користувача з ПС та контекст пошуку, щоб персоніфікувати результати пошуку та впорядковувати їх відповідно до інформаційних потреб користувача;
- забезпечити візуалізацію тих знань, що застосовуються для семантичного пошуку, щоб пояснити користувачеві шляхи отримання результатів та запобігти неправильного розуміння його інформаційних потреб;
- підтримувати збереження історії взаємодії користувача з ПС, щоб враховувати його персональні потреби та зменшити час обробки запитів (приміром, зберігати постійні інформаційні запити та відомості про онтології, що використовувалися для їх побудови та виконання);
- підтримувати переадресацію семантично збагачених запитів до інших ПС – як традиційних, так і семантичних;
- підтримувати колаборативний пошук та враховувати досвід взаємодії з іншими користувачами.

Всі ці питання були враховані в процесі проектування та реалізації ССП “МАПС” [57], яка орієнтована на користувачів, що мають у мережі постійні інформаційні інтереси. Для цього “МАПС” надає можливість зберігати й повторно виконувати запити, враховувати реакцію користувача на раніше запропоновані йому ІР (персональна фільтрація), відстежувати появу аналогічних запитів у інших користувачів (колаборативна фільтрація), зберігати формальний опис сфери інтересів користувача у вигляді онтології (семантична фільтрація) тощо.

Крім того, у “МАПС” у процесі профілювання користувачів використовується специфічний для природномовних ІР критерій оцінювання – складність тексту для розуміння. Особливістю системи є використання оригінального знання-орієнтованого алгоритму, що дає змогу визначити складність розуміння тексту для конкретного користувача (для того, щоб формалізувати рівень обізнаності користувача в

певних ПрО, використовуються тезауруси тих предметних областей, що цікавлять користувачів) [58]. Основою “МАПС” є технології Semantic Web, зокрема мова подання онтологій OWL, і засоби його обробки. Для подання знань, що цікавлять користувача, використовуються онтології ПрО та тезауруси задач, що базуються на них: тезаурус будується користувачем за відповідною онтологією самостійно, а онтологія вибирається з набору онтологій, запропонованих на сайті розробниками системи.

“МАПС” базується на онтологічній моделі, що здійснює опис семантики взаємодії користувачів і ресурсів “МАПС” в інформаційному просторі Web [59]. Користувач має вибрати онтологію, що характеризує ПрО, яка його цікавить (якщо онтології немає, то її слід сформувати самостійно, наприклад, побудувати в Protege чи знайти у Web).

В обраній онтології користувач відбирає множину термінів, що стосуються його запиту, та формує з них *тезаурус запиту*. Користувач може позначити терміни, наявність яких у шуканому документі є бажаною або небажаною, а також задати більш складні операції, наприклад, автоматично позначати всі терміни, що знаходяться у певних відношеннях із термінами, позначеними раніше. Це дає змогу, зокрема, легко враховувати під час пошуку синоніми чи близькі за значенням слова, а також здійснювати пошук відразу кількома мовами. Унаслідок цього формується непорожня множина слів (чи словосполучень)  $W = \{w_1, \dots, w_m\}$ , кожне з яких може мати свою позитивну або негативну вагу  $v_k$ ,  $k = \overline{1, m}$ . Після цього для кожного документа  $i_j$ ,  $j = \overline{0, k}$  з множини  $\Gamma, \Gamma \subseteq I$  формується коефіцієнт відповідності контексту пошуку:

$$s_j, j = \overline{0, k}, s_j = \sum_{k=1}^m v_k * f(i_j, w_k),$$

$$\text{де } f(i_j, w_k) = \begin{cases} 1, w_k \in i_j \\ 0, w_k \notin i_j \end{cases}.$$

Чим вищим є цей коефіцієнт, тим, імовірно, вищою буде релевантність документа запиту користувача.

У “МАПС” реалізуються такі базові операції над тезаурусами, як об’єднання, перетинання, різниця. Крім того, підтримується побудова тезауруса за документом, тобто формується множина слів, що використовуються у визначеному Пр. Також користувач може редагувати тезауруси вручну. Кожне слово в тезаурусі має визначену вагу (ціле позитивне або негативне число), що визначає важливість слова для запиту. Негативні значення відповідають термінам, які користувач вважає небажаними. Тезаурус може відображатися у вигляді хмари тегів (розмір шрифту відображає вагу терміну в тезаурусі, а терміни з негативною вагою позначені червоним кольором) [7]. Спосіб виконання пошуку залежить від специфіки конкретних Пр (рис. 1).

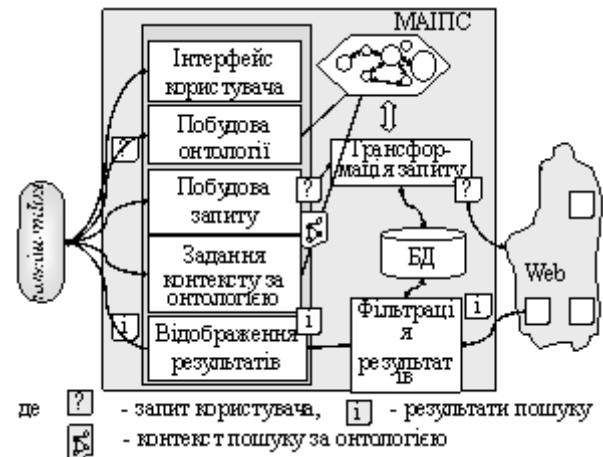


Рис. 1. Пошук в МАПС на основі онтології

Проаналізувавши наявні відомості про інформаційну потребу користувача (для того, щоб обробка виконувалася на семантичному рівні, “МАПС” використовує знання, що містяться у відповідних зовнішніх і внутрішніх онтологіях), система передає запит до зовнішніх ПС. Сьогодні в ролі такої ПС використовується Google, як найдосконаліша з сучасних пошукових систем. Отримавши у відповідь від зовнішньої ПС набір інформаційних ресурсів, “МАПС” намагається здобути з них потрібні користувачеві ві-

домості. У найпростішому варіанті, якщо потрібний користувачеві ІО є документом (можливо, певного типу), то система перепорядковує отримані посилання на ІР з урахуванням персональних особливостей користувача та збережених у БД системи відомостей щодо цих ІР. В результаті виконання пошуку користувачу надається така інформація: тезаурусний рейтинг ІР, посилання на ІР, його назва, анотація й оцінки легкості читання (рис. 2). Припускається, що ІР, у яких зустрічається більше слів з тезаурусу, більш релевантні потребам користувача.



Рис. 2. Результати пошуку в системі “МАПС”

“МАПС” реалізована як серверне Інтернет-застосування мовою PHP версії 5. 0. Для збереження внутрішніх даних використовується XML (надалі планується використання СУБД MySQL). Онтології зберігаються у форматах RTF і OWL, тезауруси – у форматі XML.

### Висновки

У дослідженні, результати виконання якого наведено в цій роботі, проаналізовано сучасний стан та перспективи розвитку систем семантичного пошуку, орієнтованих на обробку інформаційних ресурсів Web, розглянуто критерії їх класифікації. На основі цього аналізу запропоновано засоби та методи персоніфікації семантичного пошуку та зовнішні джерела інформації, з яких доцільно здобувати знання, які дозволяють підвищити перти-

нентність пошуку і більш повно та точно задовольняти сталі інформаційні потреби користувачів. Ці підходи реалізовано в системі семантичного пошуку “МАПС”, яку також описано у даній роботі.

1. *Hendler J.* Web 3.0: The dawn of semantic search. *Computer*. 2010. 43(1)/. P. 77–80.
2. *Baeza-Yates R., A. Raghavan R.* Next generation Web search. S. Ceri and M. Brambilla, editors, *Search Computing*, Springer. 2010. P. 11–23.
3. *Lawrence S.* Context in the Web Search – <http://citeser.nj.nec.com/lawrence00context.html>.
4. *Janowicz K., Wilkes M., Lutz M.* Similarity-based information retrieval and its role within spatial data infrastructures. *Proc. GIScience-2008*, Springer. 2008. P. 151–167.
5. *Broder A.* A taxonomy of web search, IBM Research, ACM SIGIR Forum archive. Vol. 36 , Issue 2 (Fall 2002). P. 3–10.
6. *Гришанова І.Ю.* Аналітичний огляд методів і засобів інформаційного пошуку в Semantic Web. *Проблеми програмування*. 2016. № 1. С. 51–72.
7. *Розушина Ю. В.* Семантичний пошук у Web на основі онтологій: розробка моделей, засобів і методів. Мелітополь: МДУ-ПУ ім. Б.Хмельницького. 2015. 291 с.
8. *Гладун А.Я., Розушина Ю.В.* Семантичні технології: принципи та практики. К.: ТОВ "АДЕФ-Україна". 2016. 308 с.
9. *Brachman R., Schmolze J.* An overview of the KL-ONE knowledge representation system. *Cognitive Science*, 1985. 9(2).
10. *Bobrow D., Winograd T.* An overview of KRL, a knowledge representation language. *Cognitive Science* 1(1) (1977).
11. *Antoniou G., Van Harmelen F.* Web ontology language: Owl. In *Handbook on ontologies*. Springer Berlin Heidelberg. 2004. P. 67–92.
12. *Gruber T.* A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*. 1993. N 5. P. 199–220.
13. *Cyganik R., Wood D., Lanthaler M.* RDF 1. 1 Concepts and Abstract Syntax. W3C Recommendation 25 February 2014. <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>.
14. *Rogushina J.* Means of the semantic search personification on base of ontological approach. *International Journal of Mathematical Sciences and Computing (IJMSC)*. 2016. Vol. 2. N 3. P. 1–20. – <http://www.mecspress.org/ijmsc/ijmsc-v2-n3/IJMSC-V2-N3-1.pdf>.



15. Rogushina J. Use of the Ontological Model for Personification of the Semantic Search. *International Journal of Mathematical Sciences and Computing (IJMSC)*. 2016. Vol. 2, N 1. <http://www.mecs-press.org/ijmsc/ijmsc-v2-n1/IJMSC-V2-N1-1.pdf>
16. Corby O., Dieng-Kuntz R., Faron-Zucker C. Querying the Semantic Web with Corese search engine. *Proc. ECAI-2004*, IOS Press, 2004. P. 705–709.
17. Finin T.W., Ding L., Pan R., Joshi A., Kolari P., Java A., Peng Y. Swoogle: Searching for knowledge on the Semantic Web. *Proc. AAAI-2005*, AAAI Press / MIT Press, 2005. P. 1682–1683.
18. Heflin J., Hendler J. A., Luke S. SHOE: A blueprint for the Semantic Web. D. Fensel, W. Wahlster, and H. Lieberman, editors. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, MIT Press, 2003. P. 29–63.
19. Kasneci G., Suchanek F.M., Ifrim G., Ramnath M., Weikum G. NAGA: Searching and ranking knowledge. *Proc. ICDE-2008*, IEEE Computer Society, 2008. P. 953–962.
20. Oren E., Gueret C., Schlobach S. Anytime query answering in RDF through evolutionary algorithms. *Proc. ISWC- 2008*, LNCS 5318, Springer, 2008. P. 98–113.
21. Buitelaar P., Eigner T., Declerck T. OntoSelect: A dynamic ontology library with support for ontology selection. *Proc. Demo Session at ISWC-2004*, 2004.
22. Cheng G., Ge W., Qu Y. Falcons: Searching and browsing entities on the Semantic Web. *Proc. WWW-2008*, ACM Press, 2008. P. 1101–1102.
23. Harth A., Hogan A., Delbru R., Umbrich J., O’Riain S., Decker S. SWSE: Answers before links. *Proc. Semantic Web Challenge 2007*, CEUR Workshop Proceedings 295. CEUR-WS.org, 2007.
24. Lei Y., Uren V. S., Motta E. SemSearch: A search engine for the Semantic Web. *Proc. EKAU-2006*, LNCS 4248, Springer, 2006. P. 238–245.
25. Tran T., Cimiano P., Rudolph S., Studer R. Ontology-based interpretation of keywords for semantic search. *Proc. ISWC/ASWC-2007*, LNCS 4825. Springer, 2007. P. 523–536.
26. Zenz G., Zhou X., Minack E., Siberski W., Nejd W. From keywords to semantic queries. Incremental query construction on the Semantic Web. *J. Web Sem.*, 7(3):, 2009. P. 166–176.
27. Cimiano P., Haase P., Heizmann J., Mantel M., Studer R. Towards portable natural language interfaces to knowledge bases – The case of the ORAKEL system. *Data Knowl. Eng.*, 65(2), 2008. P. 325–354.
28. Damjanovic D., Agatonovic M., Cunningham H. Natural language interface to ontologies: Combining syntactic analysis and ontology-based lookup through the user interaction. *Proc. ESWC-2010*, Part I, LNCS 6088, 2010. P. 106–120.
29. Fernandez M., Lopez V., Sabou M., Uren V. S., Vallet D., Motta E., Castells P. Semantic search meets the Web. *Proc. ICSC-2008*, 2008. P. 253–260.
30. Fazzingaa B., Lukasiewicz T. Semantic search on the Web. *Semantic Web – Interoperability, Usability, Applicability*. 2010. N 1. – P. 1–7. – [https://www.researchgate.net/profile/Thomas\\_Lukasiewicz/publication/220575552\\_Semantic\\_search\\_on\\_the\\_Web/links/0046351e94ee8994bd000000.pdf](https://www.researchgate.net/profile/Thomas_Lukasiewicz/publication/220575552_Semantic_search_on_the_Web/links/0046351e94ee8994bd000000.pdf).
31. Heflin J., Hendler J. A., Luke S. SHOE: A blueprint for the Semantic Web. In D. Fensel, W. Wahlster, and H. Lieberman, editors. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, MIT Press, 2003. P. 29–63. – <http://www.cse.lehigh.edu/~heflin/pubs/swbook03.pdf>.
32. Finin T.W., Ding L., Pan R., Joshi A., Kolari P., Java A., Peng Y. Swoogle: Searching for knowledge on the Semantic Web. *Proc. AAAI-2005*, AAAI Press / MIT Press, 2005. P. 1682–1683.
33. Corby O., Dieng-Kuntz R., Faron-Zucker C. Querying the Semantic Web with Corese search engine. *Proc. ECAI-2004*, IOS Press, 2004. P. 705–709.
34. Oren E., Gueret C., Schlobach S. Anytime query answering in RDF through evolutionary algorithms. *Proc. ISWC- 2008*, LNCS 5318, Springer, 2008. P. 98–113.
35. W3C. SPARQL Query Language for RDF, 2008. W3C Recommendation (15 January 2008). – <http://www.w3.org/TR/rdf-sparql-query/>.
36. Thomas E., Pan J.Z., Sleeman D. H. ONTOSEARCH2: Searching ontologies semantically. *Proc. OWLED-2007*, CEUR Workshop Proceedings 258. CEUR-WS.org, 2007.
37. Novacek V., Groza T., Handschuh S. CORAAL – Towards deep exploitation of textual resources in life sciences. *Proc. ALME-2009*, LNCS 5651, Springer, 2009. P. 206–215.
38. Kasneci G., Suchanek F.M., Ifrim G., Ramnath M., Weikum G. NAGA: Searching and

- ranking knowledge. Proc. ICDE-2008, IEEE Computer Society, 2008. P. 953–962.
39. Suchanek F.M., Kasneci G., Weikum G. Yago: A core of semantic knowledge. Proc. WWW-2007, ACM Press, 2007. P. 697–706.
  40. Buitelaar P., Eigner T., Declerck T. On-toSelect: A dynamic ontology library with support for ontology selection. Proc. Demo Session at ISWC-2004, 2004.
  41. Guha R.V., McCool R., Miller E. Semantic search. Proc. WWW-2003, ACM Press, 2003. P. 700–709.
  42. Lei Y., Uren V.S., Motta E. SemSearch: A search engine for the Semantic Web. Proc. EKAW-2006, LNCS 4248, Springer, 2006. P. 238–245.
  43. Tran T., Cimiano P., Rudolph S., Studer R. Ontology-based interpretation of keywords for semantic search. Proc. ISWC/ASWC-2007, LNCS 4825, Springer, 2007. P. 523–536.
  44. Tummarello G., Cyganiak R., Catasta M., Danielczyk S., Delbru R., Decker S. Sig.ma: Live views on the Web of data. Proc. WWW-2010, ACM Press, 2010. P. 1301–1304.
  45. Harth A., Umbrich J., Hogan A., Decker S. YARS2: A federated repository for querying graph structured data from the Web. Proc. ISWC/ASWC-2007, LNCS 4825, Springer, 2007. P. 211–224.
  46. YahooSearchBoss. <http://developer.yahoo.com/search/boss/>.
  47. Delbru R., Polleres A., Tummarello G., Decker S. Context dependent reasoning for semantic documents in Sindice. Proc. SSWS-2008, 2008.
  48. Zenz G., Zhou X., Minack E., Siberski W., Nejdil W. From keywords to semantic queries – Incremental query construction on the Semantic Web. J. Web Sem., 2009. N 7(3). P. 166–176.
  49. YahooISearchMonkey. – <http://developer.yahoo.com/searchmonkey>.
  50. Lopez V., Sabou M., Motta E. PowerMap: Mapping the real Semantic Web on the fly. Proc. ISWC-2006, LNCS 4273, Springer, 2006. P. 414–427.
  51. Lopez V., Pasin M., Motta E. AquaLog: An ontology- portable question answering system for the Semantic Web. Proc. ESWC-2005, LNCS 3532, Springer, 2005. P. 546–562.
  52. Castells P., Ferndndez M., Vallet D. An adaptation of the vector-space model for ontology-based information retrieval. IEEE Trans. Knowl. Data Eng. 2007. N 19(2). P. 261–272.
  53. Damljanovic D., Agatonovic M., Cunningham H. Natural language interface to ontologies: Combining syntactic analysis and ontology-based lookup through the user interaction. Proc. ESWC-2010, Part I, LNCS 6088, Springer, 2010. P. 106–120.
  54. Fazzinga B., Gianforme G., Gottlob G., Lukasiewicz T. Semantic Web search based on ontological conjunctive queries. Proc. FoIKS-2010, LNCS 5956, Springer, 2010. P. 153–172.
  55. D’Amato C., Esposito F., Fanizzi N., Fazzinga B., Gottlob G., Lukasiewicz T. Inductive reasoning and Semantic Web search. Proc. SAC-2010, ACM Press, 2010. P. 1446–1447.
  56. Mangol C. A survey and classification of semantic search approaches. International Journal of Metadata, Semantics and Ontologies. 2007. N 2(1). P. 23–34.
  57. Рогушина Ю. В. Семантический поиск как составляющая управления знаниями в Semantic Web. Материалы международной научно-технической конференции OSTIS-2012. Минск БГУИР. С. 239–244.
  58. Rogushina J. Use of the Ontological Model for Personification of the Semantic Search. International Journal of Mathematical Sciences and Computing(IJMSC). 2016. Vol. 2. N 1, – <http://www.mecs-press.org/ijmsc/ijmsc-v2-n1/IJMSC-V2-N1-1.pdf>
  59. Рогушина Ю.В. Разработка онтологической модели информационной потребности пользователя при семантическом поиске. Онтология проектирования. 2014. № 2(12). С. 61–82. – [http://agora.guru.ru/scientific\\_journal/files/Ontology\\_Of\\_Designing\\_2\\_2014\\_shot.pdf](http://agora.guru.ru/scientific_journal/files/Ontology_Of_Designing_2_2014_shot.pdf).

## References

1. Hendler J. Web 3.0: The dawn of semantic search. Computer, 2010, 43(1)/. P. 77–80.
2. Baeza-Yates R., A. Raghavan R. Next generation Web search // S. Ceri and M. Brambilla, editors, Search Computing, Springer, 2010. P. 11–23.
3. Lawrence S. Context in the Web Search – <http://citeser.nj.nec.com/lawrence00context.html>.
4. Janowicz K., Wilkes M., Lutz M. Similarity-based information retrieval and its role within

- spatial data infrastructures. Proc. GIScience-2008, Springer, 2008. P. 151–167.
5. Broder A. A taxonomy of web search, IBM Research, ACM SIGIR Forum archive, Vol. 36, Issue 2 (Fall 2002), P. 3–10.
  6. Grishanova I.Y. Analytic review of methods and tools of information search for Semantic Web. Problems in programming, 2016. N 1. P. 51–72. [in Ukrainian].
  7. Rogushina J.V. Semantic retrieval for Web on base of ontologies: design of models, tools and methods. Melitopol: Bogdan Hmelniysky MDUPU, 2015. 291 p. [in Ukrainian].
  8. Gladun A.Y., Rogushina J.V. Semantic technologies: principles and practices. – K.: ADEF-Ukraine, 2016. 308 p. [in Ukrainian]
  9. Brachman R., Schmolze J. An overview of the KL-ONE knowledge representation system. Cognitive Science, 1985, 9(2).
  10. Bobrow D., Winograd T. An overview of KRL, a knowledge representation language. Cognitive Science 1(1) (1977).
  11. Antoniou G., Van Harmelen F. Web ontology language: Owl. In Handbook on ontologies. Springer Berlin Heidelberg, 2004. P. 67–92.
  12. Gruber T. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition. 1993. N 5. P. 199–220.
  13. Cyganiak R., Wood D., Lanthaler M. RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation 25 February 2014. <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>.
  14. Rogushina J. Means of the semantic search personification on base of ontological approach. International Journal of Mathematical Sciences and Computing (IJMSC), Vol. 2, N 3. 2016. P. 1–20. – <http://www.mecs-press.org/ijmsc/ijmsc-v2-n3/IJMSC-V2-N3-1.pdf>.
  15. Rogushina J. Use of the Ontological Model for Personification of the Semantic Search. International Journal of Mathematical Sciences and Computing (IJMSC), Vol. 2, N 1, 2016. <http://www.mecs-press.org/ijmsc/ijmsc-v2-n1/IJMSC-V2-N1-1.pdf/>
  16. Corby O., Dieng-Kuntz R., Faron-Zucker C. Querying the Semantic Web with Corese search engine. Proc. ECAI-2004, IOS Press, 2004. – P. 705–709.
  17. Finin T. W., Ding L., Pan R., Joshi A., Kolari P., Java A., Peng Y. Swoogle: Searching for knowledge on the Semantic Web. Proc. AAAI-2005., AAAI Press / MIT Press, 2005. P. 1682–1683.
  18. Heflin J., Hendler J. A., Luke S. SHOE: A blueprint for the Semantic Web. D. Fensel, W. Wahlster, and H. Lieberman, editors. Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential, MIT Press, 2003. P. 29–63.
  19. Kasneci G., Suchanek F.M., Ifrim G., Ramanath M., Weikum G. NAGA: Searching and ranking knowledge. Proc. ICDE-2008, IEEE Computer Society, 2008. P. 953–962.
  20. Oren E., Gueret C., Schlobach S. Anytime query answering in RDF through evolutionary algorithms. Proc. ISWC- 2008, LNCS 5318, Springer, 2008. P. 98–113.
  21. Buitelaar P., Eigner T., Declerck T. OntoSelect: A dynamic ontology library with support for ontology selection. Proc. Demo Session at ISWC-2004, 2004.
  22. Cheng G., Ge W., Qu Y. Falcons: Searching and browsing entities on the Semantic Web. Proc. WWW-2008, ACM Press, 2008. P. 1101–1102.
  23. Harth A., Hogan A., Delbru R., Umbrich J., O’Riain S., Decker S. SWSE: Answers before links. Proc. Semantic Web Challenge 2007, CEUR Workshop Proceedings 295. CEUR-WS.org, 2007.
  24. Lei Y., Uren V. S., Motta E. SemSearch: A search engine for the Semantic Web. Proc. EKAW-2006, LNCS 4248, Springer, 2006. P. 238–245.
  25. Tran T., Cimiano P., Rudolph S., Studer R. Ontology-based interpretation of keywords for semantic search. Proc. ISWC/ASWC-2007, LNCS 4825, Springer, 2007. P. 523–536.
  26. Zenz G., Zhou X., Minack E., Siberski W., Nejdil W. From keywords to semantic queries. Incremental query construction on the Semantic Web. J. Web Sem., 7(3):,2009. P. 166–176.
  27. Cimiano P., Haase P., Heizmann J., Mantel M., Studer R. Towards portable natural language interfaces to knowledge bases — The case of the ORAKEL system. Data Knowl. Eng., 65(2), 2008. P. 325–354.
  28. Damjanovic D., Agatonovic M., Cunningham H. Natural language interface to ontologies: Combining syntactic analysis and ontology-based lookup through the user interaction. Proc. ESWC-2010, Part I, LNCS 6088, 2010. P. 106–120.
  29. Fernandez M., Lopez V., Sabou M., Uren V. S., Vallet D., Motta E., Castells P. Semantic

- search meets the Web. Proc. ICSC-2008, 2008. – P. 253–260.
30. Fazzingaa B., Lukasiewicz T. Semantic search on the Web. *Semantic Web – Interoperability, Usability, Applicability*, N 1, 2010. P. 1–7. [https://www.researchgate.net/profile/Thomas\\_Lukasiewicz/publication/220575552\\_Semantic\\_search\\_on\\_the\\_Web/links/0046351e94ee8994bd000000.pdf](https://www.researchgate.net/profile/Thomas_Lukasiewicz/publication/220575552_Semantic_search_on_the_Web/links/0046351e94ee8994bd000000.pdf).
  31. Heflin J., Hendler J. A., Luke S. SHOE: A blueprint for the Semantic Web. In D. Fensel, W. Wahlster, and H. Lieberman, editors. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, MIT Press, 2003. P. 29–63. <http://www.cse.lehigh.edu/~heflin/pubs/swbook03.pdf>.
  32. Finin T. W., Ding L., Pan R., Joshi A., Kolari P., Java A., Peng Y. Swoogle: Searching for knowledge on the Semantic Web. Proc. AAAI-2005, AAAI Press / MIT Press, 2005. P. 1682–1683.
  33. Corby O., Dieng-Kuntz R., Faron-Zucker C. Querying the Semantic Web with Corese search engine. Proc. ECAI-2004, IOS Press, 2004. P. 705–709.
  34. Oren E., Gueret C., Schlobach S. Anytime query answering in RDF through evolutionary algorithms. Proc. ISWC- 2008, LNCS 5318, Springer, 2008. P. 98–113.
  35. W3C. SPARQL Query Language for RDF, 2008. W3C Recommendation (15 January 2008). – <http://www.w3.org/TR/rdf-sparql-query/>.
  36. Thomas E., Pan J. Z., Sleeman D. H. ONTOSEARCH2: Searching ontologies semantically. Proc. OWLED-2007, CEUR Workshop Proceedings 258. CEUR-WS.org, 2007.
  37. Novacek V., Groza T., Handschuh S. CORAAL – Towards deep exploitation of textual resources in life sciences. Proc. ALME-2009, LNCS 5651, Springer, 2009. P. 206–215.
  38. Kasneci G., Suchanek F. M., Ifrim G., Ramanath M., Weikum G. NAGA: Searching and ranking knowledge. Proc. ICDE-2008, IEEE Computer Society, 2008. P. 953–962.
  39. Suchanek F. M., Kasneci G., Weikum G. Yago: A core of semantic knowledge. Proc. WWW-2007, ACM Press, 2007. P. 697–706.
  40. Buitelaar P., Eigner T., Declerck T. OntoSelect: A dynamic ontology library with support for ontology selection. Proc. Demo Session at ISWC-2004, 2004.
  41. Guha R. V., McCool R., Miller E. Semantic search. Proc. WWW-2003, ACM Press, 2003. – P. 700–709.
  42. Lei Y., Uren V. S., Motta E. SemSearch: A search engine for the Semantic Web. Proc. EKAW-2006, LNCS 4248, Springer, 2006. P. 238–245.
  43. Tran T., Cimiano P., Rudolph S., Studer R. Ontology-based interpretation of keywords for semantic search. Proc. ISWC/ASWC-2007, LNCS 4825, Springer, 2007. P. 523–536.
  44. Tummarello G., Cyganiak R., Catasta M., Danielczyk S., Delbru R., Decker S. Sig.ma: Live views on the Web of data. Proc. WWW-2010, ACM Press, 2010. P. 1301–1304.
  45. Harth A., Umbrich J., Hogan A., Decker S. YARS2: A federated repository for querying graph structured data from the Web. Proc. ISWC/ASWC-2007, LNCS 4825, Springer, 2007. P. 211–224.
  46. YahooSearchBoss. – <http://developer.yahoo.com/search/boss/>.
  47. Delbru R., Polleres A., Tummarello G., Decker S. Context dependent reasoning for semantic documents in Sindice. Proc. SSWS-2008, 2008.
  48. Zenz G., Zhou X., Minack E., Siberski W., Nejd W. From keywords to semantic queries – Incremental query construction on the Semantic Web. *J. Web Sem.*, N 7(3), 2009. P. 166–176.
  49. YahooISearchMonkey. – <http://developer.yahoo.com/searchmonkey>.
  50. Lopez V., Sabou M., Motta E. PowerMap: Mapping the real Semantic Web on the fly. Proc. ISWC-2006, LNCS 4273, Springer, 2006. P. 414–427.
  51. Lopez V., Pasin M., Motta E. AquaLog: An ontology- portable question answering system for the Semantic Web. Proc. ESWC-2005, LNCS 3532, Springer, 2005. P. 546–562.
  52. Castells P., Fernandez M., Vallet D. An adaptation of the vector-space model for ontology-based information retrieval. *IEEE Trans. Knowl. Data Eng.*, N 19(2), 2007. P. 261–272.
  53. Damjanovic D., Agatonovic M., Cunningham H. Natural language interface to ontologies: Combining syntactic analysis and ontology-based lookup through the user interaction. Proc. ESWC-2010, Part I, LNCS 6088, Springer, 2010. P. 106–120.
  54. Fazzingaa B., Gianforme G., Gottlob G., Lukasiewicz T. Semantic Web search based on ontological conjunctive queries. Proc.

- FoIKS-2010, LNCS 5956, Springer, 2010. P. 153–172.
55. D'Amato C., Esposito F., Fanizzi N., Fazzino B., Gottlob G., Lukasiewicz T. Inductive reasoning and Semantic Web search. Proc. SAC-2010, ACM Press, 2010. P. 1446–1447.
56. Mangol C. A survey and classification of semantic search approaches. International Journal of Metadata, Semantics and Ontologies, N 2(1), 2007. P. 23–34.
57. Rogushina J.V. Semantic retrieval as a component of the Semantic Web knowledge management. Proc. of OSTIS-2012, Minsk . P. 239–244. [in Russian].
58. Rogushina J. Use of the Ontological Model for Personification of the Semantic Search // International Journal of Mathematical Sciences and Computing(IJMSC), Vol. 2, N 1, 2016. <http://www.mecs-press.org/ijmsc/ijmsc-v2-n1/IJMSC-V2-N1-1.pdf>.
59. Rogushina J.V. Design of ontological model of user information need in semantic retrieval. Ontology of designing. P. 61–82. [http://agora.guru.ru/scientific\\_journal/files/Ontology\\_Of\\_Designing\\_2\\_2014\\_shot.pdf](http://agora.guru.ru/scientific_journal/files/Ontology_Of_Designing_2_2014_shot.pdf). [in Russian].

***Про автора:***

*Рогущина Юлія Віталіївна*,  
кандидат фізико-математичних наук,  
старший науковий співробітник.  
Кількість наукових публікацій в  
українських виданнях – 120.  
Кількість наукових публікацій в  
зарубіжних виданнях – 28.  
Індекс Хірша – 10,  
<http://orcid.org/0000-0001-7958-2557>.

***Місце роботи автора:***

Інститут програмних систем  
НАН України,  
03181, Київ-187,  
проспект Академіка Глушкова, 40.  
Тел.: 066 550 1999.  
E-mail: [ladamandraka2010@gmail.com](mailto:ladamandraka2010@gmail.com)

Одержано 07.12.2016