

ВИКОРИСТАННЯ ОНТОЛОГІЙ ДЛЯ ПЕРСОНІФІКОВАНОГО ПОШУКУ ЗНАНЬ У ПРИРОДНОМОВНИХ ТЕКСТІВ

Запропонований у роботі підхід до персоніфікації пошуку інформаційних ресурсів та інформаційних об'єктів, що базується на побудові та використанні тезаурусу задачі користувача, дозволяє використовувати знання щодо предметної області пошуку та структури інформаційних об'єктів, представлені за допомогою відповідних онтологій. Наведені визначення семантичного пошуку, його суб'єктів та компонентів дозволяють більш чітко формулювати проблеми, пов'язані з пошуком інформації у відкритому середовищі Web. Програмна реалізація запропонованого підходу підтверджує ефективність його практичного використання.

Ключові слова: семантичний пошук, інформаційний об'єкт, онтологія, тезаурус задачі.

Вступ

В процесі розвитку суспільства з великою швидкістю збільшуються обсяги інформації, що обробляється, ускладнюється її структура та методи обробки.

Для сучасного етапу розвитку інформаційних технологій (ІТ) характерні наступні тенденції: 1) все більше інформаційних систем (ІС) стають інтелектуальними та використовують знання; 2) переважна частина ІС працює у відкритому середовищі (Web, локальні та корпоративні мережі, хмари тощо) і орієнтовані на отримання відомостей із зовнішніх інформаційних ресурсів (ІР), не залежних від їх розробників; 3) все більше поширення отримують різноманітні мобільні пристрої для обробки інформації, що відповідають специфіці користувачів.

З цього випливає важливість проблеми пошуку знань та використання знань у пошуку даних, а також персоніфікація такого пошуку.

Актуальність проблеми пошуку

Значна частина знань, накопичених у результаті розвитку людського суспільства і різних предметних областей, міститься в документах у вигляді природномовного тексту. У Web представлення також велика кількість мультимедійних документів, і з ростом поширення різних підключених до Інтернет пристроїв їхній обсяг зростає значно швидше, але в них

міститься набагато менше корисної інформації (наприклад, різні фотографії і відеофайли, як правило, цікаві лише тим, хто їх знімав).

У структурованому вигляді (онтології, метаописи, семантична розмітка і т. п.) представлено набагато менше відомостей. Якщо ж порівнювати зусилля, необхідні для витягу знань (змісту), то природномовні ресурси обробляти значно легше, ніж мультимедійні, а сама форма представлення забезпечує первинний пошук релевантних проблемі ресурсів (наприклад, пошук за ключовими словами у тексті) з подальшою семантизацією [1, 2]. Крім того, при розпізнаванні мультимедійних ІР спочатку значна частина інформації перетворюється на природномовний текст.

Це обумовлює важливість розвитку методів аналізу природномовних документів, що включають пошук релевантних предметній області документів, розпізнавання їхнього змісту і поповнення відповідних онтологій [3–5]. Це потребує створення засобів отримання, збереження, пошуку та використання знань з урахуванням таких властивостей середовища Web, як динамічність та гетерогенність. Крім того, необхідно обрати засоби інтероперабельного представлення знань, які мають достатні функціональні можливості та надають можливість для їх повторного використання та обробки як комп'ютерними програмами, так і людьми.

Для цього виникає потреба у розробці моделі інформаційного середовища сучасного Web; моделі користувача, що відображає його інформаційні потреби, предметну область, яка його цікавить, та проблеми, які він прагне вирішити; і моделі інформаційних ресурсів, що має відображати не тільки їх формальні властивості, але й семантику.

Процеси глобальної інформатизації орієнтовані на побудову та інтероперабельне використання міждисциплінарних знань. Але ефективне використання знань потребує розвитку відповідних засобів їх знаходження та подання. Це вимагає розвитку інженерії знань і засобів менеджменту знань.

Відносно новим напрямком у цій сфері є онтологічний інжиніринг, що забезпечує повторне й інтероперабельне застосування накопичених у суспільстві знань [6–8]. Онтології використовуються в системах обробки знань для їхнього структурування й інтеграції [9]. Тому актуальні питання автоматизованого створення і поповнення онтологій на основі гетерогенних і динамічних ресурсів Web, їхньої інтеграції і співставлення, а також створення методів логічного виведення на них [10].

Системи семантичного пошуку

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).

Можна розглядати семантичний пошук як розвиток традиційного інформаційного пошуку, в якому з метою підвищення пертинентності пошуку (тобто для більш ефективного задоволення інформаційних потреб користувача) використовується обробка знань, що стосуються як самого користувача та його інформаційних потреб (персоніфікація пошуку), так і про інформаційні ресурси, серед яких здійснюється пошукова процедура. Системи семантичного пошуку – це певна інтелектуальна надбудова над традиційними інформаційно-пошуковими системами (ІПС) – як

загального призначення, так і спеціалізованими.

Система семантичного пошуку (ССП) – це інформаційна система, що забезпечує пошук та розпізнавання інформаційних об'єктів (ІО) різних типів з використанням знань для співставлення запиту з наявними інформаційними ресурсами на семантичному рівні.

Відкрита ССП – це ССП, в якій використовуються не тільки внутрішні, але й зовнішні бази знань, структура та контент яких не залежать від розробника ССП.

Інформаційний об'єкт – модель об'єкта предметної області (ПрО) в інформаційному просторі, яка визначає структуру, атрибути, обмеження цілісності і, можливо, поведження цього об'єкта.

Результатом семантичного пошуку може бути як здобуття інформації щодо ІО, неявно присутньої у 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.

Те, які саме знання використовуються, як вони представлені і як вони обробляються, залежить як від специфіки розроблювальної ІПС, так і від концепції, обраної її розроблювачами, але в загальному випадку результати семантичного пошуку – множина з n елементів є функцією від запиту користувача, індексу ІПС та знань, що містяться в базі знань (БЗ) ІПС:

$$I_s = \{i_j, j = \overline{1, n}\} = f(z, DB_{inc}, KB_{inc}) .$$

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 ІПС. Тому результати пошуку в Web є функцією ще й від вмісту зовнішніх БЗ:

$$I_{web_s} = \{j, j = \overline{1, n}\} = f \left(\begin{matrix} z, DB_{inc}, KB_{inc}, \\ \{KB_{web_k}, k = \overline{1, m}\} \end{matrix} \right).$$

При рішенні задачі семантичного пошуку, пов'язаного з розпізнаванням набору складних ІО, виникає ряд різних видів проблем, для опису яких необхідно визначити використовувані при цьому терміни, зокрема, визначити, які відомості є результатом такого пошуку, які – його умовами. Тому потрібна класифікація ІО та пов'язаних з ними пошукових ситуацій.

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.

Для сучасних семантичних Web-застосувань характерні ІО зі складною структурою, які пов'язані з певними об'єктами реального або віртуального світу (програмні агенти, Web-сервіси, семантично розмічені ресурси, елементи Web of Things, онтологічні описи тощо). Оцінка складності структури атомарних елементів залежить від їхньої кількості і кількості зв'язків між ними.

Приміром, ПС, що впорядковує знайдені за запитом користувача документи, враховуючи його персональні властивості, і використовує для цього інформацію з історії взаємодії з користувачем, менш інтелектуальна, ніж ПС, що впорядковує знайдені за запитом користувача документи, використовуючи для формалізації сфери інтересів користувача онтологію відповідної ПрО, та виконує семантичну розмітку знайдених документів термінами з онтології ПрО [11].

Як показує аналіз публікацій, один з перспективних підходів до завдання контексту пошуку ґрунтується на онтологіях, що містять перелік основних термінів,

зв'язки між ними і правила виведення (так, у проєкті Semantic Web, спрямованому на аналіз семантики ІР, саме онтологічний підхід є основою для подання знань про різні ПрО).

Онтологія ІО O_{IO} – онтологічна структура, що містить клас ІО $t_{IO} \in T_{IO}$ 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 ІО, який його цікавить, йому потрібно послатися на клас довільної формально описаної онтології.

Таким чином, ІО можна розглядати як клас певної онтології, який має набір характеристик, які описують його структуру і можливі зв'язки з іншими класами й екземплярами класів. Для більш точного визначення проблеми доцільно ввести кілька визначень

Екземпляр ІО – екземпляр якогонебудь підкласу ІО відповідної онтології, який можна однозначно ідентифікувати і який має власне ім'я.

Ситуація – непорожня множина ІО та екземплярів 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 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 працювали 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. Персональні ситуації можуть мати стандартизований опис, в якому відомості щодо користувача є параметрами.

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 автора В” значно простіший для користувача запит “знайти в 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 й їх екземплярів 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.

Постановка задачі

Щоб підвищити ефективність семантичного пошуку, необхідно забезпечити його *персоніфікацію*, тобто використання знань про інформаційні потреби, сферу інтересів та здатність до сприйняття інформації окремих користувачів. Тому виникає потреба у розробці формалізованої моделі інформаційних потреб користувача, у засобах її поповнення інформацією та у методах її співставлення з наявними інформаційними ресурсами.

Використання онтологій для персоніфікованого аналізу природномовних текстів

Щоб використовувати онтологічні знання в процесі семантичного пошуку, потрібно забезпечити як механізми автоматизованого створення онтологічних моделей предметних областей та інформаційних потреб, так і методи їх співставлення. Пропонується в якості такого механізму використовувати тезаурус задачі, який відображає поточні інформаційні потреби користувача на основі онтології ПрО, обраної користувачем.

У загальному випадку тезаурус – це словник основних понять мови, що позначаються окремими словами чи словосполученнями, з визначеними семантичними зв'язками між ними [12]. Тезаурус можна розглядати як окремий випадок онтології [13]. Лексика тезауруса включає

множину слів і/чи множину фраз [14–16]. Види підтримуваних семантичних зв'язків між ними можуть бути залежними чи незалежними від конкретної ПрО. Звичайно такі зв'язки визначають синоніми, омоніми, антоніми понять мови, підтримують між ними відношення виду «ціле – частина», «рід – вид», «використовується для», «працює в» тощо. Надалі в ССП будемо розуміти під тезаурусом задачі наступне:

Тезаурус задачі – це множина термінів ПрО, необхідних для опису та вирішення задачі, для якої користувач намагається за допомогою ССП знайти певну інформацію. Для кожного з них може бути визначена їх вага, що дозволяє охарактеризувати важливість та пертинентність терміну для поточної задачі, та онтологія, з якої імпортовано відповідний термін.

Для кожного тезаурусу задачі існує хоча б одна онтологія ПрО, на якій він базується. В такому тезаурусі онтологічні зв'язки між термінами не відображаються явно, проте вони використовуються в процесі побудови тезаурусу задачі за онтологією ПрО. Приміром, можна побудувати тезаурус, який містить визначену підмножину термінів $T_0 \subseteq X$ та терміни, пов'язані з ними відношенням $r \in R$.

Тезаурус задачі є персоніфікованим, тобто для рішення однієї й тої ж задачі різні користувачі можуть використовувати тезауруси, які значно різняться один від одного. Це залежить не тільки від тих природних мов, на які розраховує користувач, і не тільки від використаних онтологій, але й від індивідуальних переконань та переваг користувача в обраній ПрО.

Слід відмітити, що побудова тезаурусу задачі є відносно складною та трудомісткою, тому доцільно виконувати цю операцію тільки в тих випадках, коли задача, що вирішується, відноситься до сфери постійних та складних інформаційних потреб користувача, а пошук відомостей для задоволення відповідної інформаційної потреби має враховувати багато умов та обмежень. Приміром, до таких задач може віднести пошук нової наукової літератури або інструментальних засобів з певного питання, що має відкрити

мити визначені в тезаурусі напрямки досліджень.

Недоцільно будувати тезаурус задачі для одноразових запитів у сфері, де користувач не є експертом і тому не може сам враховувати достатню кількість зовнішніх знань. В такому випадку зусилля з побудови тезаурусу будуть більшими від отриманого ефекту [17].

Інформацію, що міститься у тезаурусі задачі, можна поділити на дві частини – операційну та службову. Операційна інформація безпосередньо використовується у семантичному пошуку за допомогою такого тезаурусу, а службова інформація описує шляхи побудови тезаурусу, його інформаційні джерела, та може використовуватися для подальших операцій з цим тезаурусом (приміром, якщо відомо, з якої онтології експортовано певний термін, то можна експортувати й його підкласи або екземпляри).

Простий тезаурус задачі – це тезаурус, який базується на термінах однієї онтології ПрО.

Складений тезаурус задачі – це тезаурус, який базується на термінах двох або більш онтологій ПрО.

Складений тезаурус може бути побудований як поєднання двох або більше простих тезаурусів. Слід відмітити, що складений тезаурус задачі може містити терміни з однаковими іменами, отримані з різних онтологій, які не будуть еквівалентними.

Формальна модель простого тезаурусу задачі $Th = \langle T, R_{Th}, O \rangle$, де T – множина термінів, $T \subseteq X$ а $R_{Th} \subseteq R$ – множина відношень між цими термінами, що використовувалися для побудови тезаурусу. Множини T й R скінчені.

Формальна модель складеного тезаурусу задачі

$$Th = \langle T = \{ \langle x_{i,j} \in X_i, O_i \rangle \mid i = \overline{1, n}, j = \overline{1, m_i} \},$$

$$R_{Th} = \bigcup_{i=1}^n R_{Th_i}, O_1, \dots, O_n \rangle,$$

де T – множина пар термінів онтологій та посилань на відповідну онтологію, R_{Th} –

об'єднання множин відношень між термінами онтологій, що використовувалися для побудови тезауруса, та перелік всіх онтологій, що застосовуються для його побудови.

Перехід від онтологій до тезаурусів дозволяє значно спростити структуру знань, що обробляються, забезпечуючи прийнятну для практичної реалізації швидкість оброблення. Але використання оригінальних онтологій ПрО як основи для побудови тезауруса забезпечує доступність всіх наявних знань щодо ПрО, з яких користувач (вручну або автоматизовано) може обрати саме ту частку, що безпосередньо пов'язана з конкретною задачею.

Крім того, тезаурус задачі можна розглядати як спрощену та персоналізовану онтологію, що характеризує термінологічну основу поточної задачі користувача.

Природномовний опис задачі може містити:

– набір слів та словосполучень, які користувач вважає важливими для задачі (якщо користувач здатний самостійно побудувати весь такий набір, то можна вважати проблему побудови тезауруса вирішеною, але у більшості випадків цей набір треба поповнювати та фільтрувати);

– опис ІО, що входять до складу ситуації, що має стати результатом пошуку;

– природномовний текст, що характеризує цю ситуацію (постановку задачі, опис проекту; технічне завдання тощо);

– набір природномовних текстів, який користувач вважає пертинентними задачі (з цієї інформації потрібно здобути інформацію щодо того, які саме поняття ПрО є значущими для поточної задачі).

Таким чином, для побудови тезауруса задачі потрібно вирішити наступну задачу – побудувати метод, який забезпечує співставлення термінів онтології з природномовним текстом.

Повністю автоматизувати цю задачу неможливо, тому що в процесі її вирішення потрібно використовувати неформалізовані знання та переконання користувача щодо того, що саме його цікавить і може бути корисним для його проблеми. Але можливо спростити цю роботу, част-

ково автоматизувавши пошук потенційно цікавих фрагментів.

Алгоритм побудови тезауруса задачі

Вхідними даними для побудови тезауруса задачі є природномовний опис задачі, онтологія ПрО, до якої користувач відносить свою задачу. Якщо результатом пошуку має стати не документ, а ситуація, то доцільно також використовувати онтологію ІО, в якій містяться відомості щодо властивостей, структури, елементів та екземплярів такого ІО. У випадку обробки конкретизованої ситуації наявність онтології ІО дозволяє чітко структурувати умови користувача [18].

Якщо обробляється персональна ситуація, то для надання знань щодо користувача може використовуватися внутрішня онтологія ССП та її клас “користувач”. Це позбавляє користувача від введення відомостей, вже відомих ССП щодо його індивідуальних властивостей. При цьому можуть оброблятися як формальні характеристики, такі як вік, місце проживання, так і семантичні, такі як сфера інтересів або компетентність в певній ПрО.

У деяких складних випадках для опису ІО можуть використовуватися не одна, а кілька онтологій.

Якщо якась одна онтологія не формалізує ПрО задовільно до задачі користувача, то можна використовувати сукупність онтологій, вважаючи їх незалежними, а їх множини термінів – такими, що не перетинаються (якщо дві онтології містять терміни з однаковими назвами, то ці терміни вважаються різними). Крім того, в онтології (чи множині онтологій) не всі елементи є корисними для конкретної задачі.

Етап 1. Відбір множини пертинентних онтологій.

Цей відбір виконується користувачем на основі його знань та переконань. Як правило, вибирати потрібно серед тих онтологій, які запропоновані в ССП. Це забезпечує не тільки якість онтологій для відображення знань ПрО, але й їх придатність для обробки засобами системи. У

більш загальному випадку можна шукати онтології у різноманітних репозиторіях, де властивості онтологій та їх домени охарактеризовані на семантичному рівні [19, 20].

При цьому можуть використовуватися як онтології ПрО пошуку, так і онтології Ю. Приміром, якщо користувач прагне знайти відомості щодо Web-сервісу для індуктивного здобуття знань з даних та їх візуалізації, то він може відібрати онтологію ПрО “Data Mining” та онтологію Ю “Web-сервіси”.

Для цього необхідно виконати аналіз пошукової ситуації, визначити, які Ю в ній використовуються та до яких ПрО належать умови пошуку.

Етап 2. Відбір у множині онтологій термінів, пертинентних задачі.

Такий відбір може виконуватися користувачем безпосередньо або будуватися за певними правилами (приміром, підкласи обраного класу чи його екземпляри, поняття на визначеній семантичній відстані від обраного поняття). За різними онтологіями можна побудувати кілька простих тезаурусів задачі а потім поєднати їх у складений тезаурус.

На цьому етапі створюється початковий варіант тезауруса задачі, який надалі потрібно вдосконалювати та поповнювати лінгвістичною інформацією, потрібною для співставлення з природномовними текстами.

Етап 3. Розробка лінгвістичної БЗ (ЛБЗ) тезауруса задачі, яка має містити фрагменти природномовного тексту, що відповідають його термінам.

Спочатку для кожного терміну тезауруса задачі до ЛБЗ додаються всі словоформи цього терміну, що відповідають різним відмінкам та множині слова або словосполучення, що використовується як його ім'я в обраній природній мові. Для цього можуть використовуватися словозміни (флексії), тобто системні засоби утворення різних форм того самого слова відповідно до його синтаксичних пов'язань з іншими словами в реченні або словосполученні без зміни його лексичного значення. Для цього в українській мові, приміром, можуть використовуватися за-

кінчення змінних морфем-афіксів і постфіксів, зміни основи слова для слів з внутрішньою флексією (“стіл”–“столи”), зміни суфікса (“швикий”–“швидший”) або зміни префікса (“більший”–“найбільший”) та сполучення флексійної форми слова з прийменником (“атрибут”–“з атрибутом”) [20].

Словозміна застосовується при відмінюванні іменника, прикметника, займенника й числівника за відмінками, а в них відмінками, числами й родами та дієслова за відмінками, способами, часами, особами, числами, родами (в минулому часі й умовному способі), видами й станами тощо.

Побудувати такі конструкції на основі відповідних знань щодо правил словозміни досить просто для окремих слів, а для словосполучень потребує додаткового лінгвістичного аналізу, метою якого є визначення головного слова, що змінюється, та поділу пов'язаних з ним слів на ті, що змінюються, та ті, що залишаються незмінними за правилами природної мови. Наприклад, у словосполученні “онтологічна модель предметної області” основним є друге слово, змінюваним – перше, а незмінюваними – третє та четверте.

Інший метод здобуття словоформ для термінів тезауруса задачі базується на використанні Wiki-ресурсів [21, 22]. Такі ресурси можуть не тільки допомогти у формуванні онтології ПрО, що відповідає потребам користувача [23], але й стати джерелом лінгвістичної інформації. Якщо вдається знайти сторінку Вікіпедії, що відповідає певному терміну тезауруса, то з коду сторінок, що на неї посилаються можна імпортувати конструкції-посилання, пов'язані з цим терміном: [[ім'я сторінки|словоформа терміну, доречна в тексті]]. Приміром, [[технічна інформатика|технічній інформатиці]], [[Глушков_В_М|Віктора Михайловича Глушкова]]. Додаткову інформацію можна здобувати і з категоризації сторінок. Крім того, для семантичних Wiki-ресурсів можна аналізувати семантичні властивості термінів.

На наступному кроці до основного слова чи словосполучення, що визначає термін тезауруса, додаються його синоні-

мічні варіанти, актуальні для визначеної ПрО. Приміром, для словосполучення “онтологічна модель предметної області” – це словосполучення “онтологічний опис ПрО” та “онтологія домену”. Слід відмітити, що для побудови синонімічних виразів використовуються не тільки відомості щодо природної мови, але й знання ПрО (приміром, експортовані з онтології ПрО відношення “є еквівалентним класом”).

Ще один крок – переклад словосполучень, що відповідають термінам тезауруса, на інші мови, які знає користувач. Наприклад, для словосполучення “онтологічна модель предметної області” – це словосполучення “domain ontology” англійською мовою та “онтологическая модель предметной области” – російською. До отриманих перекладів застосовуються аналогічні операції побудови слівформ, що відповідають правилам обраних мов.

Наступний крок – варіанти порядку слів у словосполученні та його елементи, які припустимі в рамках кожної з обраних природних мов. Приміром, екземпляр певного класу “Іванов Олександр” може описуватися також як “Олександр Іванов” або “Іванов О.”.

Остання операція цього етапу – користувач може вручну додати або видалити певні словосполучення, які він вважає відповідними термінам тезауруса. Після цього тезаурус задачі зберігається та може використовуватися для семантичного аналізу довільних природномовних текстів.

Етап 4. Анотування тезауруса задачі.

На цьому етапі доцільно створити опис тезауруса задачі, який описує як його формальні властивості (кількість термінів, оброблювані природні мови), так і семантичні характеристики – ПрО (через посилання на відповідні онтології), призначення, проблеми, для рішення яких він може застосовуватися. Це забезпечує повторне використання знань, що відображені в такому тезаурусі.

Етап 5. Вдосконалення тезауруса задачі.

У багатьох випадках користувач може застосовувати для власних цілей раніше створені тезауруси задач – як розроблені ним самим, так і розроблені іншими користувачами (якщо ті надають свої тезауруси у відкритий доступ).

Це доцільно у тих випадках, коли нова задача користувача є дещо зміненим варіантом задачі, що вирішувалася раніше (в такому випадку можна вручну відредагувати тезаурус, додавши чи видаливши кілька термінів). Якщо ж задача є узагальненням або уточненням попередніх задач, то доцільно застосовувати теоретико-множинні операції над тезаурусами, такі як перетин, об’єднання та доповнення. Приміром, до тезауруса ПрО “Онтологічний аналіз” в одному випадку можна додати тезаурус “Логічне виведення”, а в іншому – “Візуалізація знань”. Такий підхід значно зменшує час на модифікацію тезаурусів задач, але потребує створення досить докладних анотацій створюваних тезаурусів.

Етап 6. Визначення ваги термінів тезауруса задачі.

Після того, як формування множини термінів, що входять до складу тезауруса задачі, закінчується, користувачеві необхідно вказати, яку вагу для поточної задачі має кожен термін [24].

У найпростішому випадку (за замовчанням) можна вважати, що всі терміни мають однакову вагу, що дорівнює одиниці. Але на практиці зазвичай користувачеві зрозуміло, що деякі терміни значно важливіші за інші. Крім того, існують ситуації, коли наявність певного терміну в ІР свідчить про його низьку релевантність для задачі (хоча сам термін може бути важливим для ПрО в цілому). Приміром, користувачеві потрібно знайти мови подання онтологій, які не базуються на XML. Тоді термін “XML” буде присутній в онтології задачі, але з негативною вагою.

Такий тезаурус задачі надалі можна використовувати для того, щоб знаходити ІР, що найбільш пертинентні задачі користувача (ці ІР можуть бути знайдені довільною зовнішньою інформаційно-пошуковою системою (ІПС) за набором ключових слів, які вводить користувач для

попереднього опису своєї інформаційної потреби, або ж набір IP може запропонувати сам користувач – як з зовнішніх джерел, таких як тематична бібліотека або сайт, так і з внутрішніх архівів).

Надалі саме з цих IP слід здобувати знання, що будуть корисні для побудови пошукової ситуації, але методи здобуття таких знань знаходяться поза розглядом даної роботи.

Використовувати тезаурус задачі можна двома способами: 1) для кожного з аналізованих IP будувати тезаурус, а потім порівнювати його з тезаурусом задачі, або 2) безпосередньо порівнювати тезаурус задачі з контентом кожного IP, використовуючи для цього співставлення знання, накопичені у ЛБЗ.

Перший підхід доцільно використовувати, якщо IP з відносно невеликої фіксованої множини багаторазово аналізуються на пертинентність потребі користувача в фіксованій ПрО та для однієї або подібних задач. Приміром, це можна використовувати для індексації та швидкого пошуку у власних архівах або у власній електронній бібліотеці. Тоді можна будувати тезауруси всіх IP для певної ПрО і використовувати їх багаторазово.

Надалі будемо розуміти під *зв'язаним тезаурусом задачі* спрощений варіант складеного тезауруса, в якому інформація щодо походження кожного терміну не прив'язується до певної онтології.

$$Th_w = \langle X, W, \{\cup O_z\} \rangle,$$

таке, що $x_i \in X$, $i = \overline{1, n}$ – термін тезауруса задачі, w_i – вага цього терміну в тезаурусі задачі, а $\{\cup O_z\}$ – множина онтологій, які використовувалися при побудові цього тезауруса. X є скінченою непорожньою множиною.

Кожному такому тезаурусу задачі відповідає ЛБЗ L така, що для кожного $x_i \in X$ існує $L_i = \{l_{ij}, j = \overline{1, m_i}\}$ – непорожня скінчена множина фрагментів природномовних текстів, що відповідає терміну тезауруса задачі $x_i \in X$. Ці множини не перетинаються:

$$\forall i \neq k L_i \cap L_k = \emptyset.$$

Якщо ж інформація за ключовими словами імпортується зовнішньою ПС з інформаційного простору Web, то через динамічність цього середовища в будь-якому разі потрібно кожного разу заново аналізувати контент кожного IP, і тому більш придатним є другий підхід, який детальніше розглянутий далі.

Використання тезауруса задачі для фільтрації IP

Алгоритм персоніфікованої семантичної фільтрації IP за допомогою тезауруса задачі складається з наступних кроків:

Користувач вводить запит, ідентифікуючи свою інформаційну потребу за допомогою набору ключових слів. Потрібно відмітити, що на цей момент система семантичного пошуку (ССП) має вже певні відомості про цього користувача, отримані в процесі взаємодії з ним та відображені в його моделі [25, 26]. Приміром, для цього може використовуватися онтологічна модель, що персоніфікує взаємодію користувача з різними інтелектуальними застосуваннями [27, 28].

Запит передається до зовнішнього пошукового механізму, який відбирає з набору IP ті, що містять введені ключові слова (у виродженому випадку набір ключових слів – порожня множина, і подальшому аналізу підлягає весь набір приступних IP). Результати виконання запиту – n посилань на IP і їхні короткі описи

$$I = \{ \langle Ref_r, D_r \rangle \}, r = \overline{1, p},$$

де Ref_r – ідентифікатор (приміром, http-адреса для IP, знайдених в Web) відповідного IP, а d_r – інформація про цей IP, що зовнішня ПС надає користувачу у відповідь на запит.

Якщо множина I не порожня, тобто вдалося знайти хоча б один IP ($p \geq 1$), то потрібно встановити порядок, в якому пропонувати користувачу відомості про знайдений IP. Тоді для всіх IP з цієї множини Ref_r , $r = \overline{1, p}$ потрібно виконати

наступну процедуру – спробувати знайти в них фрагменти тексту з множини L , що відповідають кожному з термінів тезауруса задачі. Можна проводити швидкий аналіз – пошук лише в d_r або повний аналіз – пошук в повному контенті ІР. Перший варіант значно швидший, але другий дає значно більш релевантні результати.

В результаті цього аналізу для кожного r -го ІР формується вектор співставлення

$$a_r = \langle a_{r_1}, \dots, a_{r_n} \rangle$$

такий, що a_{r_i} , $i = \overline{1, n}$ – кількість співставлень у тексті r -го ІР з i -м терміном тезауруса задачі.

Співставлення виконується з використанням ЛБЗ: якщо l_{ij} , що співвідноситься з терміном тезауруса задачі $x_i \in X$, входить до складу r -го ІР, тоді a_{r_i} збільшується на 1 (у спрощеному варіанті порівняння елементів ЛДЗ виконується з контентом короткого опису відповідного ІР).

Значущість кожного ІР для задачі оцінюється як функція від цього вектора та вектора ваги кожного терміну.

Для більшості задач може використовуватися наступна формула:

$$f_r = \sum_{i=1}^n a_{r_i} * w_i. \quad (1)$$

Якщо потрібно порівнювати оцінки ІР для різних тезаурусів задачі (приміром, для оцінки досліджування ПрО та як її відображення у різних наборах ІР – електронних бібліотеках, сайтах тощо), то виникає необхідність у використанні замість (1) нормованої оцінки. Але таку нормовану оцінку можна використовувати тільки в тому випадку, якщо не застосовуються негативні ваги термінів тезауруса задачі.

$$fn_r = \frac{\sum_{i=1}^n a_{r_i} * w_i}{\sum_{i=1}^n a_{r_i} * \sum_{i=1}^n w_{r_i}}. \quad (2)$$

Значення оцінки (2) завжди буде знаходитися в діапазоні між 0 та 1, в той час як значення оцінки (1) може приймати довільне значення, як позитивне, так і негативне.

Отримані оцінки використовуються для перевпорядкування знайдених ІР: користувач отримує у першу чергу ІР з більш високими коефіцієнтами відповідності поточній задачі користувача.

Програмна реалізація

Вищезапропонований метод побудови тезауруса задачі застосовується в системі семантичного пошуку “МАПС” [29].

Ця система має забезпечити виконання складних багаторазових запитів у спеціалізованих ПрО, пов’язаних з професійними або науковими інтересами користувачів. Запити таких користувачів можуть повторюватися від сеансу до сеансу або змінюватися, але залишатися у рамках певної ПрО пошуку, в якій користувачі є експертами. Система надає користувачу ті результати пошуку, що відносяться до предметних областей, які його цікавлять і відповідають його інформаційним потребам.

“МАПС” дозволяє зберігати і повторно виконувати запити, зберігати формальний опис області інтересів користувача у вигляді тезаурусів задачі та онтологій ПрО. Отримані результати перевпорядковуються з урахуванням цих знань, а також персоніфікованого індексу легкості читання природномовних ІР [30, 31].

“МАПС” не замінює собою ІПС. Вона є посередником між користувачем та існуючими засобами пошуку. Її призначення – зробити звертання користувача до ІР більш ефективним, зручним та швидким.

Крім того, у “МАПС” при профілюванні користувачів використовується специфічний для природномовних ІР критерій оцінювання – складність тексту для розуміння. Особливістю системи є використання оригінального знання-орієнтованого алгоритму, що дозволяє 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).

Наукова новизна “МАПС” полягає в інтегрованому використанні онтологічного подання знань, агентної парадигми та технологій Semantic Web для пошуку інформації на семантичному рівні.

Основні технології та методи, інтегровані в “МАПС”:

- застосування технологій Semantic Web [32]: використання OWL [33] для інтероперабельного представлення онтологій та тезаурусів, що описують ПрО;
- реалізація теоретико-множинних операцій над тезаурусам;
- методи генерації тезаурусів за природномовними текстами;
- використання технологій Web 2.0 [34] (хмар тегів – для візуалізації пошукових тезаурусів; соціальних сервісів – для взаємодії між користувачами);
- оригінальні алгоритми впорядкування інформаційних ресурсів, знайдених системою, з урахуванням ваги онтологічних термінів;
- використання критеріїв оцінки читабельності тексту для пошуку інформації, що відповідає персональним потребам користувача;
- використання методів індуктивного виведення для узагальнення досвіду роботи “МАПС”;
- застосування мультиагентного підходу до створення моделі інтелектуальної інформаційно-пошукової системи та представлення компонентів системи як інтелектуальних BDI-агентів для формалізації поведінки системи в цілому [35];
- використання парадигми інтелектуальних Web-сервісів для опису функцій агентів системи, що дозволяє їх інтероперабельне багаторазове використання [36].

Основою “МАПС” є технології Semantic Web, зокрема, мова представлення онтологій OWL і засоби його обробки. Для представлення знань щодо того, що цікавить користувача, використовуються

онтології ПрО та базовані на них тезауруси задач. При цьому тезаурус будується користувачем за відповідною онтологією самостійно, а онтологія обирається з набору онтологій, запропонованих на сайті розробниками системи.

В процесі розвитку “МАПС” виникла потреба в підключенні репозиторію онтологій, щоб користувачі могли повторно використовувати знання ПрО, доступні в Web. При цьому пошук може здійснюватися не тільки за ключовими словами, а і за іншими важливими властивостями онтологій – обсяг, розробники, кількість та типи відношень, базові DL, діалекти мов подання тощо. Тому надалі представляється доцільним реалізувати в “МАПС” засоби взаємодії з репозиторіями онтологій, що підтримують пошук потрібної користувачу онтології, виявлення схожих на обрану користувачем онтологій, а також зіставлення побудованого користувачем тезауруса з іншими онтологіями і тезаурусами.

“МАПС” базується на онтологічній моделі, що описує семантику взаємодії користувачів і ресурсів “МАПС” в інформаційному просторі Web. Ця модель також може застосовуватися у побудові тезауруса задачі, якщо пошукова ситуація є персональною (тобто стосується саме даного користувача) [37].

Отримавши у відповідь від зовнішньої ІПС набір інформаційних ресурсів, “МАПС” намагається здобути з них потрібні користувачеві відомості. У найпростішому випадку, якщо потрібний користувачеві ІО є документом (можливо, певного типу), система перевпорядковує отримані посилання на ІР з урахуванням персональних особливостей користувача та збережених у БД системи відомостей про ці ІР. У більш складних випадках з ІР здобуваються відомості про властивості атрибутів шуканого ІО. Приміром, якщо користувачеві був потрібен Web-сервіс з певними властивостями, то він отримує опис вхідних і вихідних даних наявних Web-сервісів, що відповідають його умовам та опис їх роботи.

“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.

При редагуванні вже створеного тезауруса можна вводити вагу різних термінів, що позначають їх важливість для пошуку (як позитивну, так і негативну), цілі числа від –9 до +9. Ця інформація дозволяє відображати тезаурус у вигляді хмари тегів (червоним кольором відмічені терміни з негативною вагою, синім – з позитивною, розмір шрифта відображає числові значення ваги) (рис. 1).



Рис. 1. Створення тезауруса задачі в “МАНПС”

Результати пошуку, отримані від зовнішньої ІПС (наприклад, від Google) перевпорядковуються за допомогою критерію (1). Крім цього, “МАНПС” пропонує користувачеві адресу та анотацію кожного текстового ІР та оцінки легкості його читання (рис. 2), обчислені з урахуванням термінів тезауруса задачі – слова та словосполучення, пов’язані з термінами цього тезауруса, не вважаються складними для користувача.

Слід зазначити, що для представлення знань у “МАНПС” використовуються онтології, які можна поділити на дві окремі групи – внутрішні і зовнішні. Внутрішні онтології створюються безпосередньо розроблявачами “МАНПС” і можуть поповнюватися в процесі взаємодії

“МАНПС” із користувачами. Основною особливістю такі онтології є те, що розроблявачам цілком відома їхня структура і зміст, тому можна прогнозувати кінцівку обчислень.



Рис. 2. Впорядкування результатів пошуку в системі “МАНПС”

Зовнішні онтології дозволяють інтегрувати в “МАНПС” динамічні і розподілені знання, доступ до яких забезпечує Web. Пошук таких онтологій може здійснюватися в різноманітних репозиторіях, чи ж вони можуть бути сформовані в процесі роботи користувача з іншими інтелектуальними додатками. У загальному випадку інформація про складність структури таких онтологій, про те, на яких дескриптивних логіках вони базуються, наскільки повними є такі знання, відсутня, і це не дозволяє прогнозувати час роботи алгоритмів для глибокого аналізу й обробки таких онтологій. Тому в “МАНПС” для обробки зовнішніх онтологій застосовуються спрощені алгоритми, що використовують тільки найбільш прості властивості онтологій (наприклад, обробляються тільки відношення “клас-підклас”). Для більшості задач інформаційного пошуку цього досить, але для пошуку сукупностей складних ІО – наприклад, для дослідження і композиції семантичних Web-сервісів чи формування мультиагентних систем – потрібно використовувати більш складні структурні зв’язки. Саме для таких ситуацій до складу “МАНПС” входять онтології складних ІО.

Висновки

Використання онтологічного аналізу для інтелектуалізації пошукових процедур забезпечує створення семантичної надбудови над традиційними інформаційно-пошуковими системами та дозволяє використовувати знання щодо індивідуальних інформаційних потреб користувачів.

Запропонований у роботі підхід орієнтований на користувачів з постійними та складними інформаційними інтересами, такими як науково-дослідницька діяльність у певній сфері. Це припускає здатність таких користувачів до аналізу відповідної предметної області, обізнаності в її основних поняттях та зв'язках між ними та їх потребу в створенні складних багаторазових запитів.

Для персоніфікації пошуку інформаційних об'єктів та інформаційних ресурсів, які містять відомості, потрібні користувачам для розв'язання поточних задач, запропоновано будувати та застосовувати тезауруси таких задач. Це дозволяє використовувати знання щодо предметної області пошуку та структури інформаційних об'єктів, які мають самі користувачі або які отримуються з відповідних онтологій. Визначення термінів, пов'язаних з семантичним пошуком, дозволяє більш чітко формулювати проблеми, пов'язані з пошуком інформації у відкритому середовищі Web.

Програмна реалізація запропонованого підходу підтверджує ефективність його практичного використання.

1. *Amerland D.* Google Semantic Search: Search Engine Optimization (SEO) Techniques That Gets Your Company More Traffic, Increases Brand Impact and Amplifies Your Online Presence. – Que Publishing. – 2013. – 230 p.
2. *Lawrence S.* Context in the Web Search. – <http://citeser.nj.nec.com/lawrence00context.html>.
3. *Berry M.W.* Survey of text mining // *Computing Reviews* 45.9, 2004. – 244 p.
4. *Andon P., Deretsky V.* Approach to Automatic Creation of Ontology from Documents for Improving Existent Information Retrieval // *Proc. of 2-nd Balkan Conference in Informatics (BCI'2005)* November 17–19, 2005. – P. 236–241.
5. *Cimiano P.* *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications.* – Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2006. – 347 p.
6. *Fensel D., Harmelen F., Horrocks I., McGuinness D., Patel-Schneider P.* OIL: An Ontology Infrastructure for the Semantic Web. – <http://www.cs.man.ac.uk/~7Ehorrocks/Publications/download/2001/IEEE-IS01.pdf>.
7. *Gruber T.R.* What is an Ontology? – <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>.
8. *Uschold M., Grüninger M.* Ontologies: Principles, Methods and Applications // *Knowledge Engineering Review* 11(2), 1996. – P. 93–155.
9. *Guarino N.* Formal Ontology in Information Systems // *Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 1998.* – P. 3–15.
10. *Боргест Н.М., Коровин М.Д.* Онтологии: современное состояние, стандарты, средства поддержки. Уч. пособие. СРАУ. – Самара, 2013. – 84 с.
11. *Клецев А.С., Артемьева И.Л.* Математические модели онтологий предметных областей. Часть 1. Существующие подходы к определению понятия «онтология» // *Научно-техническая информация, серия 2.* – 2001. – С. 20–27.
12. *ISO 25964-1:2011, Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval.* – Geneva: International Organization for Standards, 2011.
13. *Нариньяни А.С.* Кентавр по имени ТЕОН: Тезаурус + Онтология. – <http://www.artint.ru/articles/narin/teon.htm>.
14. *Браславский П.И., Гольдштейн С.Л., Ткаченко Т.Я.* Тезаурус как средство описания систем знаний // *Информационные процессы та системы.* – 1997. – № 11, Серия 2. – С. 16–22.
15. *Величко В., Волошин П., Свитла С.* Автоматизированное создание тезауруса терминов предметной области для локальных поисковых систем. – www.foibg.com/ibs_isc/ibs-15/ibs-15.pdf.
16. *Добров Б.В., Иванов В.В., Лукашевич Н.В., Соловьев В.Д.* Онтологии и тезаурусы: модели, инструменты, приложения. – Элект-

- ронная книга, 2006. – 220 с. – http://catscpp.googlecode.com/svn-history/r146/trunk/diploma/materials/ontologies_tesauruses.pdf.
17. Gladun A., Rogushina J. Use of Semantic Web Technologies and Multilingual Thesauri for Knowledge-Based Access to Biomedical Resources // International Journal of Intelligent Systems and Applications. – 2012, N 1. – P. 11–20. – <http://www.mecspress.org/ijisa/ijisa-v4-n1/IJISA-V4-N1-2.pdf>
 18. Gladun A., Rogushina J. Use of Semantic Web technologies in design of informational retrieval systems // in Book “Building and Environment”, Nova Scientific Publishing, New-York, USA. – 2009. – P.89–103.
 19. Hartmann J., Palma R., Gomez-Perez A. Ontology Repositories // in Book “Handbook on Ontologies”, Edt. by S. Staab, R. Studer, Springer, 2009. – P. 551–572.
 20. Baclawski K., Schneider T. The Open Ontology Repository Initiative: Requirements and Research Challenges // Proceedings of the 8th International Semantic Web Conference ISWC-2009, October 25, 2009, USA.
 21. Лесько О.Н., Рогушина Ю.В. Использование онтологий для анализа семантики естественно-языковых текстов // Проблемы програмування. – 2009. – № 3. – С. 59–65.
 22. Leuf B., Cunningham W. The Wiki way: collaboration and sharing on the Internet. – 2001. – <http://www.citeulike.org/group/13847/article/7659081>.
 23. Wagner C. Wiki: A technology for conversational knowledge management and group collaboration // The Communications of the Association for Information Systems. – 2004. – V. 13(1). – P. 264–289.
 24. Рогушина Ю.В., Гладун А.Я. Семантическая Википедия как источник онтологий для интеллектуальных поисковых систем // В кн.: Advanced Research in Artificial Intelligence. International Book Series "Information Science and Computing". ИТНЕА, Sofia, 2008. – P. 172–178.
 25. Гладун А.Я., Рогушина Ю.В. Основы методологии формування тезаурусів з використанням онтологічного та мереологічного аналізу // Искусственный интеллект. – 2008. – № 5. – С.112–124.
 26. Jansen B.J., Spink A., Saracevic T. Real life, real users, and real needs: a study and analysis of user queries on the Web. – <http://citeseer.nj.nec.com/jansen00real.html>.
 27. Kobsa A. User modeling: recent work, prospects and hazards. – <http://zeus.gmd.de/~kobsa/papers/1993-aui-kobsa.pdf>.
 28. Рогушина Ю.В. Разработка средств персонализации интеллектуальных Web-приложений // Материалы V Международной научно-технической конференции «Открытые семантические технологии проектирования интеллектуальных систем» OSTIS-2015. – Минск: БГУИР, 2015. – С. 265–270. – <http://www.conf.ostis.net/images/8/8b/OSTIS-2015.compressed.pdf>.
 29. Rogushina J., Gladun A. Ontology-based competency analyses in new research domains // Journal of Computing and Information Technology. – 2012. – V. 20, N. 4. – P. 277–293.
 30. Рогушина Ю.В., Гришанова І.Ю. Літературний твір наукового характеру "Модель мультиагентної інформаційно-пошукової системи "МАПС"("Модель МАПС"). – Свідоцтво про реєстрацію авторського права на твір № 32068.
 31. Flesch Reading Ease Readability Formula. – <http://oleandersolutions.com/fleschreadingease.html>.
 32. McLaughlin H. SMOG grading a new readability formula // Journal of Reading. – 1969. – N 22. – P. 639–646.
 33. Davies J., Fensel D., van Harmelen F. Towards the Semantic Web: Ontology-driven knowledge management // John Wiley & Sons Ltd., England. – 2002. – 288 p.
 34. OWL 2 Web Ontology Language Document Overview. W3C. 2009. – <http://www.w3.org/TR/owl2-overview/>.
 35. McCann R., Shen W., Doan A. Matching schemas in online communities: A web 2.0 approach // Proc. of ICDE, 2008.
 36. Rao A.S., Georgeff M.P. Modeling rational agents within a BDI-architecture // In R. Pikes and E. Sandewall, eds.. Proc. of Knowledge Representation and Reasoning (KR&R-91), Morgan Kaufmann Publishers: San Mateo, CA, April 1991. – P. 473–484.
 37. Cowles P. Web Services and the Semantic Web. – http://ezolin.pisem.net/logic/ws_and_sw_rus.html.

References

1. Amerland D. Google Semantic Search: Search Engine Optimization (SEO) Techniques That Gets Your Company More Traffic, Increases Brand Impact and

- Amplifies Your Online Presence. – Que Publishing, 2013. – 230 p.
2. *Lawrence S.* Context in the Web Search. – <http://citeser.nj.nec.com/lawrence00context.html>.
 3. *Berry M.W.* Survey of text mining // *Computing Reviews* 45.9, 2004. – 244 p.
 4. *Andon P., Deretsky V.* Approach to Automatic Creation of Ontology from Documents for Improving Existent Information Retrieval // *Proc.of 2-nd Balkan Conference in Informatics (BCI'2005)* November 17–19, 2005. – P. 236–241.
 5. *Cimiano P.* Ontology Learning and Population from Text: Algorithms, Evaluation and Applications. *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications.* – Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2006. – 347 p.
 6. *Fensel D., Harmelen F., Horrocks I., McGuinness D., Patel-Schneider P.* OIL: An Ontology Infrastructure for the Semantic Web. – <http://www.cs.man.ac.uk/%7Ehorrocks/Publications/download/2001/IE-EE-IS01.pdf>.
 7. *Gruber T.R.* What is an Ontology? – <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>.
 8. *Uschold M., Grüninger M.* Ontologies: Principles, Methods and Applications // *Knowledge Engineering Review.* – 11(2). – 1996. – P. 93–155.
 9. *Guarino N.* Formal Ontology in Information Systems // *Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 1998.* – P. 3–15.
 10. *Borgest N.M., Korovin M.D.* Ontologies: modern state, standards and support means, SRAU. – Samara, 2013. – 84 p. (in Russian)
 11. *Keshev A.S., Artemieva I.L.* Mathematical models of domain ontologies. Part 1. Existing approaches for definition of «ontology» concept Science-technical information, series 2, 2001. – P. 20–27. (in Russian)
 12. *ISO 25964-1:2011, Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval* – Geneva: International Organization for Standards, 2011.
 13. *Nariniani A.C.* Centaurus named TEON: thesaurus + ontology. – <http://www.artint.ru/articles/narin/teon.htm>. (in Russian)
 14. *Braslavsky P.I., Goldshtein S.L., Tkachenko T.Ya.* Thesaurus as a mean of knowledge system deskribing // *Information processes and systems, 1997.* – N 11, series 2. – P. 16–22. (in Russian)
 15. *Velichko V., Voloshin P., Svitla S.* Automated creation of thesaurus of domain terms for local retrieval systems. – www.foibg.com/ibs_isc/ibs-15/ibs-15.pdf. (in Russian)
 16. *Dobrov B.V., Ivanov V.V., Lukashevich N.V., Soloviev V.D.* Ontologies and thesauri: models, instruments, applications. 2006. – 220 p. – http://window.edu.ru/resource/583/64583/files/Dobrov_978-5-9963-0007-5%2F1-2-3_cC0007-5.pdf. (in Russian)
 17. *Gladun A., Rogushina J.* Use of Semantic Web Technologies and Multilinguistic Thesauri for Knowledge-Based Access to Biomedical Resources // *International Journal of Intelligent Systems and Applications.* – 2012. – N 1. – P. 11–20. – <http://www.mecspress.org/ijisa/ijisa-v4-n1/IJISA-V4-N1-2.pdf>
 18. *Gladun A., Rogushina J.* Use of Semantic Web technologies in design of informational retrieval systems // in Book “Building and Environment”, 2009 Nova Scientific Publishing, New-York, USA. – P. 89–103.
 19. *Hartmann J., Palma R., Gomez-Perez A.* Ontology Repositories // in Book “Handbook on Ontologies”, Edt. by S.Staab, R.Studer, Springer, 2009. – P. 551–572.
 20. *Baclawski K., Schneider T.* The Open Ontology Repository Initiative: Requirements and Research Challenges//*Proceedings of the 8th International Semantic Web Conference ISWC-2009, October 25, 2009, USA.*
 21. *Lesko O.N., Rogushina Y.V.* Use of ontologies for analysis of natural language texts semantics // *Problems in programming.* – 2009. N 3. – P. 59–65. (in Russian)
 22. *Leuf B., Cunningham W.* The Wiki way: collaboration and sharing on the Internet. – 2001. – <http://www.citeulike.org/group/13847/article/7659081>.
 23. *Wagner C.* Wiki: A technology for conversational knowledge management and group collaboration // *The Communications of the Association for Information Systems.* – 2004. – V. 13(1). – P. 264–289.
 24. *Rogushina Y.V., Gladun A.Ya.* Semantic Wikipedia as a source of ontologies for intelligent retrieval systems // *Advanced Research in Artificial Intelligence. International Book Series "Information Science and Computing".* ITHEA, Sofia, 2008. – P. 172–178. (in Russian)
 25. *Gladun A.Ya., Rogushina Y.V.* Methodology bases of thesauri creation with use of ontological and mereological analysis // *Artificial intelligence.* – 2008. – N 5. – P. 112–124. (in Ukrainian)

26. *Jansen B.J., Spink A., Saracevic T.* Real life, real users, and real needs: a study and analysis of user queries on the Web. – <http://citeseer.nj.nec.com/jansen00real.html>.
27. *Kobsa A.* User modeling: recent work, prospects and hazards. – <http://zeus.gmd.de/~kobsa/papers/1993-aui-kobsa.pdf>.
28. *Rogushina Y.V.* Design of personification means of intelligent Web applications // Proc. of V scientific and technical conf. OSTIS-2015. – Minsk, 2015. – P. 265–270. (in Russian) – <http://www.conf.ostis.net/images/8/8b/OSTIS-2015.compressed.pdf>.
29. *Rogushina J., Gladun A.* Ontology-based competency analyses in new research domains // Journal of Computing and Information Technology. V.20, N. 4, 2012. – P. 277–293.
30. *Rogushina Y.V., Grishanova I.Y.* Literary work “Model of multiagent information retrieval system MAIPS (“MAIPS model”). – Copyright certificate of product registration N 32068. (in Ukrainian).
31. *Flesch* Reading Ease Readability Formula. – <http://oleandersolutions.com/fleschreadingease.html>.
32. *McLaughlin H.* SMOG grading a new readability formula // Journal of Reading. – 1969. – N 22. – P.639–646.
33. *Davies J., Fensel D., van Harmelen F.* Towards the Semantic Web: Ontology-driven knowledge management. – John Wiley & Sons Ltd., England. – 2002. – 288 p.
34. *OWL 2* Web Ontology Language Document Overview. W3C. 2009. – <http://www.w3.org/TR/owl2-overview/>.
35. *McCann R., Shen W., Doan A.* Matching schemas in online communities: A web 2.0 approach // Proc. of ICDE, 2008.
36. *Rao A.S., Georgeff M.P.* Modeling rational agents within a BDI-architecture // In R. Pikes and E. Sandewall, eds.. Proc. of Knowledge Representation and Reasoning (KR&R-91), Morgan Kaufmann Publishers: San Mateo, CA, April 1991. – P. 473–484.
37. *Cowles P.* Web Services and the Semantic Web. – http://ezolin.pisem.net/logic/ws_and_sw_rus.html.

Одержано 07.12.2015

Про автора:

Рогущина Юлія Віталіївна,
кандидат фізико-математичних наук,
старший науковий співробітник.
Кількість наукових публікацій в
українських виданнях – 100.
Кількість наукових публікацій в
іноземних виданнях – 25.
Індекс Гірша – 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