

О. І. Кушнірецька<sup>1</sup>, І. І. Кушнірецька<sup>1</sup>, А. Ю. Берко<sup>2</sup>  
Національний університет “Львівська політехніка”,  
<sup>1</sup>кафедра інформаційних систем та мереж,  
<sup>2</sup>кафедра загальної екології та екоінформаційних систем

## СЕМАНТИЧНИЙ ПОШУК І ЗБЕРІГАННЯ ДАНИХ НАУКОВО-ТЕХНІЧНОЇ ІНФОРМАЦІЙНОЇ СИСТЕМИ

© Кушнірецька О. І., Кушнірецька І. І., Берко А. Ю., 2015

Описано семантичний пошук і зберігання даних науково-технічної інформаційної системи. Наведено пропозиції щодо семантичного структурування контенту науково-технічної інформаційної системи з явним структурованим представленням семантичних зв'язків між інформаційними об'єктами, що містяться в системі. Визначено основні складові математичної моделі онтології науково-технічної інформаційної системи для семантичного пошуку і зберігання науково-технічного інформаційного ресурсу.

**Ключові слова:** науково-технічна інформаційна система, семантичний пошук, науково-технічний інформаційний ресурс, зберігання науково-технічних інформаційних ресурсів, онтології.

This paper describes the semantic search and storage of data of scientific and technical information system. The proposals of semantic structuring of the content of scientific and technical information system with explicitly structured representation of semantic relations between information objects contained in the system have been presented. The main components of the mathematical model of ontology of scientific and technical information system for semantic search and storage of scientific and technical information resource have been determined.

**Key words:** scientific and technical information system, semantic search, scientific and technical information resources, storage of scientific and technical information resources, ontology.

### Вступ. Загальна постановка проблеми

Сьогодні накопичено величезний обсяг знань і інформаційних ресурсів, що стосуються науково-технічної тематики. Однак доступ до цих ресурсів дуже ускладнений через те, що сучасні інформаційні системи використовують доволі обмежений набір методів пошуку, подання, зберігання, інтерпретації та обробки інформації. Такі системи переважно представляють і видають користувачеві знання і дані у вигляді окремих текстових документів, тоді як для людини найбільш природною формою подачі інформації є подання її у вигляді мережі взаємопов'язаних фактів. При цьому велика частина інформації, представленої в Інтернеті і локальних сховищах даних (електронних бібліотеках, архівах і т.п.), стає практично недоступною через неефективну роботу пошукових машин, які переважно застосовують примітивні механізми пошуку за ключовими словами, що не враховують ані семантики слів, що входять до запиту, ані його контексту. Невирішеною залишається проблема зручного подання і зберігання інформації. Навіть вже представлені методи пошуку, обробки, подання і зберігання інформації залишаються недоступними широкому колу користувачів через відсутність змістовної інформації про них.

### Зв'язок висвітленої проблеми із важливими науковими та практичними завданнями

Завдання пошуку і зберігання інформації через недостатню формалізованість предметної області, несистематизованість самих ресурсів і їх слабку структурованість у різних інтернет-сайтах, електронних бібліотеках та архівах стало фокусом широкої теоретичної роботи, а численні

проблеми залишаються невирішеними. Особливо актуальним напрямком є дослідження можливостей формування в науково-технічній інформаційній системі явно представлених структурованим чином багатоаспектних семантичних зв'язків між науковими документами, що в них містяться, в поєднанні з методами моніторингу змін структури цих зв'язків і основаними на такій структурі новими функціональними можливостями.

### **Аналіз останніх досліджень та публікацій**

Сьогодні накопичено величезну кількість інформації з різних галузей знань, а із бурхливим розвитком інформаційних технологій значна її частина представлена безпосередньо в мережі Інтернет різноманітними інформаційними ресурсами. Це бази даних, веб-сайти (або окремі розділи, якщо ці розділи є сукупністю документів), колекції електронних бібліотек (включаючи електронні музеї, галереї та ін.), веб-форуми, соціальні мережі, електронні видання, спеціальні масиви наукових даних (електронні моделі об'єктів, масиви експериментальних даних, бази знань, діагностичні, експертні системи тощо), масиви конструкторських, технологічних та інших виробничих документів (створювані в межах CALS-технологій), ГІС-системи та електронні карти (масиви геоданих) [1]. Сьогодні вже розроблено кілька методів для вирішення проблеми зберігання інформаційних ресурсів та автоматичного вилучення структурованої інформації з веб-сайтів і додатків, які через специфіку характеру HTML мови, як правило, надають неструктуровану інформацію. Ці методи переважно розроблені на основі певних правил видобування даних, які можуть бути менш або більш витонченими і визначаються або в напівавтоматичному режимі за допомогою демонстрації, або автоматично за допомогою машино-навчальних алгоритмів або інших методів штучного інтелекту. Огляд декількох засобів видобування веб-даних наведено в роботі [5].

Завдяки опису семантики змісту інформаційного ресурсу пропонується підвищувати якість послуг сервісів систем, використовуючи семантичні мережі, фрейми або онтології. Сьогодні більш досконалими і виразними вважають онтологічні моделі [3, 6]. Це пояснюється тим, що вони ґрунтуються на формальних (дескриптивних) логіках, і для роботи з ними в рамках концепції Semantic Web розроблено набір стандартних технологій, що дають змогу створювати, описувати і використовувати онтологічні моделі.

До технологій Semantic Web належать такі мови, як RDF (Resource Description Framework), RDFS (RDF Schema), OWL (Ontology Web Language) і SPARQL (Simple Protocol And RDF Query Language). Крім цього, розроблено багато програмних систем з роботи з онтологічними моделями (редактори, системи логічного висновку, RDF-сховища і т.п.).

У світі вже активно ведуться дослідження з використання технологій Semantic Web у таких галузях, як електронні бібліотеки, інтеграція та пошук інформації в мережі Інтернет та системи управління знаннями [2, 4, 6]. Проте досі все ж немає узгодженого і обґрунтованого набору моделей і методів для створення і підтримання роботи з явним описом семантики науково-технічного інформаційного ресурсу за допомогою технологій Semantic Web. Тому проблема семантичного пошуку і зберігання даних науково-технічного інформаційного ресурсу є надзвичайно актуальною. Інтернет багато в чому змінив умови використання систем текстового пошуку і висунув до них нові вимоги. У стислому вигляді основні з цих вимог можна сформулювати так [7]:

- ефективне опрацювання дуже великих колекцій документів;
- поліпшене відображення смислового змісту документів і користувацьких пошукових запитів;
- реалізація мультимедійної обробки, тобто спільної обробки документів різних форматів і уявлень – текстових документів, зображень, аудіо, відео та ін.;
- реалізація ефективних методів пошуку в потоках документів (задачі фільтрування).

Окремо варто зазначити підвищення вимог до пошукових систем відносно так званого людського фактора, що визначає ефективність взаємодії людини і пошукової системи. Це стосується насамперед проектування користувацьких інтерфейсів і організації роботи людини з пошуковою системою. Останнім часом з'явилася тенденція до інтелектуалізації інформаційних

систем, орієнтація систем на людину, що, безумовно, зачіпає і системи інформаційного пошуку. Це розвиток таких актуальних напрямків досліджень, як:

- багатомовний і крос-мовний пошук;
- фактографічний пошук (відповіді пошукової системи на задані користувачем питання);
- пошук відеоданих (за змістом відеоданих, пошук відомих об'єктів, задачі розпізнавання і т.д.);
- інтерактивний пошук (реалізація діалогу з користувачем під час пошуку, уточнення запитів і т.д.);
- пошук за документом-зразком тощо.

У роботі [8] описано основні вимоги до подання і зберігання науково-технічних інформаційних ресурсів:

- 1) збирання і оновлення інформації – режим збирання та актуалізації повинен реалізувати технологію поточного збирання у режимі перевірки;
- 2) введення інформації – передбачає здійснення операцій контролю введення, зокрема за словниками і класифікаторами, а також відповідно до формату і структури;
- 3) зберігання інформації. Зберігатися основні масиви даних мають у вигляді, доступному для клієнта з використанням стандартного браузера;
- 4) пошук. У єдиному інтерфейсі під стандартним браузером повинно бути передбачено реалізацію декількох режимів пошуку:
  - атрибутивний;
  - тематичний;
  - лексичний.
- 5) подавати дані необхідно, використовуючи якісні сучасні технології.
- 6) інформаційно-лінгвістичне забезпечення. Система повинна містити дві складові: базу даних організацій (відомості про наукові організації) і пов'язану з нею базу даних ресурсів, що містить метайнформацію про науково-технічні ресурси;
- 7) набір елементів даних, що описують ресурси, повинен бути необхідним і достатнім для виконання операцій над ресурсами.

Виникає багато протиріч і проблем внаслідок технічної, методологічної та організаційної складності розглянутих завдань.

Водночас роботи в галузі інформаційного пошуку успішно розвиваються, накопичується необхідна теоретична і практична база. Сьогодні існують рішення, що добре зарекомендували себе в області текстового пошуку. Розглянемо деякі з них в ракурсі тематичного аналізу й ідентифікації документів.

Важливим поняттям, що характеризує вибір того чи іншого методу аналізу текстової інформації, а також реалізацію конкретного варіанта пошуку, є модель пошуку [7]. Модель пошуку являє собою поєднання таких складових:

- спосіб подання документів;
- спосіб представлення пошукових запитів;
- вид критерію релевантності документів.

Варіації цих складових визначають велику кількість всіляких реалізацій систем текстового пошуку. Найпопулярнішими сьогодні є:

1. Найпростіші моделі пошуку. Це моделі, в яких документ наведено у вигляді набору асоційованих з ним зовнішніх атрибутів. До найпростіших моделей пошуку належать:

1) модель дескриптивного пошуку. У цих системах представлення документа описується сукупністю слів чи словосполучень лексики предметної області, що характеризують зміст документа. Такі слова/словосполучення називаються дескрипторами. Індексують документ в системах дескриптивного пошуку за допомогою призначення для нього сукупності дескрипторів. Дескриптори можуть приписуватися документу або на основі його змісту, або на основі його назви. Ці два процеси називають відповідно індексуванням за змістом і індексуванням за заголовками документів. У деяких дескрипторних системах індексують документи вручну експерти в

предметній області системи, в інших це виконується автоматично [7]. Дескрипторні системи можна зарахувати до класу систем, орієнтованих на бібліографічний пошук або пошук “за каталогом”.

2) модель, основана на дублінському ядрі. Дублінське ядро (Dublin Core) [8] – це набір елементів метаданих, зміст яких зафіксовано в специфікації, що визначає його стандарт. У термінах значень цих елементів можна описувати зміст різного роду текстових документів.

Початкову версію дублінського ядра було запропоновано в 1995 році на проведену в Дубліні (США) симпозиумі, організованому Online Computer Library Center (OCLC) і National Center for Supercomputing Applications (NCSA) для опису інформаційних ресурсів бібліотечних систем [8].

У моделі пошуку, основаній на дублінському ядрі, представленням  $k$ -го документа є множина пар:

$$D_k = \{(N_{ik}, V_{ik})\}, \quad (1)$$

де  $N_{ik}$  – ім'я  $i$ -го елемента метаданих дублінського ядра в описі змісту  $k$ -го документа;  $V_{ik}$  – значення цього елемента метаданих.

Зображенням запиту також є множина пар деяких елементів дублінського ядра та їхніх значень:

$$D_j = \{(N_{jk}, V_{jk})\}, \quad (2)$$

де  $N_{jk}$  – ім'я  $j$ -го елемента метаданих дублінського ядра в описі користувацького запиту;  $V_{jk}$  – значення цього елемента метаданих.

Критерій релевантності  $k$ -го документа виглядає так:

$$Q \subseteq D . \quad (3)$$

2. Моделі, основані на класифікаторах. Документ у цій моделі набуває вигляду сукупності асоційованих з ним атрибутів. Атрибутами є ідентифікатори класів, до яких належить цей документ. Класи формують ієрархічну структуру класифікатора. Запит можна подати двома способами:

– простий варіант – запитом є ідентифікатор якого-небудь класу із заданого класифікатора. Критерій релевантності документа запиту – клас документа збігається з класом в представленні запиту або є його підкласом;

– складний варіант – у запиті можна вказати кілька класів класифікатора. Критерій релевантності документа запиту – клас документа збігається з яким-небудь із зазначених у запиті класів або є його підкласом.

3. Булеві моделі. В булевих моделях пошуку користувач може формулювати запит у вигляді булевого виразу, використовуючи для цього оператори І, АБО, НІ. Терми запиту залежать від конкретного варіанта моделі пошуку. В булевій моделі, орієнтованій на пошук “за текстом”, термами будуть слова, відповідно, критерієм релевантності буде умова входження деякого слова або словосполучення до тексту документа. В булевій моделі, орієнтованій на пошук за класифікаторами, термами виразу будуть ідентифікатори класів класифікатора. В булевій моделі пошуку з використанням дублінського ядра термом буде значення елементів метаданих. Документ, що має значення елементів метаданих, збіжні зі значеннями, заданими в запиті, вважається релевантним.

У загальному випадку критерієм релевантності документа запиту в булевих моделях пошуку є істинність булевого виразу, заданого в запиті.

Однією з безперечних переваг булевої моделі пошуку є простота її реалізації. Головними недоліками вважають:

– відсутність можливості ранжирування знайдених документів за ступенем релевантності, оскільки відсутні критерії її оцінювання;

– складність використання – далеко не кожен користувач може вільно оперувати булевськими операторами, формулюючи свої запити.

Варто зазначити, що робилися спроби ускладнення булевої моделі пошуку для забезпечення можливості ранжирування безлічі виданих користувачеві документів. А саме, запропоновано декілька варіантів так званих розширених булевих моделей [9]. У цих моделях вводяться спеціальні

узагальнення булевих операторів, що дають змогу додати підвищеної ваги документам, що задовольняють булевий запит, і пониженої ваги – всім іншим документам [7].

4. Векторні моделі. Сьогодні векторні моделі є найпоширенішими і вживаними на практиці моделями пошуку. Векторні моделі, на відміну від булевих, дають змогу без великих зусиль ранжувати результуючу множину документів запиту. Суть таких моделей зводиться до подання документів і запитів у вигляді векторів.

5. Імовірнісні моделі. Вперше ідеї таких моделей було запропоновано в 1960 році [10]. В їх основу покладено принцип імовірнісного ранжування (Probabilistic Ranking Principle, PRP). Цей принцип полягає в наступному: найвищої загальної ефективності пошуку досягають в разі, коли результуючі документи ранжуються за зменшенням імовірності їх релевантності запиту. Спочатку для кожного документа оцінюється ймовірність того, що він релевантний запиту, а потім за цими оцінками ранжують документи.

Існують різні способи отримання цих оцінок, а також додаткові припущення і гіпотези на основі апріорних відомостей щодо документів колекції, які й визначають конкретну реалізацію ймовірнісної моделі пошуку.

Наприклад, цю оцінку можна обчислити, відповідно до теореми Байеса, за деякою функцією ймовірностей входження термів цього документа в релевантні і нерелевантні документи. За допомогою запиту визначають ймовірність входження заданого терма до релевантних документів, а за повної колекції документів визначається ймовірність входження цього терма до нерелевантних документів [7].

6. Мережі виведення. Так само, як і імовірнісні моделі, мережі виведення основані на принципі імовірнісного ранжування результуючих документів пошуку [11]. Головна їхня відмінність від імовірнісних моделей полягає в тому, що використовується оцінка неймовірності релевантності документа запиту, а ймовірності того, що він задовольняє інформаційні потреби користувача. За цією моделлю процес пошуку документів описується як процес міркувань в умовах невизначеності. В процесі такого міркування оцінюється ймовірність того, що інформаційні потреби користувача, виражені за допомогою одного або кількох запитів, задоволено. Мережа виведення основана на байєсівській мережі, яка містить вузли чотирьох видів. Вузлами першого виду є документи колекції, вивчені користувачем у процесі пошуку. Вузлами другого виду є терми, якими описується зміст документів. Вузлами третього виду є запити, що складаються з термів, якими описується зміст документів. Вузол четвертого типу в мережі тільки один, і він відповідає інформаційним потребам користувача, які не відомі пошуковій системі. Всі вузли першого і другого видів формуються заздалегідь для заданої колекції. Вузли третього виду і їх зв'язку з вузлами термів, що описують документи, і вузлом інформаційних потреб формуються для кожного конкретного запиту.

Після того, як мережу побудовано, оцінюють документи колекції, поширюючи мережею оцінки ймовірності вузла конкретного документа. Результатом поширення є обчислення ймовірності вузла інформаційних потреб. При цьому оцінка для кожного документа будується незалежно від оцінок інших документів, з урахуванням матриць описують зв'язки між вузлами документів і вузлами термів, вузлами термів і вузлами запитів.

Оцінювання повторюють для кожного документа, потім вони ранжуються на основі обчислених оцінок ймовірності вузла інформаційних потреб [7].

Для подолання проблем природних мов необхідно описувати семантику термінів і враховувати її під час оцінювання близькості між запитом і документом для підвищення релевантності одержуваних результатів. Для цього використовують різні методи, серед яких можна виділити методи на основі зустрічальності термінів і методи на основі словників предметних областей (тезауруси, онтології).

### **Виділення проблем**

Розглядаючи пошук і зберігання науково-технічної інформації в класичних інформаційних ресурсах, дані зазвичай описуються тільки за допомогою різних метаданих їх контексту, а їх зміст (контент ресурсу) представляється тільки у вигляді набору термінів. Це призводить до таких проблем, як:

- відсутність у системи розуміння сенсу ресурсів через такі особливості природних мов, як синонімія, полісемія і омонімія;
- слабка інтеграція інформаційних ресурсів: система може містити велику кількість різномірних інформаційних ресурсів, описаних з використанням різних метаданих, з якими потрібно працювати як з єдиним цілим, що вимагає вирішення завдання їх інтеграції;
- недостатньо висока точність виконання основних функцій науково-технічних систем: для задоволення потреб користувачів системи потрібно забезпечити високу точність результатів основних функцій, таких як пошук, автоматична категоризація і формування рекомендацій;
- відсутність зручних засобів навігації між інформаційними ресурсами.

Вирішити перераховані проблеми і підвищити якість послуг сервісів науково-технічних систем можна завдяки опису семантики змісту інформаційного ресурсу і реалізації роботи з ними, використовуючи передусім ефективні моделі пошуку та подання інформаційного ресурсу.

### **Формулювання мети**

Мета роботи полягає у використанні існуючих підходів та методів для вирішення проблеми семантичного пошуку і зберігання даних науково-технічної інформаційної системи, виділенням семантики змісту інформаційного ресурсу та побудовою математичної моделі структурного подання тексту науково-технічної інформаційної системи.

### **Аналіз отриманих наукових результатів**

Науково-технічні інформаційні ресурси являють собою джерела знань для вирішення необхідних задач і проблем, відповіді на які зазвичай необхідно знайти серед цих ресурсів. У зв'язку з цим в науково-технічній інформаційній системі необхідно зробити доступними знання, що містяться в інформаційних ресурсах, для ефективної їх обробки та повторного використання. В результаті цього відповіді на питання можна отримати на основі описуваних знань науково-технічного інформаційного ресурсу.

Існуючі технології та перспективність нового напрямку дослідження мотивують дослідити та розробити такий новий вид науково-технічної інформаційної системи, як системи управління знаннями науково-технічного електронного ресурсу.

Загалом будь-яка науково-технічна інформаційна система повинна містити такі підсистеми:

- пошуку і представлення інформаційних ресурсів;
- адміністрування користувачів інформаційних ресурсів;
- адміністрування інформаційних ресурсів;
- забезпечення спільної роботи користувачів.

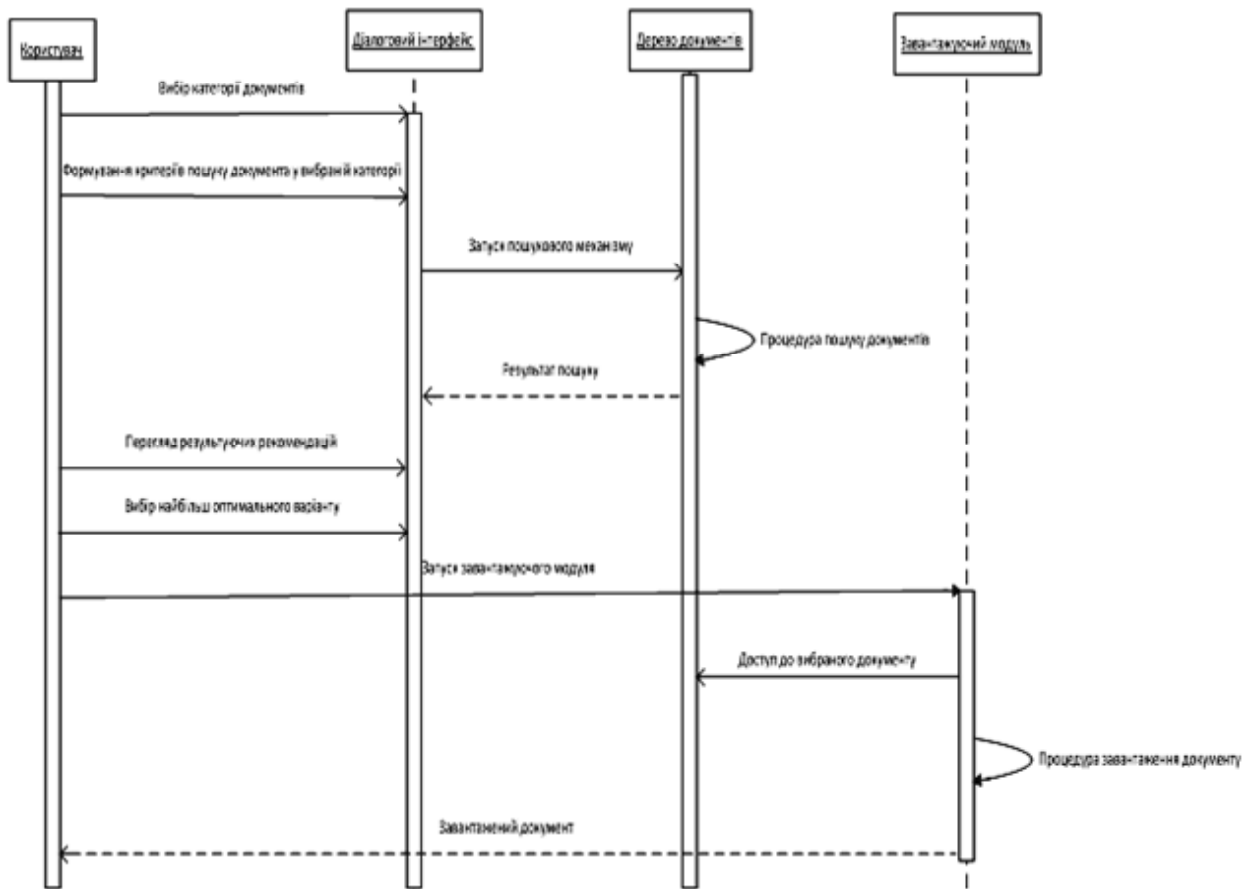
Основне завдання будь-якої науково-технічної системи – якісне надання потрібної інформації користувачу системи. Розглянемо процес пошуку та завантаження документа в науково-технічній інформаційній системі, зображений на рисунку у вигляді UML-діаграми послідовностей.

Основними об'єктами послідовності процесу пошуку та завантаження документа є: користувач, діалоговий інтерфейс, дерево документів та завантажувальний модуль.

Користувач виконує такі дії: вибір категорії документів, формування критеріїв пошуку документа у вибраній категорії, перегляд результуючих рекомендацій, вибір найбільш оптимального варіанта, запуск завантажувального модуля, після чого користувачу виводиться завантажений документ. Користувач взаємодіє з системою через діалоговий інтерфейс. Діалоговий інтерфейс, своєю чергою, взаємодіє із деревом документів, запускаючи механізми пошуку, після яких йому повертається результат пошуку. У дереві документів відбувається сама процедура

пошуку документів. Також із завантажувального модуля повинно прийти підтвердження наміру завантаження вибраного документа.

Завантажувальний модуль забезпечує власне процедуру завантаження документа. Для цього він звертається за доступом до певного документа у дереві документів і завантажує його.



Послідовність процесу пошуку та завантаження документа

Науково-технічні інформаційні ресурси можна поділити на суб'єкти знань (користувачі) і ресурси знань (електронні ресурси, каталоги, профілі користувачів, схеми класифікації знань предметних областей).

Всі ресурси мають контент — зміст, тобто знання, які в них утримуються, та контекст – набір відношень з іншими ресурсами науково-технічної системи.

Контекст – це зовнішні зв'язки розглянутого ресурсу з іншими ресурсами або даними. А контент – це смисловий зміст ресурсу або твердження про поняття і примірники понять моделі, які містяться в ресурсі. Наприклад, для деякого ресурсу (документа) його контекстом будуть стосунки з його авторами, ключовими словами, категорією, часом створення, правами доступу до нього тощо, а його контентом буде те, що в ньому описується.

Контекст і контент всіх науково-технічних інформаційних ресурсів повинні бути описані на основі деякої єдиної моделі знань і об'єднані в єдиний простір знань науково-технічної системи.

Використовуючи онтології, опишемо загальну модель науково-технічних інформаційних ресурсів:

$$O = \{O_b, O_d\}, \quad (4)$$

де  $O_b$  – базові онтології науково-технічної інформаційної системи;  $O_d$  – онтології основних предметних областей знань для опису змісту науково-технічних інформаційних ресурсів.

До базових онтологій належать онтології [4]:

- користувачів;

- ресурсів;
- системи.

Визначимо основні складові математичної моделі онтології науково-технічної інформаційної системи для семантичного пошуку і зберігання науково-технічного інформаційного ресурсу:

1) множина понять (класів)  $C$ , для кожного з яких визначають такі характеристики:

- URI – ідентифікатор поняття (URI – уніфікований ідентифікатор ресурсів), заданий за певними правилами;

- рядкові літеральні значення ( $L$ ) – тестові мітки для їх інтерпретації (терміни, коментарі та ін. певною природною мовою), наприклад: співробітник, документ тощо;

- початкову множину його примірників ( $I$ ).

2) множина відношень (предикатів)  $P$ , для кожного з яких задаються:

- URI – ідентифікатор предикату;

- рядкові літеральні значення ( $L$ );

- ваговий коефіцієнт  $p_w$ ;

- область дії (domain) і область значення (range), задані за допомогою функцій ( $F$ );

- можливі властивості: функціональності, обернено-функціональності, транзитивності, рефлексивності, симетричності, асиметричності;

3) множина часткових порядків  $T$ , яка називається відношенням успадкування  $C \times C$ ,  $P \times P$ ;

4) множина екземплярів  $I$ , для кожного з яких задаються:

- URI;

- рядкові літеральні значення ( $L$ ).

Розглядаємо онтологію науково-технічної інформаційної системи як знакову систему в такому вигляді:

$$O = \langle C, P, T, I, F, L, A \rangle, \quad (5)$$

де  $C$  – множина понять;  $P$  – множина властивостей (двомісні предикати);  $T$  – частковий порядок на множинах  $C$  і  $P$ , що задає відношення “підклас” і “суперклас”;  $I$  – множина екземплярів понять з  $C$ ;  $F$  – функція, яка призначає кожному елементу з  $P$  множину елементів з  $C$  (з урахуванням їх ієрархії в  $T$ ), до яких воно застосовується (область дії, domain), і множину елементів з  $C$  або літералів (примірників примітивних типів, таких як рядки і числа), які можуть бути їх значеннями (область можливих значень, range);  $L$  – множина текстових міток для понять, властивостей та примірників;  $A$  – набір аксіом онтології – тверджень про елементи предметної області, які вважаються правильними, виражених з використанням відповідної логічної мови.

Будь-якій властивості  $p \in P$  може бути задано значення вагового коефіцієнта  $p_w \in [0,1]$ , яке вказує на смислову близькість між суб'єктом і об'єктом твердження (триплетом), який використовує цю властивість. Отже, якщо  $p_w = 0$ , то суб'єкт і об'єкт абсолютно не пов'язані між собою за смисловою близькістю (семантикою), а якщо  $p_w = 1$ , то вони вважаються абсолютно аналогічними за семантикою. Предикати онтології, у яких  $0 \leq p_w \leq 1$ , називатимемо семантичними властивостями (відношеннями).

У єдиній онтологічній моделі науково-технічної інформаційної системи всі поняття множини  $C$  вважаються спадкоємцями універсального поняття Thing, яке розглядається як суперклас усіх понять онтології. Всі твердження (триплети) про зв'язок між поняттями і предикатами включаються до бази знань компонент. Крім цього, до бази знань також вводять такий компонент, як набір тверджень, що описують екземпляри понять.

Використовуючи онтологічну модель ресурсів науково-технічної інформаційної системи, можна описати всі ресурси з погляду їх контексту і контенту на основі визначених для них моделей.

Виділимо основні операції з даними науково-технічної інформаційної системи, використовуючи онтологію:

1) отримання даних:



- вхідні дані: будь-який ідентифікатор компоненти або шаблонний триплет із змінними суб'єкта; предиката; об'єкта; або їх комбінацій;
- вихідні дані: набір триплетів в різних форматах або окремі значення цих змінних;
- 2) видалення даних:
  - вхідні дані: набір триплетів задається на вході або отримується за допомогою функції “Отримання даних” на основі її вхідних даних;
  - вихідні дані: повідомлення про успішне виконання операції.;
- 3) додавання даних:
  - вхідні дані: кінцевий набір триплетів;
  - вихідні дані: повідомлення про результат виконання операції.
- 4) оновлення даних: комбінація функції “видалення даних” та функції “додавання даних”;
- 5) завантаження даних:
  - вхідні дані: документи в різних RDF-форматах;
  - вихідні дані: повідомлення про результат виконання операції.

### **Висновки та перспективи подальших наукових розвідок**

Описано семантичний пошук і зберігання даних науково-технічної інформаційної системи. Розглянуто найпопулярніші сьогодні моделі пошуку інформаційних ресурсів. Наведено основні складові науково-технічної інформаційної системи. Зображено послідовність процесу пошуку та завантаження документа науково-технічної інформаційної системи. Використовуючи онтології, описано загальну модель науково-технічних інформаційних ресурсів. Визначено основні складові математичної моделі онтології науково-технічної інформаційної системи для семантичного пошуку і зберігання науково-технічного інформаційного ресурсу. Виділено основні операції з даними науково-технічної інформаційної системи, використовуючи онтології. Результати досліджень, наведені в статті, відкривають нові можливості для наукометричного аналізу і досліджень історії розвитку наукового знання в конкретних напрямках науки.

1. Когаловский М. Р. Семантическое структурирование контента научных электронных библиотек на основе онтологий / М. Р. Когаловский, С. И. Паринов // В сб. “Современные технологии интеграции информационных ресурсов: сборник научных трудов”, 2011. – Вып. 2.
2. Когаловский М. Р. Перспективные технологии информационных систем / М. Р. Когаловский. – М.: ДМК Пресс; М.: Компания АйТи, 2003. – 288 с.
3. Кушнірецька О. І. Принципи семантичного пошуку науково-технічних інформаційних ресурсів / О. І. Кушнірецька, І. І. Кушнірецька, А. Ю. Берко // “Інформаційні управляючі системи та технології” (ІУСТ-Одеса-2014); Мат-ли III Міжн. наук.-практ. конф. 23–25 вересня 2014 р., Одеса / відп. ред. В. В. Вичужанін. – “ВМВ”, 2014. – 328 с.
4. Ле Хоай. Формирование рекомендаций в семантических электронных библиотеках / Хоай Ле, А. Ф. Тузовский // Проблемы информатики. – 2012. – № 3. – С.113–119.
5. Нгуен Б. Н. Метод семантического поиска в коллекции интеллектуальных документов / Б. Н. Нгуен, А. Ф. Тузовский // Научно-технический вестник Поволжья. – 2012. – № 2. – С. 246–254.
6. Agichtein E., Lawrence S., Gravano L. Learning search engine specific query transformations for question answering. In Proc. of the WWW10, pp. 169-178, 2001.
7. Bulskov H. Ontology-based Information Retrieval. Doctoral Dissertation / Henrik Bulskov Styltsvig. – Roskilde University, Denmark, 2006.
8. Laender A. A brief survey of web data extraction tools / A. Laender, B. Ribeiro-Neto, A. Silva // SIGMOD record, 31(2), 2002.
9. Maron M. E., Kuhns J. L. On relevance, probabilistic indexing and information retrieval. Journal of the ACM, No. 7, 1960, pp. 216–244.
10. Salton G., Fox E., and Wu H. Extended Boolean information retrieval. Communications of the ACM, Vol. 26, No. 4, December 2001, pp. 35-43.
11. Singhal A. Modern Information Retrieval: A Brief Overview. Data Engineering Bulletin, IEEE Computer Society, Vol. 24, No. 4, December 2001, pp. 35–43.