

КВАЗІСЕМАНТИЧНИЙ ПОШУК ТЕКСТОВИХ ДАНИХ В ЕЛЕКТРОННОМУ ІНФОРМАЦІЙНОМУ РЕСУРСІ

*МИХАЙЛЮК А.Ю., ПИЛИПЧУК О.В.,
СНІЖКО М.В., ТАРАСЕНКО В.П.*

Описується оригінальний, квазісемантичний, підхід до пошуку текстових даних за змістом. Орієнтація пошуку природномовних інформаційних об'єктів на зміст досягається шляхом застосування інтелектуального редактора пошукового запиту, який функціонує на базі онтології відповідної предметної галузі. Аналізується ефективність запропонованого підходу при реалізації комп'ютерних засобів інструментальної підтримки інформаційно-аналітичної діяльності, зокрема у науково-освітній галузі.

1. Вступ

Сучасний етап розвитку науково-освітньої галузі суспільства характеризується загальною інтенсифікацією процесів інтеграції знань у всесвітню інформаційну мережу. Завдяки глобальній мережі Internet з'являється доступ до колосальної кількості електронних

джерел інформації, серед яких бібліотеки наукових робіт, енциклопедично-довідкові ресурси, спеціалізовані ресурси з певних галузей науки чи культури. Вузким місцем в роботі з цими джерелами залишається складність пошуку конкретної інформації серед великого обсягу даних. Для вирішення цієї задачі зазвичай використовуються інформаційно-пошукові та інформаційно-моніторингові системи. Незважаючи на те, що цими системами охоплюється достатньо великий сегмент електронних природномовних інформаційних об'єктів, існує ряд недоліків у механізмах їх функціонування, які суттєво зменшують ефективність застосування в освітній та науково-дослідницькій діяльності.

По-перше, відчутним недоліком переважної більшості систем названого класу є низька впорядкованість пошукового відгуку. При відпрацюванні багатьох запитів пошукові системи надають користувачеві більше мільйона посилань, і якщо потрібний об'єкт буде не в першій сотні результуючого списку, то фактично його можна вважати недоступним користувачеві. Хоча пошукові системи виконують певне ранжування елементів пошукової вибірки, втім ранг часто відображає не стільки рівень релевантності документа, скільки популярність ресурсу, якому він належить.

По-друге, до пошукового відгуку в багатьох випадках потрапляють текстові інформаційні об'єкти, що фактично не відповідають запиту через наявні неоднозначності у можливостях щодо трактування окремих ключових слів. Пошукові системи часто неспроможні коректно розпізнати та розв'язати таку неоднозначність навіть опираючись на контекст запиту.

По-третє, знайдені інформаційні об'єкти можуть не відповідати пошуковому інтересу користувача. Іншими словами результуюча вибірка документів, навіть задовольняючи вимогам щодо релевантності, може характеризуватися низькою пертинентністю. Так, одні й ті самі поняття можуть набувати різного змістовного відтінку за різної прагматики. Наприклад, можна шукати інформацію про бази даних, як про продукт на ринку програмного забезпечення, а можна як про специфічну парадигму побудови інструментів накопичення даних. Пошуковій системі важко автоматично визначити необхідний напрямок, орієнтуючись лише на пошуковий запит.

Очевидно, що сучасним інформаційно-пошуковим системам (ІПС) не вистачає здатності до "осмисленого" пошуку, тобто пошуку за змістом. Широкодоступні семантичні ІПС могли б якісно розширити пакет «ступенів інформаційної свободи» користувача, надавши передусім можливість шукати текстові дані не по простому збігу слів, а по їх семантичній близькості до запиту. Семантичний аналіз документів дозволив би виділяти також головний зміст інформаційного об'єкта і використовувати його при ранжуванні пошукового відгуку. Також реалізувалася б можливість використання ефективного багатомовного пошуку, оскільки документи порівнювалися б із запитом на рівні понять, які є практично універсальними для всіх мов. Однак застосування алгоритмів семантичного аналізу в контексті пошукових задач стикається з рядом проблем, що ускладнюють, а інколи унеможливають реалізацію семантичного пошуку в його "класичному" варіанті [1,2]. Серед основних перешкод можна виділити зокрема такі:

- оскільки природна мова є надзвичайно складним об'єктом формалізації, більшість алгоритмів власне змістовного аналізу тексту опирається на спеціальні службові лінгвістичні ресурси, які попри колосальні обсяги і надзвичайну логічну складність все ж не можуть служити адекватною моделлю природної мови і тому на рівні реальних застосувань не забезпечують прийнятної ефективності;

- мала кількість слів у пересічному пошуковому запиті [3] без відповідного контексту фактично не дозволяє провести його якісний семантичний аналіз, наслідком чого є проблеми зі встановленням релевантності;

- процедури проведення семантичного аналізу та встановлення семантичної близькості документів і запиту з технічної точки зору базуються на використанні спеціальних структур даних (для зберігання та обробки семантичних образів), на реалізації специфічних

методів доступу до них, а також на виконанні обчислювально складних алгоритмів обробки і тому потребують тривалого залучення масштабних системних ресурсів, що входить у протиріччя з необхідністю вести семантичний пошук у реальному часі;

- «класичний» семантичний пошук передбачає систематичне проведення спеціальної семантичної переіндексації всього текстового інформаційного ресурсу, причому з огляду на необхідність попереднього максимального врахування множини прагматик наступного пошуку масштаби цієї задачі досягають рівня, практична реалізація котрого є проблематичною навіть на фоні широких можливостей сучасних інформаційних технологій;

Таким чином, актуальною сьогодні є задача розробки нових методів пошуку текстових інформаційних об'єктів за змістом, а також створення відповідних моделей ІПС, які б дозволили конструктивно вирішити наведені вище проблеми чи звести їх вплив до мінімуму і досягти оптимального балансу між точністю пошуку та швидкістю і складністю процедур його реалізації.

2. Концепція квазісемантичного пошуку

Складність реалізації «класичної» концепції семантичного пошуку в реальних ІПС спонукає до реалізації якісно відмінного підходу, який би, по-перше, дозволяв використати основні переваги пошуку за змістом і разом з тим забезпечити алгоритмічну простоту і помірну ресурсоемність та, по-друге, не вимагав би суттєвої модифікації вже наявної на сьогодні інфраструктури пошукових індексів. Таким чином, сама пошукова база та основні механізми пошуку зможуть залишатись незмінними, а врахування власне семантики буде виконуватись або на етапі взаємодії з пошуковими індексами (наприклад, так звана концептуальна індексація на основі тезаурусу [4]), або на етапі формування запиту.

Запит — це основне джерело інформації для ІПС щодо пошукових інтересів користувача. Інформаційна потреба є певним психологічним станом людини [5]. Оскільки в даний час не існує спеціальних носіїв для подібної інформації, то найбільш природним способом вираження інформаційної потреби є її вкладання у формі, доступній для обробки машиною, наприклад, у формі запиту, сформульованого природною мовою. Такий спосіб означення мети пошуку, безумовно, зручний для людини, але створює певні проблеми для ефективної роботи пошукової машини. Одна з найсуттєвіших проблем полягає в тому, що запит може мати декілька різних тлумачень (трактувань), він може бути неповним, або надмірним, може містити багатозначні слова, сильно залежати від контексту, погано відображати інформаційну потребу і т.п. Хоча запит рідко точно відповідає власне інформаційній потребі, проте сьогодні це єдиний спосіб зв'язку між користувачем і пошуковою машиною. Отже, оптимізація пошукового запиту з метою забезпечення якомога більш точного опису інформаційного образу об'єкта пошуку та з

врахуванням можливостей наявного пошукового індексу є на сьогодні єдиним реально доступним шляхом підвищення ефективності процесу пошуку стандартними ІПС. В більшості випадків для цього використовують спеціальні словники, тезауруси, онтологічні бази знань, класифікатори тощо [3]. Серед наявних сьогодні методів оптимізації запиту на пошук з метою його «семантизації» доцільно звернути увагу на такі.

1. Розширення пошукового запиту синонімами, рідше антонімами (з використанням словників відповідно синонімів та антонімів) [5]. Це перший і найпростіший крок, що був зроблений у напрямку покращення результатів звичного повнотекстового пошуку шляхом залучення семантики. Власне модифікація пошукового запиту синонімами та антонімами стала можливою після появи спеціальних електронних словників — так званих електронних тезаурусів, які були першими носіями семантики у комп'ютерній лінгвістиці. Втім даний метод має суттєвий недолік: у випадку, якщо термін у складі запиту має багато значень, розширення запиту синонімами може лише «зашумити» пошуковий відгук, збільшивши у ньому кількість об'єктів, що не відповідають пошуковим інтересам користувача. Введення антонімів до складу пошукового запиту у випадку, коли трактування початкового терміну було неправильним, у свою чергу пов'язано з ризиком вилучення з результуючої вибірки потенційно релевантних текстів. Таким чином модифікація пошукового запиту синонімами та антонімами корисна лише у випадку наявності можливості однозначного трактування всіх термінів вихідного запиту.

2. Обмеження поля пошуку шляхом уточнення тематичної спрямованості бажаного результату. Даний метод потенційно спроможний суттєво підвищити пертинентність документів результуючої вибірки, проте слід пам'ятати, що його реалізація вимагає залучення інструментарію автоматичної тематичної класифікації текстів, який за складністю хоч і значно поступається засобам семантичного аналізу, та все ж на сьогодні ще не забезпечує необхідної достовірності результату [6,7,8].

3. Уточнення пошукового запиту шляхом долучення саме тих атрибутів об'єкта пошуку, які є вагомими з огляду на пошуковий інтерес. Процедура виконується із залученням об'єктних онтологій [5]. Наприклад, якщо необхідно провести пошук за терміном «комп'ютер», то можна скористатися тим, що в структурі сучасних онтологій знань цей концепт буде мати ряд атрибутів, як-от: «тип», «виробник», технічні параметри тощо. Таким чином, для обмеження інформаційного поля пошуку необхідно лише запропонувати користувачеві заповнити деякі атрибути об'єкта необхідними значеннями. Результати пошуку за таким модифікованим запитом будуть значно краще задовольняти інформаційні потреби користувача. Однак розробка відповідних онтологій становить

окрему досить складну задачу, яка на даний час знаходиться у стадії розв'язання.

4. Модифікація пошукового запиту за допомогою так званих стратегій пошуку та відповідних спеціальних тезаурусів [9]. Стратегії являють собою шаблони для поєднання понять запиту із взаємопов'язаними поняттями тезаурусу. Кожна стратегія спрямована на модифікацію пошукового запиту за певним тематичним напрямком. Такий метод використовується для пошуку в межах вузьких предметних галузей і базується на використанні специфічних тезаурусів з сильно диференційованими зв'язками.

Підсумовуючи, можна сказати, що наявні реалізації згаданих методів оптимізації пошукового запиту лише частково використовують семантичні можливості онтологічних баз знань і не дають бажаного ефекту тому, що здебільшого принципово орієнтовані на повну автоматичність. Проблема полягає в тому, що процедура оптимізації пошукового запиту, яка по суті являє собою семантично-прагматичний синтез тексту, навіть у спрощеному варіанті (з огляду на незначні обсяги пересічного запиту: 2-3 слова [10]) за своєю складністю мало поступається семантично-прагматичному аналізу. Таким чином, отримання у близькому майбутньому «робочих» інформаційно-пошукових систем, придатних для роботи з семантикою може бути досягнуто в першу чергу за рахунок технічної реалізації парадигм, які ґрунтуються, зокрема, на залученні людини-користувача до виконання найбільш інтелектуальних і найменш формалізованих на сьогоднішній день технологічних етапів семантичного пошуку. Будемо називати квазісемантичною таку концепцію пошуку за змістом, згідно з якою:

а) власне пошук ведеться згідно з алгоритмами стандартного повнотекстового пошуку, за допомогою відповідного інструментарію та інформаційної інфраструктури;

б) орієнтація пошуку на семантику реалізується шляхом формування пошукового запиту, відповідним чином оптимізованого за допомогою онтологічної бази знань;

в) процедура формування пошукового запиту виконується напівавтоматично за безпосередньої участі користувача як єдиного достовірного джерела пошукового інтересу.

3. Редактор запиту на квазісемантичний пошук

Формування запиту на квазісемантичний пошук навіть за наявності досить повної комп'ютерної онтологічної бази знань являє собою досить складну процедуру. Проблема може додатково ускладнюватись неповною визначеністю пошукового інтересу, недостатнім рівнем володіння користувачем відповідною предметною галуззю, його необізнаністю щодо характеру інформаційного ресурсу, в якому ведеться пошук, тощо [11].

Тому цілком закономірною є необхідність озброєння користувача відповідним інструментарієм – редактором пошукового запиту. Оскільки формування запиту є, по суті, процесом прийняття рішення, редактор запиту має бути інтелектуальною діалоговою системою на кшталт експертної системи (ЕС), база знань котрої має серед іншого містити онтологію предметних галузей, в межах котрих має вестись квазісемантичний пошук. На рис. 1 подано узагальнену структурну організацію інтелектуального редактора пошукового запиту.

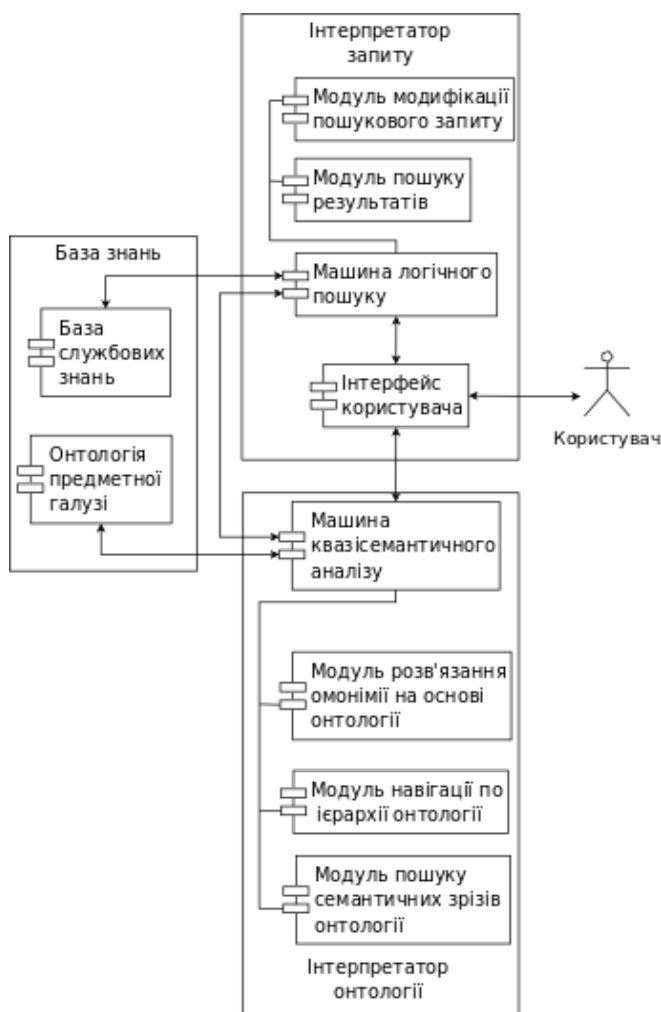


Рис. 1. Узагальнена структура редактора запиту на квазісемантичний пошук

База знань ЕС складається з двох сегментів — бази службових знань та онтології предметної галузі. База службових знань містить основні правила, щодо організації процесу формування та модифікації пошукового запиту. Онтологія вміщує формалізовану систему понять відповідної предметної галузі [12].

З точки зору організації механізмів виведення, особливістю інтелектуального редактора запиту є наявність у ньому двох інтерпретаторів - інтерпретатора запиту та інтерпретатора онтології. Перший, шляхом інтерпретації вмісту бази службових знань, реалізує загальною стандартну для ЕС логіку вибору правил по організації процедури модифікації пошукового запиту. Інтерпретатор онтології здійснює підбір категорій

предметної галузі, які після відповідного підтвердження користувачем включаються до запиту на квазісемантичний пошук.

Алгоритми функціонування інтерпретатора онтології в цілому та його відповідних модулів визначаються особливостями структурно-логічної організації онтологічної бази знань предметної галузі. З точки зору організації прототипом онтологічної бази знань редактора запиту може слугувати найбільш популярна сьогодні універсальна тезаурусно-онтологічна база WordNet та її локалізовані похідні (EuroWordNet, BalkaNet, RusNet тощо), в основі яких лежать принципи ієрархічності та асоціативної зв'язаності [5]. На рис. 2 запропоновано фрагмент побудованої згідно згаданих принципів експериментальної предметної онтології для підгалузі «Основи баз даних».

Даний фрагмент дозволить проілюструвати логіку функціонування інтерпретатора для WordNet-подібної онтології (див. діаграму на рис. 3).

Основу онтології складають поняття, або синсети. Синсет можна трактувати як загальне поняття, що відповідає дуже близьким синонімічним словам, скороченням тощо. Необхідно наголосити на тій особливості, що синсет може виражати не лише одиничне слово або термін, але й стійкі словосполучення. У поданому на рис. 2 фрагменті онтології прикладом такого синсету може бути, наприклад, “база даних”. Очевидно, це поняття об'єднує словосполучення “база даних” та скорочення “БД”. Синсет можна трактувати як основний елемент фіксації змісту термінів за рахунок надання йому певного тлумачення по аналогії з тезаурусом.

Неважко дійти до висновку, що одне й те саме слово може бути віднесеним до кількох понять (синсетів), що мають різне тлумачення. На цій особливості організації синсетів ґрунтується алгоритм функціонування модуля розв'язання омонімії. На діаграмі функціонування редактора запиту (рис. 3) видно, що відразу після розбору запиту інтерпретатор ініціює процес зв'язування термінів запиту з онтологією. На виході процедури зв'язування замість переліку слів отримується набір синсетів, з якими вони асоційовані. Це, по-перше, одразу розширює запит близькими синонімами із синонімічних груп кожного синсету, по-друге, у разі виявлення багатозначності терміна (якщо він задовольняє правилам приналежності до двох і більше синсетів одночасно) модуль розв'язання омонімії пропонує користувачеві варіанти тлумачення. Прикладом може служити слово “рядок”, яке в контексті наведеної онтології може бути як типом даних, так і фрагментом реляційної таблиці (див. рис. 2). Розв'язання омонімії розглядається як закріплення рішенням користувача одного із значень слова запиту за певним синсетом (на основі його опису і аналізу близьких синонімів). Перелік відповідних синсетів пропонується експертною системою. Таким чином, першим напрямком застосування редактора запиту є автоматичне виявлення неоднозначності понять запиту і розробка можливих варіантів ліквідації цієї

неоднозначності шляхом вибору користувачем конкретного значення відповідного слова.

Наступною характерною особливістю онтології, яка використовується редактором запиту, є її зв'язність. Всі синсети пов'язані між собою семантичними зв'язками різноманітних типів. Основним типом зв'язку є ієрархія понять, котру можна подати у вигляді дерева множини концептів мови, причому в корені цього дерева міститься найзагальніше поняття. Нарис.2 ці зв'язки показані суцільними лініями. Щонижче спускатись ієрархією даного дерева, то більш конкретними будуть відповідні поняття. Таке дерево можна вважати основою будь-якої онтології. Структурованість онтології, а саме впорядкованість понять, які становлять її основу, у певну ієрархію дозволяє зміщувати акценти пошуку, відповідно «зато-чуючи» його. Так, долучення до запиту понять, які в ієрархії онтології знаходяться вище від поняття, наявного у запиті на поточний момент, очевидно може суттєво збільшити обсяг вибірки текстових об'єктів. Обґрунтованість такого припущення є очевидною, оскільки більш загальні поняття явно і неявно фігурують у більшій кількості документів. З іншого боку є можливість «звужити» поле пошуку, конкретизуючи запит поняттями, що стоять у дереві онтології нижче по ієрархії. Редактор запиту в даному випадку виконує роль своєрідного навігатора ієрархією онтології. Модифікацію запиту користувачем при допомозі такого навігатора можна вважати другим напрямком застосування інтерпретатора запиту (рис.3).

Окрім вертикальних зв'язків, в онтології широко використовуються так звані горизонтальні семантичні зв'язки, які поєднують синсети і, таким чином, вказують на наявність певного виду семантичних відношень між відповідними поняттями (на рис.2 вони вказані пунктирними лініями). До горизонтальних зв'язків можна віднести, наприклад, відношення «частина-ціле», відношення властивості, відношення дії, спосо-

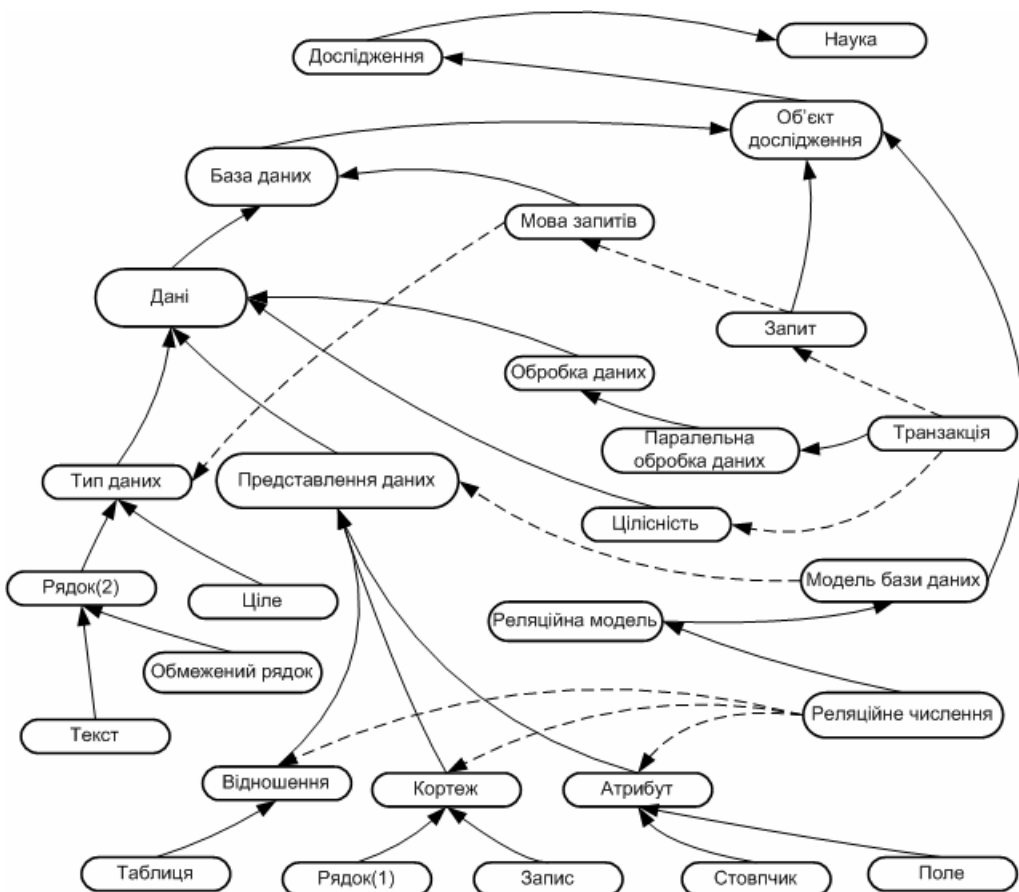


Рис.2. Фрагмент експериментальної онтології для предметної галузі «Основи баз даних»

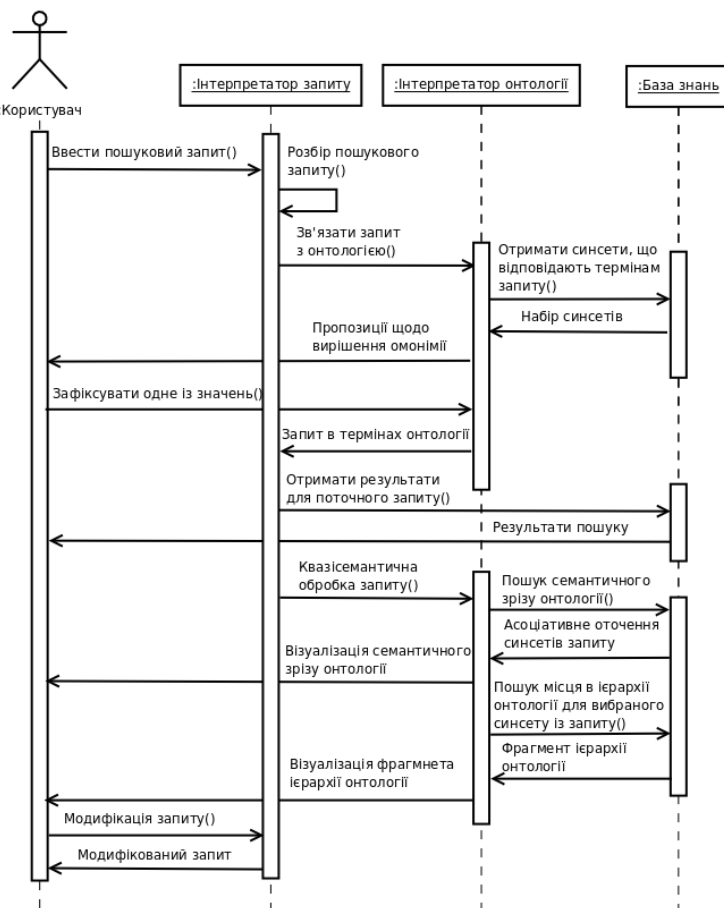


Рис. 3. Діаграма функціонування редактора запиту

бу дії та інші правила взаємодії об'єктів навколишнього світу [5]. Ці відношення можуть бути як загальними, тобто такими, що присутні в більшості універсальних онтологій (наприклад, відношення в онтологіях типу WordNet), так і специфічними, характерними передусім для певної предметної галузі [9].

Останні, здебільшого, присутні у спеціальних предметно-орієнтованих онтологіях з сильно диференційованими зв'язками. Більшість зв'язків онтології в загальному випадку можна розглядати як відношення асоціації. Тому саму онтологію можна трактувати як семантичну мережу (див. рис. 2). Редактор запиту використовує таку структуру онтології під час аналізу пошукового запиту, щоб знайти асоціативно близькі поняття. Для цього в онтології предметної галузі виділяються підмережі певного радіусу, в центрі кожної з котрих знаходиться відповідне поняття із запиту. Перетин цих підмереж формує зріз онтології по поняттях запиту і може вказати на області онтології, що семантично найбільше тяжіють до пошукового інтересу. Механізм автоматичного пошуку та візуалізації зрізів онтології предметної галузі, що найбільше тяжіють до групи понять із запиту, дає можливість нелінійно (тобто за прихованими асоціативними зв'язками) модифікувати запит для уточнення вектора пошуку. Цей, третій, напрямок застосування інтерпретатора запиту особливо корисний у контексті навчальної інформаційно-пошукової роботи, оскільки крім більш повного задоволення пошукового інтересу дозволяє досягти ще й певного дидактичного ефекту, оскільки наочно у всіх вимірах демонструє користувачеві систему понять відповідної предметної галузі.

Наведені способи взаємодії з онтологією в процесі функціонування редактора запиту (зв'язування запиту з онтологією та розв'язання неоднозначностей, навігація по ієрархії онтології та пошук семантичних зрізів з метою модифікації пошукового запиту) складають основу парадигми квазісемантичної концепції пошуку. На основі пропозицій, сформованих відповідними модулями ЕС, користувач може прийняти рішення щодо подальшої модифікації пошукового запиту відповідно до своїх пошукових інтересів. Слід також зазначити, що процедура модифікації запиту в більшості випадків може мати ітераційний характер. Кожна ітерація включає в себе як етап модифікації запиту, так і формування пошукового відгуку згідно з вмістом індексних масивів. Залежно від змісту та обсягу пошукового відгуку, що видає ППС за поточним запитом, на кожній з ітерацій може бути задіяно як один, так і декілька з наведених механізмів модифікації пошукового запиту, виходячи зі співвідношення поточного результату пошуку з пошуковим інтересом користувача.

4. Висновки

Таким чином, квазісемантичний пошук являє собою дієвий засіб підвищення ефективності знаходження текстових даних в електронному інформаційному ресурсі за рахунок привнесення до пошукового запиту

семантики, закладеної у онтологію предметної галузі. Інтелектуалізований програмний інструментарій редактора запиту дозволяє користувачеві в автоматизованому режимі виконувати модифікації запиту для підвищення його змістовності, що дає можливість значно підвищити пертинентність результатів пошуку.

Як вже зазначалось, реалізація концепції квазісемантичного пошуку в мінімальному варіанті, тобто в такому, що передбачає привнесення семантики лише на етапі формування запиту, не вимагає жодних змін у організації та вмісті індексних масивів. Тому редактор запиту, побудований згідно з викладеними у роботі засадами, може ефективно використовуватись як надбудова над існуючими стандартними інформаційно-пошуковими системами. Втім, у випадку розробки також спеціального квазісемантичного пошукового індексу (пошуковий індекс, побудований не на словесних формах, а на відповідних синсетах) або індексу концептів [3] ефект від використання редактора запиту значно зросте, оскільки пошук в індексі буде вестись по чітко зафіксованих поняттях, а не на основі збігу слів. Ще одним реальним кроком для підвищення ефективності використання квазісемантичного пошуку може служити так звана вибіркова модифікація, коли у ході вибору понять для розширення запиту (у тому числі з тезаурусу чи онтології) виконується перевірка їх наявності у пошуковому індексі.

Особливого значення концепція квазісемантичного пошуку та відповідний програмний інструментарій набувають у контексті їх застосування для інформаційного забезпечення навчальної та навчально-наукової роботи. Оскільки креативна діяльність в освіті тісно пов'язана з постійним пошуком нової інформації з малознайомих для користувача галузей знань, навігація по семантичній мережі онтології дає можливість «у фоновому режимі» адаптуватись до термінології та внутрішніх змістовних зв'язків відповідного предмета. Це не лише дозволяє тим, хто навчається, самостійно, без залучення фахівців, отримувати безпосередньо з глобального електронного інформаційного простору інформацію, необхідну для розв'язання поточних учбових задач, а й сприяє чіткій систематизації предметних знань, інтегруванню міжпредметних зв'язків, а відтак - формуванню наукового світогляду в цілому.

Література: 1. Сокирко А.В. Семантические словари в автоматической обработке текста (по материалам системы ДИАЛИНГ) // Дисс. канд. физ.-мат наук: 05.13.11. М., 2001. 120с. 2. Лезин Г.В., Тузов В.А. Семантический анализ текста на русском языке: семантико-синтаксическая модель предложения // Экономико-математические исследования: математические модели и информационные технологии. СПб.: Наука. 2003. Вып. 3. С.282–303. 3. Добрынин В.Ю., Некрестьянов И.С. Расширение запросов с помощью вероятностного латентного семантического индексирования // Труды третьей всероссийской научной конференции “Электронные библиотеки: перспективные методы и технологии, электронные коллекции”. Петрозаводск, 2001. С.151-155. 4. Добров Б.В., Лукашевич Н.В. Тезаурус и автоматическое концептуальное индек-

сирование в университетской информационной системе РОССИЯ // Третья Всероссийская конференция по Электронным Библиотекам "Электронные библиотеки: перспективные методы и технологии, электронные коллекции". Петрозаводск, 2001 С.78-82. **5. Соловьев В.Д., Добров Б.В., Иванов В.В., Лукашевич Н.В.** Онтологии и тезаурусы // Учебное пособие. Казань, Москва, 2006. 157с. **6. Додонов А. Г., Ландэ Д. В.** Выявление понятий и их взаимосвязей в рамках технологии контент-мониторинга // Реестрация, зберігання і обробка даних. 2006. Т. 8, №4. С.45-52. **7. Ландэ Д.В.** Основы интеграции информационных потоков: Монография. К.:Инжиниринг, 2006. 240 с. **8. Yang Y.** An Evaluation of Statistical Approaches to Text Categorization//Journal of Information Retrieval. 1997. №1. P. 67-88. **9. Браславский П.И.** Методы повышения эффективности поиска научной информации (на материале Internet) // Дисс. канд. техн. наук: 05.13.16. Екатеринбург, 2000. 154с. **10. Ландэ Д.В., Фурашев В.Н., Брайчевский С.М., Григорьев А.Н.** Основы моделирования и оценки электронных информационных потоков: Монография. К.: Инжиниринг, 2006. 176с. **11. Альшанский Г.В., Браславский П.И., Титов П.В.** Формирование информационных запросов к машинам поиска интернета на основе тезауруса // Доклад на VIII Международной конференции по электронным публикациям "EL-Pub2003". Новосибирск, 2003. Режим доступа: <http://www.ict.nsc.ru/ws/elpub2003/5964/>. **12. Литвин В.В., Пасічник В.В., Яцишин Ю.В.** Інтелектуальні системи: Підручник. Львів: "Новий світ-2000", 2009. 406с.

Надійшла до редколегії 11.09.2009

Рецензент: д-р техн. наук, проф. Хаханов В.І.

Михайлюк Антон Юрійович, кандидат технічних наук, старший науковий співробітник, завідувач науково-дослідної лабораторії прикладної інформатики Київського університету імені Бориса Грінченка. Адреса: Україна, 04053, Київ, вул. Воровського, 18/2, тел. служб. 461-02-38. E-mail: may-62@ukr.net

Пилипчук Олексій Васильович, аспірант кафедри спеціалізованих комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут». Адреса: Україна, 03056, Київ, просп. Перемоги, 37, тел. служб. 406-84-76. E-mail: ilexcorp@ukr.net

Сніжко Микола Васильович, молодший науковий співробітник науково-дослідної лабораторії прикладної інформатики Київського університету імені Бориса Грінченка. Адреса: Україна, 04053, Київ, вул. Воровського, 18/2, тел. служб. 461-02-38. E-mail: snizko-n@rambler.ru

Тарасенко Володимир Петрович, д-р техн. наук, професор, завідувач кафедри спеціалізованих комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут». Адреса: Україна, 03056, Київ, просп. Перемоги, 37, тел. служб. 236-32-02. E-mail: vtarasen@scs.ntu-kpi.kiev.ua