УДК 004.9

А. Ю. Михайлюк, О. С. Михайлюк, О. В. Пилипчук, В. П. Тарасенко

КОНТЕКСТНООРІЄНТОВАНИЙ ПІДХІД ДО РОЗВ'ЯЗАННЯ ПРОБЛЕМИ НЕОДНОЗНАЧНОСТІ ТЕРМІНІВ В РАМКАХ КВАЗІСЕМАНТИЧНИХ ЗАСТОСУВАНЬ

Київський Університет ім. Б. Грінченка, Київ

НТУУ "КПІ імені Ігоря Сікорського", Київ

Анотація: Розглядається проблема розв'язання неоднозначності термінів, що виникає фактично при будь-якій формі роботи з текстовими масивами даних: в інформаційному пошуку, під час аналітичної обробки текстових корпусів, в автоматичному перекладі, в інженерії онтологій та словників тощо. Пропонується контекстноорієнтований підхід до автоматизованого розв'язання задачі неоднозначності термінів, із застосуванням лінгвістичної онтології, розглядаються переваги та недоліки такого підходу та основні сфери і способи застосування.

Ключові слова: лінгвістична онтологія, текстові дані, квазісемантичний підхід.

Аннотация: Рассматривается проблема решения неоднозначности терминов, возникает фактически при любой форме работы с текстовыми массивами данных: в информационном поиске, во время аналитической обработки текстовых корпусов, в автоматическом переводе, в инженерии онтологий и словарей и тому подобное. Предлагается контекстно-ориентированный подход к автоматизированному решению задачи неоднозначности терминов, с применением лингвистической онтологии, рассматриваются преимущества и недостатки такого подхода и основные сферы и способы применения.

Ключевые слова: лингвистическая онтология, текстовые данные, квазисемантический подход.

Abstract: A problem of solving the ambiguity of terms which appears in fact at any form of working with text data sets: in information search, during analytical processing of text corps, automatic translation, in engineering ontologies and vocabularies, etc. is considered. A context-oriented approach to automated solving of terms ambiguity problem using a linguistic ontology is proposed. Advantages and disadvantages of the approach and main areas and ways of using are considered.

Key words: linguistic ontology, text data sets, quasi-semantic approach.

Вступ

Проблема розв'язання неоднозначності термінів виникає фактично при будь-якій формі роботи з текстовими масивами даних: в інформаційному пошуку, під час аналітичної обробки текстових корпусів [1], в автоматичному перекладі [2], в інженерії онтологій та словників [3] тощо.

Можна виділити два способи організації фіксації значення багатозначного терміна. «Ручний» спосіб передбачає фіксацію значення людиною (користувачем інформаційно-пошукової машини, редактором онтології, перекладачем тощо). Такий спосіб характеризується в багатьох випадках найвищою точністю фіксації значення, проте вимагає залучення зовнішнього суб'єкта, що не завжди є можливим, і, крім того, вимагає переривання автоматизованого процесу обробки даних. Інший спосіб — це автоматичне розв'язання неоднозначності терміна, який може характеризуватися різним ступенем точності фіксації значення, втім дозволяє зробити процес обробки інформаційних даних безперервним. Найбільшу цінність становлять, очевидно, методи автоматичного розв'язання проблеми неоднозначності термінів, оскільки вони дають можливість максимально автоматизувати весь процес обробки текстової інформації.

Стаття присвячена розширенню теоретичного апарату побудови сучасних комп'ютерних інструментів для усунення неоднозначності термінів у природномовному тексті в ході автоматизованого перекладу, в процесі навчання або при здійсненні аналітичної діяльності.

Актуальність

Квазісемантичний підхід до пошуку [4], як один із видів інформаційно-пошукової діяльності, що активно взаємодіє з текстовими даними на поняттєвому рівні, особливо чутливий щодо точної фіксації значення термінів, оскільки квазісемантичні процедури залежать від правильності трансляції текстових даних у форму понять, що їм відповідають. Квазісемантика активно використовує в якості джерела семантичних даних допоміжний ресурс у вигляді лінгвістичної онтології, процес розв'язання омонімії термінів в квазісемантичних процедурах відбувається із застосуванням цього виду баз знань. Тому створення прикладного інструментарію компенсації неоднозначності термінів у ході використання лінгвістичної онтології є надзвичайно актуальною задачею сьогоднішнього дня.

Мета

В рамках теорії квазісемантичного аналізу природномовних текстових даних створити дієвий апарат розв'язання проблеми омонімії.

Задачі

1. Розробити контекстноорієнтований підхід до автоматизованого розв'язання задачі неоднозначності термінів, із застосуванням лінгвістичної онтології.
2. Проаналізувати переваги та недоліки запропонованого підходу та основні сфери і способи застосування.

Основні способи розв'язання проблеми неоднозначності термінів

Під багатозначністю будемо розуміти наступне: оскільки термін виражає певне поняття у вигляді слова або словосполучення, можливі випадки, коли одному і тому ж терміну природної мови відповідають одразу декілька понять (об'єктів, явищ, процесів тощо), іншими словами, ці поняття будуть відображати різні аспекти людського знання, втім транслюватимуться у певну мову однаково. Виходячи з вищесказаного, розв'язанням проблеми багатозначності терміну буде фіксація за цим терміном певного значення. Під значенням терміну необхідно розуміти концепт людського знання (поняття), що має конкретне визначення (дефініцію), яке відрізняє його від інших концептів [5]. Такий концепт характеризується рядом синонімів, через які визначається в тій чи іншій природній мові. Очевидно, що для формалізації значень термінів цілком підходить лінгвістична онтологія [6] з її набором понять, що мають як формалізовані визначення, так і лексичні подання у вигляді слів чи словосполучень однією із мов.

Визначення або фіксація конкретного значення для багатозначного терміна може відбуватися в різний спосіб [7], що в значній мірі залежить від особливостей задачі, в рамках якої розглядається процес розв'язання неоднозначності. Зокрема, найчастіше можна зустріти такі способи:

- визначення значення на основі контексту, в якому вживається багатозначний термін;
- визначення значення на основі найбільш поширених форм вживання (використовується за відсутності контексту, або коли контекст не дозволяє визначити фіксоване значення);
- визначення значення на основі зворотнього зв'язку з користувачем (напр., після видачі деяких результатів пошуку відслідковувати, по яких з них користувач переходив і таким чином коригувати значення терміну із його пошукового запиту);
- визначення значення на основі попереднього досвіду (напр., використання пошукового профілю користувача, профілю інтересів, загальної статистики пошукових запитів тощо).

Звичайно, кожний з цих способів може використовуватись як самостійно, так і в поєднанні з іншими, що часто і зустрічається на практиці.

В основі підходу до розв'язання омонімії, що розглядатиметься надалі, лежить спосіб визначення значення терміну на основі контексту, в якому цей термін використовується. Така задача часто виникає на етапі аналізу текстових даних, наприклад, під час індексації текстових потоків даних або розбору пошукового запиту під час квазісемантичного пошуку тощо. Контекст в рамках сформованої задачі – це деякі опорні терміни, що уже зафіксовані за певними поняттями (будемо називати їх опорними поняттями) і які вживаються разом із багатозначним терміном або знаходяться з ним у деякому зв'язку.

Особливості вибору опорних понять із контекстного оточення неоднозначних термінів

Вибір опорних понять може залежати від багатьох аспектів постановки задачі, а також доступних засобів аналізу текстових даних та певних допоміжних лінгвістичних ресурсів, або наявності результатів попередньої обробки текстових даних тощо. В загальному випадку, в залежності від ситуації, в якості таких понять можна вибирати наступні:

- поняття онтології, що характеризують тематику, в рамках якої розглядаються текстові дані, це можуть бути як категорії, так і певні ключові терміни, якими ці данні позначив автор або редактор;
- уже зафіксовані поняття із околу багатозначного терміна (із даного речення, абзацу, пошукового запиту тощо);
- уже зафіксовані поняття, що є найбільш уживаними в тексті;
- також опорними можуть стати багатозначні терміни, значення яких уже зафіксовані за певним поняттям на попередніх етапах обробки.

Хоча суть підходу до розв'язання неоднозначності терміна в значній мірі не буде залежати від вибору будь-яких із вище згаданих опорних понять, точність результату визначення значення може сильно варіюватись від вдалого або навпаки невдалого їх вибору.

Контекстноорієнтований підхід до розв'язання омонімії терміну на основі лінгвістичної онтології

В загальному випадку суть підходу зводиться до послідовного виконання наступних кроків:

- визначити опорні поняття;
- зафіксувати опорні поняття в онтології;
- визначити всі поняття онтології, що можуть відповідати неоднозначному терміну;

- визначити, як співвідносяться поняття із першої та другої груп;
- оцінити кожне зі співвідношень і визначити найвірогідніше значення багатозначного терміну.

Оскільки всі поняття в лінгвістичній онтології пов'язані між собою певними семантичними відношенням, це дає змогу використати відповідні зв'язки для визначення та оцінки співвідношення понять. Найбільш поширеними зв'язками лінгвістичної онтології можна вважати ієрархічні та асоціативні. Загалом для оцінки близькості між опорним та багатозначним поняттям можна використати як ієрархічний, так і асоціативний тип відношення між поняттями. Використання ієрархічного відношення виглядає більш зручним, оскільки на основі цього виду зв'язку онтологія набуває вигляду впорядкованої ієрархічної структури, що дає змогу оцінити наскільки близько одне від одного знаходяться певні поняття через пошук їх спільного предка. Природа ієрархічного зв'язку дозволяє говорити про збереження певної безпосередньої відповідності між поняттями на різних рівнях ієрархії. Таким чином, підхід, що пропонується в даній статті, базується на використанні ієрархічної впорядкованості онтології. В його основі лежить пошук спільного предка для зафіксованого опорного поняття та кожного із можливих понять, що відповідають багатозначному терміну.

Для візуалізації підходу розглянемо схему фрагмента деякої лінгвістичної онтології на рис. 1. Нехай поняття, яке позначене як S_i, буде опорним концептом, а S_k, S_p та S_b поняття, що є кандидатами на значення для деякого багатозначного терміну. Для того, щоб зафіксувати значення за одним із цих понять, необхідно визначити як співвідносяться в ієрархії онтології кожне з них із опорним поняттям S_i. Якщо подивитись на рисунок, то видно, що найближче і найглибше в ієрархії опорне поняття S_i пов'язане з поняттям S_k через спільного предка S_j. Їх найближчий спільний предок з S_p (поняття, що позначене на рис.1 як S_q) знаходиться вище по ієрархії, а з S_b опорне поняття пов'язано аж через кореневий елемент S_0. Останню ситуацію можна розглядати, як повну відсутність зв'язку між поняттями, інакше кажучи, опорне поняття і поняття - кандидат на значення терміну належать до різних предметних галузей. Інколи про приналежність до різних предметних галузей, у випадку використання онтологій широкого профілю, можна говорити навіть у випадку першого відносно кореня рівня понять ієрархії, а часом і більш низьких.

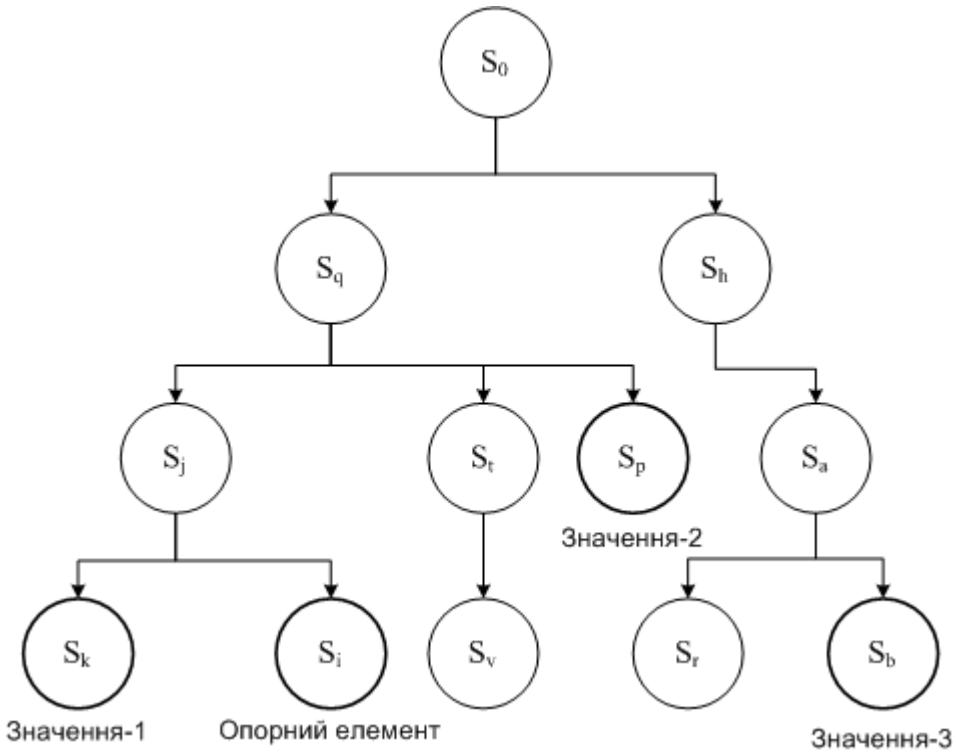

Рисунок 1 - Ілюстрація розв'язання омонімії із застосуванням ієрархічних зв'язків онтології

Оскільки зв'язок між елементами ієрархії встановлюється від найбільш узагальнених понять на найвищих рівнях до більш вузьких галузей знань чи вузько профільних понять на найнижчих рівнях, то чим нижче по ієрархії знаходиться найближчий спільний предок опорного поняття та деякого поняття-кандидата на значення терміну, тим більш вузька категорія їх поєднує. Відповідно саме за таким понят-

тям-кандидатом найвірогідніше має бути зафіксовано значення багатозначного терміна. В наведеній вище ситуації таким елементом є поняття S_k.

Для того, щоб формалізувати даний підхід, необхідно ввести деяку міру ієрархічної близькості, за допомогою якої можна було б оцінити зв'язок між опорним поняттям і кожним із понять-кандидатів на остаточне значення багатозначного терміну. З рис. 1 видно, що ця міра має відображати місцеположення в ієрархії спільного предка відносно опорного поняття та поняття кандидата. Для визначення міри введемо функцію L(S), яка буде відповідати рівню поняття S в ієрархії онтології, тобто найкоротшому шляху (найменшій кількості переходів між поняттями) від поняття S до кореня ієрархічної структури онтології. Оскільки в більшості лінгвістичних онтологій загального призначення важко виділити поняття, яке б узагальнювало всі інші поняття, як правило вводиться технічне поняття (напр., концепт «Все»), що буде служити коренем онтології. Зважаючи на технічний характер цього поняття, воно не виділяється в текстових об'єктах, які обробляються, тому для будь-якого поняття, яке було виділено під час аналізу текстових даних, L(S) завжди більше 0. І тільки $L(S_0)=0$, де S_0 – корінь онтології. Розглянемо можливі варіанти міри близькості опорного поняття S_i і поняття кандидата S_k зі спільним предком S_j:

$$M_1(S_k, S_i) = L(S_j) / L(S_i)$$

$$M_2(S_k, S_i) = 2*L(S_j) / (L(S_i) + L(S_k))$$

$$M_3(S_k, S_i) = (L(S_j))^2 / (L(S_i) * L(S_k)).$$

Як видно із формул розрахунку мір, чим глибше в ієрархії буде знаходитись спільний предок S_j, тим більшим буде значення кожної міри. Коли ж $S_j=S_0$, тобто найближчий спільний батьківський елемент співпадає з кореневим елементом, міри прийматимуть значення 0, що відповідає припущенню приналежності понять S_i та S_k до різних предметних галузей. Найпростіша міра M_1 бере до уваги лише розташування опорного поняття відносно спільного предка і не залежить від розташування поняття-кандидата. За допомогою мір M_2 і M_3 , в свою чергу, ми намагаємось також врахувати розташування і поняття-кандидата. Вони дають змогу зменшити вплив «сторонніх» опорних понять (опорні поняття, спільні предки з якими знаходяться достатньо близько до кореня ієрархічної структури і далеко від поняття-кандидата) на результат вибору значення багатозначного терміна.

Для простої ієрархії функція L(S) понять онтології буде відповідати найменшій кількості переходів від поняття до кореня онтології. При цьому вага кожного переходу розглядається як одиниця і кожний перехід вважається рівноправним, тобто вага кожного переходу однакова. У випадку більш складних ієрархічних структур, елемент онтології може мати декілька батьківських елементів, іншими словами він може належати до різних категорій. В такому разі приналежність до тієї чи іншої категорії може оцінюватись не однаково і, відповідно, кожний перехід буде отримувати деяку вагу W. Тоді розрахунок L(S) буде являти собою деяку більш складну функцію, що залежить від ваг переходів між елементами онтології. Так, у випадку, коли більшому значенню ієрархічного зв'язку відповідає менше значення ваги, найпростіша функція розрахунку рівня поняття в ієрархії може мати вигляд мінімальної суми ваг всіх зв'язків на шляху від поняття до кореня онтології.

Особливості підрахунку мір близькості понять в рамках квазісемантичного підходу до пошуку

Квазісемантичний підхід до пошуку характеризується використанням в якості бази знань онтологічного ресурсу, де кожному поняттю онтології відповідатиме декілька батьківських понять, що відображає природу людського знання, оскільки в більшості випадків важко віднести одне поняття лише до однієї категорії. Така особливість робить структуру ієрархії неоднорідною, оскільки один і той же елемент може знаходитись одразу на різних рівнях в залежності від гілки, через яку до нього спускатись від кореня онтології.

Зважаючи на той факт, що міри близькості понять розраховуються, виходячи із рівнів відповідних понять в ієрархії онтології, в рамках квазісемантичного підходу необхідно врахувати складну структуру ієрархії і відповідно неоднозначність поняття рівня елементу ієрархії. Для цього розглянемо знову фрагмент ієрархії на рис. 1, а саме: опорне поняття S_i і поняття-кандидат S_k зі спільним предком S_j. При цьому будемо вважати, що окрім наведених на малюнку шляхів із S_k в S_j, із S_i в S_j та з S_j в S_0 існують альтернативні шляхи між вказаними парами понять.

Якщо проаналізувати розрахунок кожної міри близькості, можна побачити, що вони являють собою деяке відношення між значенням $L(S_j)$ з одного боку і $L(S_i)$ і $L(S_k)$ з іншого. Причому чим ближче це відношення до 1, тим сильніша міра близькості. Перше, що необхідно врахувати під час розрахунку мір

близькості між поняттями в рамках квазісемантики, це очевидний факт, що значення $L(S_i)$ і $L(S_k)$ мають розраховуватись лише для шляхів, що проходять через S_j. Для неоднорідної ієрархії це забезпечить важливий висновок, що $L(S_i)$ і $L(S_k)$ будуть завжди більші або рівні $L(S_j)$. Очевидно, для того щоб максимально наблизити значення мір близькості до 1, необхідно обрати максимальне значення $L(S_j)$ та мінімізувати $L(S_i)$ і $L(S_k)$. Для того, щоб формалізувати це правило, введемо функцію $D(S_i,S_j)$, яка відображатиме довжину (кількість переходів від одного елемента до іншого строго в сторону кореня онтології) певного шляху між елементами онтології S_i та S_j. Теперь для спільного батьківського елемента S_j функцію $L(S_j)$ можна відобразити наступним чином:

$$L(S_j) = \max(D(S_j,S_0)).$$

А функції $L(S_i)$ і $L(S_k)$ відповідно опорного поняття і поняття-кандидата будуть мати наступний вигляд:

$$L(S_i) = L(S_j) + \min(D(S_i,S_j));$$

$$L(S_k) = L(S_j) + \min(D(S_k,S_j)).$$

Виконання таких умов розрахунку мір близькості відповідає емпіричному визначенню міри близькості понять, оскільки з одного боку забезпечує максимально глибину спільного предка двох понять в рамках ієрархії по відношенню до кореневого елементу (зі збільшенням глибини поняття стають більш вузьконаправленими), а з іншого - забезпечує знаходження саме найближчого спільного предка для опорного поняття і поняття-кандидата.

Приклад використання контекстноорієнтованого підходу до розв'язання неоднозначності термінів

Розглянемо на конкретному прикладі, як можна застосувати наведений підхід до розв'язання неоднозначності термінів під час аналізу текстових даних. В якості допоміжного ресурсу буде використовуватись лінгвістична онтологія, що застосовується в процедурах квазісемантичного пошуку [4], а формування та структура якої описані в [8]. Згадана онтологія є загальною і охоплює більшість сфер людського знання, а в її основі лежить ієрархічна структура понять, всі зв'язки якої є рівноправними, тобто мають однакову вагу. Оскільки ієрархічна структура понять, що використовується в квазісемантичному пошуку, має складну будову, тобто кожний елемент цієї структури може мати декілька батьківських елементів, необхідно враховувати особливості підрахунку мір в рамках квазісемантичного підходу. Для спрощення, ланцюжки з ієрархії, які розглядатимуться нижче, будуть фрагментами онтологічного дерева після відкидання альтернативних шляхів між поняттями, залишаючи лише ті, що задовольняють умовам підрахунку мір для складних ієрархічних структур. Таким чином, для розгляду залишиться фрагмент ієрархії, де кожному поняттю відповідає лише одне батьківське поняття.

В якості вхідних даних візьмемо два речення:

{Ядро} важкого [хімічного елемента] розпадається під час [зіткнення] з [нейтроном].

[Зовнішня мембрана] {ядра} підтримує форму [органели].

В цих двох реченнях в квадратних дужках подані опорні поняття, а в фігурних дужках подані терміни, що можуть мати багато значень, тобто можуть відповідати декільком поняттям онтології. Складемо таблиці, які будуть відображати зв'язок між всіма можливими значеннями терміну «ядро» та кожним із опорних понять, використовуючи кожну з мір близькості.

Для прикладу розглянемо, як було знайдено значення міри M_1 в табл. 1 для понять «Ядро клітини» та «Зовнішня мембрана». Для цього побудуємо ланцюжок шляху від кожного з понять до кореня онтології:

Ядро клітини – Органели – Клітинна біологія – Розділи біології – Біологія – Природничі науки – Природа – Все.

Зовнішня мембрана – Клітинна біологія – Розділи біології – Біологія – Природничі науки – Природа – Все.

Таблиця 1 – Значення міри M_1 близькості між терміном «ядро» та опорними поняттями

Поняття для терміну «ядро»	**Речення №1**				**Речення №2**		
	Хімічний елемент	*Зіткнення*	*Нейтрон*	**Σ**	*Зовнішня мембрана*	*Органела*	**Σ**
Земне ядро	0.2	0.2	0.14	0.54	0.17	0.17	0.33
Ядро операційної системи	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ядро (артилерія)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ядро клітини	0.4	0.4	0.29	1.09	0.83	1.0	1.83
Ядро атома	0.4	0.8	0.43	1.63	0.33	0.33	0.66
Ядро (математика)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ядро (спорт)	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Як видно, найближчим спільним предком для обох понять є поняття «Клітинна біологія». Відповідно, L(«Зовнішня мембрана»)=6, L(«Клітинна біологія»)=5, отже міра:

$$M_1(\text{«Ядро клітини», «Зовнішня мембрана»}) = 5 / 6 \approx 0.83.$$

Аналогічно до попереднього прикладу, розрахуємо значення міри M_2 для тих же понять в табл. 2. Для цього, крім попередніх даних, необхідно знайти L(«Ядро клітини»)=7. Отже міра:

$$M_2(\text{«Ядро клітини», «Зовнішня мембрана»}) = (2 * 5) / (6 + 7) \approx 0.77.$$

Таблиця 2 – Значення міри M_2 близькості між терміном «ядро» та опорними поняттями

Поняття для терміну «ядро»	**Речення №1**				**Речення №2**		
	Хімічний елемент	*Зіткнення*	*Нейтрон*	**Σ**	*Зовнішня мембрана*	*Органела*	**Σ**
Земне ядро	0.17	0.17	0.14	0.48	0.15	0.15	0.3
Ядро операційної системи	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ядро (артилерія)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ядро клітини	0.33	0.33	0.29	0.95	0.77	0.92	1.69
Ядро атома	0.4	0.8	0.5	1.7	0.36	0.36	0.72
Ядро (математика)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ядро (спорт)	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Значення міри M_3 для тих же понять в табл. 3 буде розраховуватись наступним чином:

$$M_3(\text{«Ядро клітини», «Зовнішня мембрана»}) = (5 * 5) / (6 * 7) \approx 0.6.$$

Заповнивши таким чином таблиці, можна знайти інтегральне значення для кожного з понять-кандидатів і (відповідно до результатів) обрати найбільш вірогідне значення терміну «ядро» для кожного з речень. Так, в розглянутому випадку, для першого речення найбільш вірогідним значенням буде поняття «Ядро атома», а для другого - «Ядро клітини», що відповідає дійсному змісту відповідних термінів.

Таблиця 3 – Значення міри M_3 близькості між терміном «ядро» та опорними поняттями

Поняття для терміну «ядро»	**Речення №1**				**Речення №2**		
	Хімічний елемент	*Зіткнення*	*Нейтрон*	**Σ**	*Зовнішня мембрана*	*Органела*	**Σ**
Земне ядро	0.03	0.03	0.02	0.08	0.02	0.02	0.04
Ядро операційної системи	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ядро (артилерія)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ядро клітини	0.11	0.11	0.08	0.3	0.6	0.86	1.46
Ядро атома	0.16	0.64	0.26	1.06	0.13	0.13	0.26
Ядро (математика)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ядро (спорт)	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Висновки

1. В залежності від задач, що розв'язуються, формуються і вимоги до результату вирішення омонімії термінів. Наприклад, в задачі автоматичного перекладу необхідна максимально однозначна фіксація значення терміна. В той же час в пошуковій сфері, навіть результати не повністю розв'язаної проблеми омонімії терміна (залишається ситуація, коли одразу декілька понять в тій чи іншій мірі відповідають контексту проблемного терміну) можуть суттєво підвищити ефективність роботи пошукових механізмів. Так, значення мір близькості понять кандидатів для багатозначного терміну до опорних понять з контексту можуть бути використані для розрахунку поправочних вагових коефіцієнтів під час квазісемантичної індексації текстових даних або на етапі ранжування результатів в рамках квазісемантичного підходу до пошуку. Співвідношення інтегральних значень міри для кожного з понять-кандидатів дає можливість квазісемантичній пошуковій машині під час формування пошукового запиту, у разі виявлення в ньому неоднозначного терміну, запропонувати користувачу для вибору лише ті варіанти понять-кандидатів (у разі неможливості однозначного розв'язання проблеми омонімії), які найбільш тяжіють до контексту.
2. Використання запропонованого в статті контекстноорієнтованого підходу до розв'язання проблеми омонімії термінів дозволяє значно підвищити ефективність застосування квазісемантичної складової в інформаційному пошуку.

Література

1. Zhi Zh. Word sense disambiguation improves information retrieval / Zhi Zhong, Hwee Tou Ng // Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers. - 2012. - P.273-282.
2. Zhang, Word Sense Disambiguation for Improving the Quality of Machine Translation / Zhang, Chun Xiang, Long Deng, Xue Yao Gao, Li Li Guo // Advanced Materials Research. - 2014. - Vol. 981. - P.153-156.
3. Wimmer H. Word Sense Disambiguation for Ontology Learning / H. Wimmer, L. Zhou // Proceedings of the Nineteenth Americas Conference on Information Systems. - 2013. - P.1–10.
4. Михайлюк А.Ю. Квазісемантичний пошук текстових даних в електронному інформаційному ресурсі / Михайлюк А.Ю., Пилипчук О.В., Сніжко М.В., Тарасенко В.П. // Радиоэлектроника и информатика. - Харьков: ХНУРЭ. - 2009. - №3. - С.61-67.
5. Комарова З.И. Методология, метод, методика и технология научных исследований в лингвистике [Электронный ресурс]: учеб. пособие / З.И. Комарова. - 3-е изд., стер. - М.: ФЛИНТА, 2014. - 820 с.
6. Лукашевич Н.В. Тезаурусы в задача информационного поиска / Н.В. Лукашевич. - М.: Издательство Московского университета, 2011. - 512 с.
7. Eneko A. Random Walks for Knowledge-Based Word Sense Disambiguation / Eneko Agirre, Oier de Lacalle, Aitor Soroa // Computational Linguistics. - 2014. - Vol. 40, No. 1. - P.57-84.

8. Михайлюк А.Ю. Автоматизоване формування лінгвістичної онтології на базі структурованого енциклопедичного ресурсу / Михайлюк А.Ю., Пилипчук О.В., Сапсай Т.Г., Тарасенко В.П. // Радіоелектронні і комп'ютерні системи. – 2012. - №4. – С.81-89.

Відомості про авторів:

Михайлюк Антон Юрійович – кандидат технічних наук, викладач кафедри інформатики, Київський Університет ім. Б. Грінченка. Київ, вул. Тимошенка 13б.

Михайлюк Олена Станіславівна – науковий співробітник кафедри СПСКС, НТУУ "КПІ імені Ігоря Сікорського". Київ, пр. Політехнічна 14а.

Пилипчук Олексій Васильович – асистент кафедри СПСКС, НТУУ "КПІ імені Ігоря Сікорського". Київ, вул. Політехнічна 14а.

Тарасенко Володимир Петрович – доктор технічних наук, професор, завідуючий кафедрою СПСКС, НТУУ "КПІ імені Ігоря Сікорського". Київ, вул. Політехнічна 14а, тел. (044) 236-32-02.