



ПІДВИЩЕННЯ ТОЧНОСТІ ПОШУКУ ТЕКСТОВИХ ДОКУМЕНТІВ НА ОСНОВІ АДАПТИВНОЇ ОНТОЛОГІЇ

Романа Даревич

Керівник науково-організаційного відділу Фізико-механічного інституту НАН України,
79601 м. Львів, Наукова, 5, darevych@ipm.lviv.ua

Резюме: Концептуальні графи – є ефективним засобом представлення семантичного змісту текстових документів та онтології предметної області. В статті запропоновано новий метод оцінювання подібності текстів за змістом, який полягає у поданні текстів у вигляді зважених концептуальних графів, доповнених контекстом з онтології, та обчисленні відстані між центрами семантичної ваги цих графів. Показано, що метод задовольняє аксіоми метрики. Розроблено процедури автоматичного налаштування онтології на задану предметну область та інформаційні потреби користувача. Результати експерименту показали, що аналіз текстів з урахуванням семантики вжитих у них понять і словосполучень та коефіцієнтів важливості взятих з адаптивної онтології в середньому на 20% підвищує точність пошуку текстових документів.

Ключові слова: Адаптивна онтологія, інформаційні потреби користувачів, зважені концептуальні графи.

1. ВСТУП

Швидкий розвиток галузі інформаційного пошуку, зокрема, побудови інтелектуальних метапошукових систем пов'язаний із появою та розбудовою глобальної комп'ютерної мережі Інтернет, яка створила принципово нові умови та можливості застосування інформаційних технологій для доступу, пошуку, опрацювання та зберігання інформації. За таких обставин опрацювання значних масивів текстових даних вимагає застосування інформаційно-пошукових систем, які ґрунтуються на інтелектуальних алгоритмах аналізу.

Огляд існуючих підходів до вирішення проблеми створення вискоелективних технологій автоматизації інформаційного пошуку текстових документів показав переваги застосування адаптивних інтелектуальних метапошукових систем здатних точно відображати інформаційні потреби своїх користувачів. Залежно від моделі пошуку інформаційні потреби представляють множиною ключових слів, текстом-прототипом чи одним з методів подання знань. Статистичні та семантичні методи порівняння текстових документів (векторно-просторова модель, міра на основі коефіцієнта Дайса, латентно-семантичне індексування, порівняння концептуальних графів), відповідні до способів подання інформаційних потреб, були запропоновані свого часу Фолцом та Думайсом, Салтоном,

Расмусеном, Монтез-Гомезом та іншими [1-3].

Загальним недоліком розроблених методів є недостатня точність інформаційного пошуку, яка оцінюється як відсоток документів, релевантних до запиту. В той же час для автоматизованих систем інформаційного пошуку, в яких не передбачено інтерактивної взаємодії системи з користувачем, точність пошуку має вирішальне значення.

Одним з способів підвищення точності пошуку є використання онтології предметної області (ПО), як для конструювання запиту [4, 5], так і для безпосереднього аналізу документів [3, 6]. Але у згаданих працях онтологія адекватно не відображає інформаційні потреби, оскільки є статичною, вагові коефіцієнти понять та зв'язків між ними призначаються вручну, що виключає її ефективну реалізацію у автоматизованих метапошукових системах.

Тому актуальною є задача розроблення методів та алгоритмів порівняння за змістом текстових документів доповнених контекстом з адаптованої до інформаційних потреб користувача онтології.

2. КОНЦЕПТУАЛЬНИЙ ГРАФ ТА ОНТОЛОГІЯ

Для аналізу текстовий документ ефективно представляти у вигляді концептуального графа, враховуючи його семантичні виразні можливості. *Концептуальний граф* - це

скінчений, орієнтований, зв'язний двосторонній граф [7]. Два різні види вершин цього графа - це поняття і зв'язки між ними.

Більшість методів порівняння концептуальних графів зосереджено в області пошуку інформації. Деякі з них визначають чи у графі який представляє документ повністю міститься граф запиту, при чому жодна міра їх подібності не застосовується [8]. Більш загальні методи, вимірюють подібність між двома концептуальними графами, але описують цю подібність як набір всіх спільних елементів дозволяючи дублювання інформації [9, 10]. Гнучкіший метод – метод Монтеса-Гомеза, в якому порівнюються концептуальні графи доповнені контекстом з тезаурусу враховуючи вагові коефіцієнти, визначені вручну [11].

Ідею такого підходу взято за основу в даній роботі та розвинуто для пошуку текстових документів забезпечуючи вищу точність. При цьому:

- Доповнення контекстом відбувається з адаптованої до потреб користувачів онтології;
- Вага понять та зв'язків онтології змінюється динамічно під час експлуатації;
- Оцінювання подібності концептуальних графів відбувається через порівняння відстані між центрами семантичної ваги таких графів.

Адаптивна онтологія - таксономія понять, пов'язаних семантичними зв'язками, здатна налаштовуватись на певну предметну область шляхом зміни своєї структури і значень параметрів, і може бути представлена зваженим концептуальним графом (КГ). Основна властивість адаптивної онтології - її здатність в процесі експлуатації інтелектуальної системи динамічно формуватися, та шляхом періодичної оптимізації структури та змісту такої онтології налаштовуватися на задану ПО.

Головним інструментом для реалізації оптимізації є автоматичне зважування понять та зв'язків між ними в онтології. Цю роль в роботі беруть на себе коефіцієнти важливості, де **коефіцієнт важливості поняття (зв'язку)** – це число, котре характеризує значимість даного поняття (зв'язку) у конкретній предметній області і змінюється за визначеним алгоритмом під час опрацювання текстових документів. При цьому коефіцієнти важливості мають задовільняти наступні вимоги:

- відображати семантичну вагу понять предметної області, в якій інтелектуальна система застосовується;

- формуватися під час наповнення онтології та коректується у відповідності з визначеними правилами;
- забезпечувати контроль цілісності онтології;
- відповідати вимогам метрики при їх використанні для порівняння семантичної близькості понять.

Детально метод розрахунку коефіцієнтів важливості, який створює передумови для адаптації онтології до інформаційних потреб користувача, тобто дозволяє сконцентрувати знання з даної ПО відкинувши такі, що мають малу вагу, а значить в даному колі задач не використовуються, розроблено авторами і описано в роботі [12].

3. АДАПТАЦІЯ ОНТОЛОГІЇ ДО ІНФОРМАЦІЙНИХ ПОТРЕБ КОРИСТУВАЧІВ

Процедури адаптації онтології полягають у нормалізації її структури на основі критерію цілісності (відсутність суперечностей і надлишковостей) та оптимізації змісту – процедури вилучення тих знань які рідко використовуються.

Під час наповнення онтології поданої зваженим графом, де вага ребра відображає важливість представленого ним зв'язку, визначена через статистику його вживання, а надлишковість проявляється у вигляді паралельних ребер, циклів та петель, процедура усунення особливостей, що виникають, полягає у послідовному вилученні ребер з мінімальною вагою при збереженні зв'язності усього графа. Така задача розв'язується методом виділення мінімального остову.

Поява взаємо-суперечливих тверджень в структурі призводить до виникнення внутрішніх логічних конфліктів, порушуючи таким чином цілісність онтології, тому система повинна мати здатність своєчасно виявляти за допомогою методу резолюцій, та вилучати їх на основі коефіцієнта достовірності джерела твердження.

Достовірність джерела твердження – означається як імовірність отримання від нього істинного твердження $D_n = P(s = 1)$. Для незнайомого джерела апріорна достовірність встановлюється рівною 0,5. Апостеріорна імовірність в процесі поступової перевірки істинності s наданих n -м джерелом тверджень визначається за формулою:

$$D_{n,i+1} = \frac{D_{n,i}}{(2-s)} + \frac{1-D_{n,i}}{2} \cdot s \quad (1)$$

де s – істинність твердження, що приймає значення 1 у випадку, якщо твердження істинне, або 0 – у протилежному випадку, i – номер кроку підтвердження/заперечення істинності одного з тверджень n -го джерела.

Після перевірки структурної узгодженості онтології при необхідності задоволення обмеження на максимальний об'єм серед істинних вибираються твердження, коефіцієнт важливості яких найменший, з метою їх подальшого вилучення.

Порядок внесення тверджень до множини тих, що підлягають вилученню не впливає на їх вагу, тому маємо лінійну задачу оптимізації змісту, для рішення якої її можна представити як задачу про рюкзак. Отже, нехай онтологія складається з n -елементів, загальним об'ємом пам'яті M . У ролі „рюкзака” виступатиме 10% об'єму, у який слід віднести найменш цінні елементи (поняття з мінімальною семантичною вагою та максимальним об'ємом) з метою подальшого їх вилучення. Тобто максимізувати сумарний зиск:

$$\sum_{i=1}^n \frac{1}{W_i} x_i \text{ таких елементів } i, \text{ для яких } \sum_{i=1}^n m_i x_i \leq N$$

та $W_i > 0, m_i > 0, i = 1, \dots, n$

де x_i – поняття онтології, $x_i = 1$ якщо поняття вносимо в „рюкзак” та 0 – у протилежному випадку; W_i – вага поняття; m_i – об'єм пам'яті, який займає цей елемент; N – частка об'єму пам'яті займаної онтологією до якої слід віднести найменш цінні елементи, наприклад, $N = 1/10M$.

Для вирішення цієї задачі ефективно застосувати жадібний алгоритм, який закладено в розробленій імітаційній моделі онтології, описаній далі.

Не менш важливою задачею в даній роботі є визначення оптимальної кількості понять онтології для конкретного користувача враховуючи два критерії: швидкодія (E) та повнота (G). Цільову функцію запишемо як лінійну згортку цих критеріїв:

$$f = E + \frac{1}{G} \rightarrow \min, \quad (2)$$

$$E = \frac{N \sum_{i=1}^k \left[\frac{e^{-\lambda} \lambda^i}{i!} (i+k) \right]}{K}, \quad K = N \sum_{i=1}^k \frac{e^{-\lambda} \lambda^i}{i!},$$

де E – швидкодія виражена як середній ексцентриситет вершин графа, що представляє онтологію, k – к-ть рівнів у графі; G – відносна кількість понять у онтології, $G = K/N$, K – кількість понять в онтології, N – кількість понять в словнику ПО.

Встановлено, що для таким чином означених критеріїв та словника який містить 100000 понять оптимальна їх кількість в онтології становить 31000. (Див. Рис. 1).

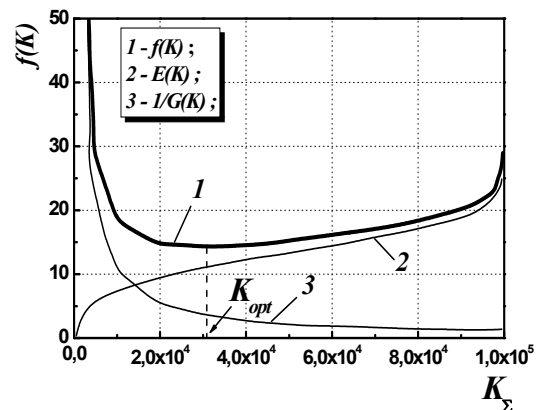


Рис. 1 - Цільова функція пошуку оптимальної кількості понять онтології

Тому онтологія в процесі експлуатації наповнюється до цього об'єму з подальшим циклічним вилученням заданої частки найменш важливих понять методом, описаним вище.

4. ІМІТАЦІЙНА МОДЕЛЬ ОНТОЛОГІЇ

З метою дослідження запропонованих в роботі підходів побудовано імітаційну модель процесу генерування та оптимізації онтології.

Для побудови адекватної моделі досліджено розподіл понять за рівнями (Див. Рис. 2) у таксономічній ієрархії англійської мови за семантичною базою даних WordNet, розробки Принстонського університету (США). Встановлено, що з достатньою для імітаційного моделювання точністю розподіл може бути поданий через закон розподілу Пуассона при $\lambda \approx 8$:

$$p_i = \frac{e^{-\lambda} \lambda^i}{i!} \quad (3)$$

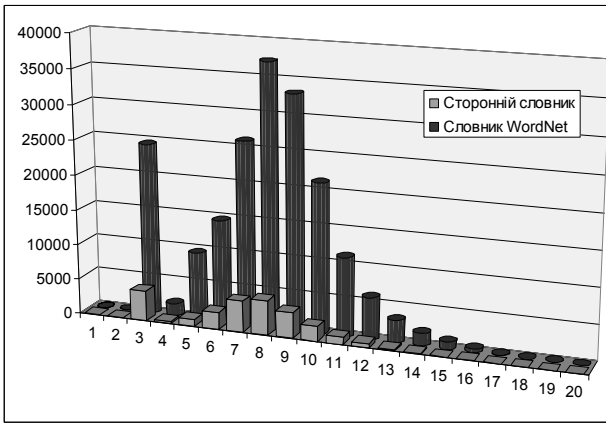


Рис. 2 - Статистика розподілу понять за рівнями в таксономічній ієрархії WordNet

Моделювання виконувалося в середовищі Delphi. Розроблене програмне забезпечення дозволяє генерувати модель онтології до заданої кількості понять та розраховувати семантичну вагу цих понять та зв'язків між ними в процесі модельної експлуатації. Модель імітує функціонування онтології в трьох режимах:

- Наповнення – додавання нових понять;
- Робота – нарощування кількості зв'язків, перерахунок ваги;
- Оптимізація – вилучення заданої частки елементів онтології з найменшою вагою.

З використанням моделі досліджено ефект адаптації інтелектуальної метапошукової системи до заданої предметної області. Встановлено, що через 40-50 циклів (наповнення, робота, оптимізація) відносна кількість видалених елементів, що були додані в попередньому циклі роботи сягає ~ 70 % і майже не змінюється (Рис. 3). Це означає, що критична маса важливих для даної ПО понять внесена в онтологію.

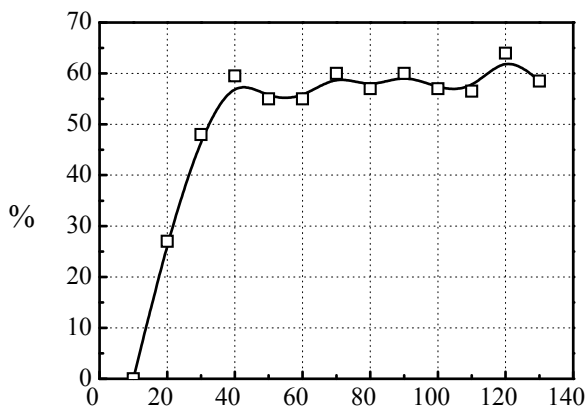


Рис. 3 - Адаптація онтології до предметної області метапошукової системи

В результаті зважений концептуальний граф онтології найбільш повно представляє дану ПО. Зважений концептуальний граф також може бути використаний для представлення текстового документа, вагові коефіцієнти при цьому беруться з налаштованої під дану предметну область онтології.

5. ОЦІНЮВАННЯ ПОДІБНОСТІ ДОКУМЕНТІВ ДОПОВНЕНИХ КОНТЕКСТОМ З ОНТОЛОГІЇ

Для порівняння текстових документів представлених у вигляді зважених концептуальних графів запропоновано визначати центр їх семантичної ваги та оцінювати подібність документів за змістом як відстань між такими центрами. Це дозволяє, по-перше, порівнювати тексти незалежно від їх розміру, по-друге, оцінювати релевантність досліджуваного тексту до заданої онтології, представленої відповідним концептуальним графом.

Алгоритм оцінювання подібності текстів за змістом полягає у наступному:

1. Порівнювані тексти подаємо у вигляді їх концептуальних графів;
2. Графи доповнюємо відповідним контекстом та коефіцієнтами важливості з адаптивної онтології;
3. Відстань між двома вершинами графа C_i та C_j , якщо ці вершини з'єднані дугою, визначаємо як:

$$d_{ij} = \frac{Q}{L_{ij}(W_i + W_j)} \quad (4)$$

де W_i та W_j – коефіцієнти важливості вершин C_i та C_j відповідно; L_{ij} – коефіцієнт важливості зв'язку між вершинами; Q – константа, яка залежить від конкретної онтології. Прийнемо, що $L_{ii} = \infty$, тоді $d_{ii} = 0$.

4. Знаходимо центр ваг концептуального графа. Це вершина C_{i^*} для якої середня відстань \bar{d}_i найменша:

$$\bar{d}_{i^*} = \min_i \bar{d}_i \quad (5)$$

Середня відстань \bar{d}_i для вершини C_i обчислюється згідно формули:

$$\bar{d}_i = \frac{\sum_{j=1, j \neq i}^n d_{ij}^*}{n-1} \quad (6)$$

де n - кількість вершин графа, d_{ij}^* - найкоротший шлях між вершинами C_i та C_j , який обчислюється за допомогою відомих алгоритмів, наприклад Форда, Дейкстри, Флойда-Уоршалла [13];

5. Накладаємо порівнювані графи.

а) Якщо вони мають спільні дуги, то відстань між вершинами, що з'єднані такими дугами, визначається як середня відстань двох графів:

$$\bar{d}^{12} = \frac{\bar{d}^1 + \bar{d}^2}{2} \quad (7)$$

б) Якщо дуги не є спільними, то відстань між вершинами береться із відповідного графа.

6. Обчислюємо найкоротший шлях між центрами ваг КГ, яка буде служити оцінкою подібності двох електронних документів.

$$d^{12} = \min d(C^1, C^2) \quad (8)$$

де C^1 - центр ваги 1-го графа, C^2 - центр ваги 2-го графа. Найкоротший шлях між вершинами обчислюємо за допомогою алгоритму Дейкстри.

За отриманою відстанню визначається подібність між двома документами, яким відповідають дані концептуальні графи.

Основною вимогою до запропонованого методу оцінювання подібності (семантичного порівняння чи рангування) електронних документів є його відповідність аксіомам метрики.

Дійсно, згідно визначення відстані, автоматично виконуються дві перші аксіоми:

$$d(C_i, C_i) = 0;$$

$$d(C_i, C_j) = d(C_j, C_i).$$

Нехай R_{ij}^* - шлях між вершинами C_i та C_j , який відповідає відстані між ними. Тоді $d_{ij} = d_{ik} + d_{kj}$, якщо вершина C_k лежить на шляху R_{ij}^* і $d_{ij} < d_{ik} + d_{kj}$, якщо вершина C_k не лежить на шляху R_{ij}^* . А це означає, що виконується третя аксіома метрики.

Визначену таким чином відстань можна використовувати для рангування текстових

документів, знаходження їх подібності до взірцевого документа тощо.

6. РЕЗУЛЬТАТИ ПОРІВНЯННЯ АНОТАЦІЙ НАУКОВИХ СТАТЕЙ

Проведено наступний експеримент з анотаціями наукових публікацій в області штучного інтелекту який засвідчив, що запропонований у цій роботі підхід на основі адаптивної онтології підвищує точність пошуку документів у середньому на 20%.

Для цього з ключових слів анотації-взірця сформовано запит в мережу Інтернет. В результаті отримано 23 анотації з сайту наукових публікацій CiteSeer. За трьома методами: методом концептуальних графів (Монтеса-Гомеса), коефіцієнтом Дайса (варіант векторно-просторової моделі) та методом розробленим в даній роботі проводився порівняльний аналіз отриманих анотацій із взірцевою. Оцінювання ефективності цих методів для інформаційного пошуку зроблено за параметром – точність пошуку.

$$\text{точність} = \frac{\text{число_знайдених_релевантних(експерт)}}{\text{число_усіх_знайдених(програма)}}$$

Отримані результати представлено в Табл. 1:

Таблиця 1 - Результати порівняння

Методи	Точність
Метод за коефіцієнтом Дайса	10/15=0,66 (66%)
Метод Монтеса-Гомеса	9/12=0,75 (75%)
Метод зважених концептуальних графів (розроблений)	11/12=0,916 (92%)

Слід також зазначити, що метод порівняння за Дайсом у 40% визначав найбільш подібними до взірця ті анотації, що мали найбільшу кількість спільних слів, при цьому найменше відповідали прототипу за змістом. Метод концептуальних графів, враховуючи лише кількість спільних зв'язків, також не дав задовільний результат. У той же час врахування апріорної інформації про предметну область, через зважування вершин та зв'язків концептуальних графів еталонної та досліджуваної анотації дозволило виділити найбільш відповідні до взірця анотації.

Цей експеримент ілюструє ефективність застосування розробленого в роботі підходу для автоматизації пошуку документів, котрі

найбільше відповідають запиту-прототипу і може бути застосований при побудові інтелектуальних метапошукових систем.

7. ВИСНОВКИ

На основі критеріїв швидкодії та повноти розроблено процедури оптимізації онтології адаптивної метапошукової системи, який полягає у періодичному її доповненні при вилученні тих елементів, семантичне значення яких для системи найменше. В методі також враховано необхідність виявлення і усунення суперечності та надлишковості під час наповнення онтології, що відповідає дотриманню критерію цілісності онтології. Запропонований метод забезпечує налаштування предметної області метапошукової системи до актуальних інформаційних потреб користувачів.

Застосування методу оцінювання подібності та процедур оптимізації онтології в архітектурі інтелектуальної метапошукової системи дає змогу автоматизувати інформаційний пошук електронних текстових документів у мережі Інтернет, тим самим підвищити ефективність наукових досліджень та вивести їх на якісно новий рівень.

8. СПИСОК ЛІТЕРАТУРИ

- [1] P. Foltz, S. Dumais. Personalised Information Delivery: Analysis of Information Filtering Methods. *Communications of the ACM* 35(12), 1992.
- [2] E. Rasmussen. Clustering Algorithms. Information Retrieval: Data Structures & Algorithms. *William B. Frakes and Ricardo Baeza-Yates (Eds.), Prentice Hall*, 1992.
- [3] M. Montes-y-Gómez, A. Gelbukh, A. López-López. Comparison of Conceptual Graphs. *Mexican International Conference on Artificial Intelligence MICA I 2000, Acapulco, Mexico, April 2000. Lecture Notes in Artificial Intelligence* N 1793, Springer-Verlag, 2000.
- [4] H. Bulskov, R. Knappe, T. Andreasen. On Querying Ontologies and Databases. *6th International Conference on Flexible Query Answering Systems. Lyon, France, June 24-26. Springer-Verlag. Lecture Notes in Artificial Intelligence*, 3055, 2004.
- [5] Д. Ночевнов. Методи та засоби адаптивного інформаційного пошуку на основі моделі користувача: Автореф. дис. канд. техн. наук: 05.13.06/Черкаський держ. технологічний ун-т. — Черкаси, 2005. — 20с.
- [6] Wang Hui-jin, Hu Hua, Li Qing. A dynamic

knowledge base based search engine. *Journal of Zhejiang University Science*, 2005 6A(7), pp. 683-688.

- [7] John F Sowa. "Knowledge Representation: Logical, Philosophical and Computational Foundations". 1-st edition, Thomson Learning, 1999.
- [8] T. Huibers, I. Ounis, J. Chevallet "Conceptual Graph Aboutness", *Proceedings of the 4th International Conference on Conceptual Structures (ICCS'96), Sydney, Australia, Lecture Notes in Artificial Intelligence, Springer*. 1996., pp. 130-144.
- [9] D. Genest, M. Chein. "An Experiment in Document Retrieval Using Conceptual Graphs". *Lecture Notes in artificial Intelligence* 1257, August 1997.
- [10] H. Myaeng, A. López-López "Conceptual Graph Matching: a Flexible Algorithm and Experiments", *Journal of Experimental and Theoretical Artificial Intelligence*, Vol. 4, 1992.
- [11] M. Montes-y-Gómez, A. Gelbukh, A. López-López, R. Baeza-Yates. Flexible Comparison of Conceptual Graphs. *12th International Conference on Database and Expert Systems Applications DEXA 2001, Munich, Germany, September 2001. Lecture Notes in Computer Science*, vol. 2113, Springer-Verlag, 2001.
- [12] Даревич Р.Р., Досин Д.Г., В.В.Литвин. Метод автоматичного визначення інформаційної ваги понять в онтології бази знань // *Відбір та обробка інформації*. 2005.-Вип. 22(98).— с.105-111
- [13] Р. Седжвик. *Фундаментальные алгоритмы на С++*. Алгоритмы на графах: Пер. с англ./Роберт Седжвик. - СПб: ООО "ДиаСофтЮП", 2002. - 496 с.



Даревич Романа Романівна – керівник науково-організаційного відділу Фізико-механічного інституту ім. Г.В. Карпенка НАН України. У 2002 році отримала диплом магістра закінчивши Інститут комп'ютерних технологій, автоматики та метрології Національного університету "Львівська політехніка". З цього ж року працюю у Фізико-механічному інституті.
Наукові інтереси: інтелектуальні метапошукові системи, онтології та бази знань – їх створення та автоматичне налаштування на інформаційні потреби користувачів.

RISING OF THE TEXT DOCUMENTS SEARCH PRECISION BY USING THE ADAPTIVE ONTOLOGY

Romana Darevych

Chief of scientific-organizational department of Karpenko Physico-Mechanical Institute of NASU
Naukova str., 5, Lviv, 79601 darevych@ipm.lviv.ua

Abstract: *Conceptual graphs are an effective tool for representation of the semantic content of text documents and domain ontology as well. In this article the new method of evaluation of text documents content similarity is proposed. The method consists in representation compared texts as its weighted conceptual graphs supplemented by related context from domain ontology and estimation of a distance between semantic weights centers of these graphs. It is shown that the method satisfies axioms of a metric. Procedures of the automatic tuning of ontology to the specified domain and information needs of user are developed. The results of experiment shows that the taking into account semantics of the used concepts, assertions and significance coefficients from adaptive ontology during the text processing rises the search precision on average 20 %.*

Key words: *Adaptive ontology, users information needs, weighted conceptual graphs.*

The rapid evolution of the information search, particularly creation of intelligent meta-searching systems is related to Internet development. The Internet create fundamentally new conditions and possibilities for widely adopt of information technologies of accessing, searching, processing and saving of information. The advantages in using adaptive ontologies for meta-searching systems were revealed by the review of present approaches for solving a problem of creation of high-effective technologies for automation of information searching of text documents. User information needs can be represented more exactly by using an adaptive ontology. It is shown that one of ways to increase the searching accuracy is using for text documents similarity estimation an adaptive ontology of domain.

In this article the new method of evaluation of text documents content similarity is proposed. The method consists in representation of compared texts as weighted conceptual graphs supplemented by related context from domain ontology and estimation of a distance between semantic weights centers of these graphs. Weights (significance coefficients) of concepts and relations between them in the ontology are changing dynamically during exploitation of the system.

The definition of an adaptive ontology as a concepts taxonomy linked by semantic relations with possibility of adjustment to the specified domain and user information needs by way of change of the structure and values of parameters is given. Also it is founded that conceptual graphs are an effective tool for representation of the semantic content of text document and domain ontology as

well.

Based on rate and completeness criteria procedures of ontology optimization of adaptive meta-searching system were developed. The optimal number of concepts in the ontology is substantiated. It is shown that for common domain vocabulary of 100 000 concepts optimal ontology volume is 31000 concepts. In proposed method the necessity of revealing and elimination of contradictions and redundancy during ontology supplement has been considered with respect to ontology integrity. Logical conflicts are revealed by means of resolution method and excluded taking into account statements reliability factor.

For a research of proposed approaches the simulation model of ontology generation and optimization processes was developed. This software enables us to generate an ontology model until given number of concepts will be reached. During generation semantic weights of these concepts and relations among them is calculated. The model simulates an ontology working in three modes:

- Supplementation – adding new elements;
- Working – increasing number of relations, weights recalculation;
- Optimization – exclusion of minimal weighted elements up to established number.

Using the model the adaptation effect of intelligent metasearching system to given knowledge domain has been investigated. It is shown that after 40-50 cycles (supplementation-working-optimization) the ontology is adjusted to the given

knowledge domain and more correctly represents user information needs.

Comparison of text documents by the content similarity is considered in details. In order to solve this problem text documents were proposed to be presented as weighted conceptual graphs. Weights of concepts are specified from the adaptive ontology. For texts comparison its conceptual graphs semantic weight centers are determined. A distance between these centers defines a similarity among texts by its content. This method make possible to compare texts or estimate a relevance of investigated text to given ontology irrespectively to its size. It is shown that the method satisfies all axioms of metrics.

In this work a comparative analysis of set of abstracts relatively to sample abstract, given by user, using three methods (Montez-Gomes method of conceptual graphs, method of Dice coefficients and developed in this work method) was carried out. Effectiveness of these information searching

methods was estimated using searching accuracy, as a ratio between numbers of found documents, identified by expert as relevant, to total found documents, obtained from user analysis (by software).

The results of experiment shows that in case of using the developed in this work method for comparison text documents with taking into account its semantics, search precision is the greatest (92 %), in contrast to other mentioned methods (66 % and 75 %, respectively). Thus using adaptive ontology significance coefficients during text analysis raises the search precision on average 20 %.

Application of text comparison method and correspondent ontology optimization procedures of inside the architecture of intelligent metasearching systems enables us to automate the information searching process of electronic text documents in the Internet and increase the effectiveness of scientific researches.