

УДК 004.89

МЕТОД КОНТЕКСТНОГО ПОШУКУ НА ОСНОВІ ТЕЗАУРУСУ ПРЕДМЕТНОЇ ОБЛАСТІ

В. В. Литвин

Доктор технічних наук, доцент
Кафедра інформаційних систем та мереж*
E-mail: vasy17.lytvyn@gmail.com

О. В. Мороз

Аспірант
Кафедра прикладної лінгвістики*

*Національний університет «Львівська політехніка»
вул. Ст. Бандери, 12, м. Львів, Україна, 79013

У статті запропоновано підхід до побудови семантичної метрики на основі тезаурусу предметної області. Описано процес побудови такого тезаурусу. Запропоновано використати побудовану таким чином базу знань для пошуку потенційних партнерів, які займаються подібними науковими проблемами в предметній області, для якої побудовано тезаурус

Ключові слова: тезаурус, семантична метрика, інтелектуальна пошукова система

В статті пропонується підхід до побудови семантичної метрики на основі тезаурусу предметної області. Описано процес побудови такого тезаурусу. Пропонується використовувати побудовану таким чином базу знань для пошуку потенційних партнерів, які займаються подібними науковими проблемами в предметній області, для якої побудовано тезаурус

Ключевые слова: тезаурус, семантическая метрика, интеллектуальная поисковая система

1. Вступ

Пошукову систему називають інтелектуальною пошуковою системою (ІПС), якщо вона здійснює пошук на основі контексту [1].

Кількість користувачів мережі Інтернет на сьогодні у світі сягає півмільярда і продовжує швидко зростати. Зараз всесвітня павутина налічує понад 3 мільярди документів, а обсяг текстової інформації, що зберігається в ній, складає сотні терабайт. Кількість даних, доступних нам, постійно збільшується, проте наша здатність сприймати і опрацьовувати цю інформацію залишається незмінною. Попит на високоефективні ІПС стрімко зростатиме і надалі, при цьому актуальною науково-технічною задачею є розроблення автоматизованих ІПС, здатних точно відображати інформаційні потреби своїх користувачів. Наприклад, у сфері наукових досліджень виникає задача знаходження партнерів для виконання різного роду грантів, наукових програм (наприклад, Tempus), або пошук опонентів під час захисту наукових робіт. Для розв'язання таких задач пошуку пропонується використати тезаурус предметної області [2 – 3].

Процес функціонування системи такий: користувач вводить набір ключових слів, які належать до тезаурусу й на думку користувача повно описують специфіку певних на-

укових досліджень, в межах яких здійснюється пошук потенційних партнерів. ІПС на основі цих ключових слів, тезаурусу й деякої семантичної метрики знаходить множину релевантних текстових документів, які рангуються згідно до значення семантичної метрики. Автори цих текстових документів будуть нашими потенційними партнерами.

Отже центральною компонентою архітектури ІПС буде тезаурус предметної області як ядро бази знань такої системи. Іншим головним модулем буде модуль обчислення відстані між запитом користувача системи й знайденими текстовими документами.

Загалом архітектура запропонованої ІПС наведена на рис. 1.

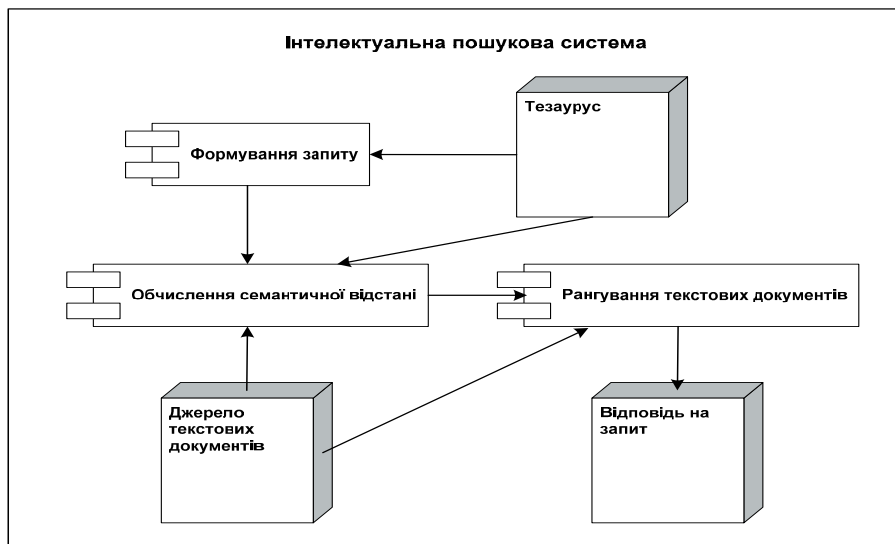


Рис. 1. Архітектура ІПС

Центральним модулем такої системи є модуль обчислення семантичної відстані, який базується на деякій метриці. Проаналізуємо відомі семантичні метрики, які використовуються для розв'язування подібних задач.

2. Аналіз літературних джерел

Виділяють кілька способів визначення семантичних метрик.

У табл. 1 наведені способи обчислення ступеня подібності текстових документів (ТД), засновані на:

- частотності слів у ТД;
- відстані в таксономії понять;
- одночасно і частотності слів, і відстані в таксономії понять.

слів: x і y , де M – це загальне число web-сторінок, проіндексованих Google; $f(x)$ і $f(y)$ – число сторінок, що містять ключові слова x і y , відповідно; $f(x, y)$ – число сторінок, що містять відразу і x , і y . Якщо x і y на всіх сторінках зустрічаються разом, то вважають $NGD=0$, якщо тільки окремо, то $NGD=\infty$.

Виділимо клас метрик, що обчислюють подібність на основі даних таксономії.

Дані метрики використовуються для обчислення подібності концептів WordNet [6], GermaNet, Вікіпедій [4].

У роботі [13] запропонована формула, що враховує як глибину концептів в ієрархії, так і глибину найближчого спільного батька lcs (least common subsumer):

$$wup(C_1, C_2) = \frac{lcs(C_1, C_2)}{\text{depth}(C_1) + \text{depth}(C_2)}$$

Таблиця 1

Класифікація семантичних метрик

| Формула / посилання на опис алгоритму | Назва |
|---|---|
| 1. Частотність слів у ТД | |
| $NGD(x, y) = \frac{\max(\log f(x), \log f(y)) - \log f(x, y)}{\log M - \min(\log f(x), \log f(y))}$ | Нормалізована відстань Google (NGD) |
| $jaccard(x, y) = \frac{\text{Hits}(x \wedge y)}{\text{Hits}(x) + \text{Hits}(y) - \text{Hits}(x \wedge y)}$ | Jaccard [4] |
| 2. Відстані в таксономії понять | |
| Відстань відповідає числу ребер найкоротшого шляху між концептами | Метрика застосовувалася для концептів тезаурусу Роже [5] |
| $lch(C_1, C_2) = -\log \frac{\text{length}(C_1, C_2)}{2D}$ | Leacock & Chodorov 1997, [6] стор. 265-283 |
| $wup(C_1, C_2) = \frac{lcs(C_1, C_2)}{\text{depth}(C_1) + \text{depth}(C_2)}$ | Wu & Palmer [7] |
| $res_{\text{hypo}}(C_1, C_2) = 1 - \frac{\log(\text{hypo}(lcs(C_1, C_2)) + 1)}{\log(C)}$ | Метрика res [8], адаптована до таксономії категорій вікіпедії |
| 3. Частотності слів і відстані в таксономії | |
| $res(C_1, C_2) = \max_{C \in S(C_1, C_2)} [-\log(P(C))]$ | Відстань res [9] |
| $lin(C_1, C_2) = \frac{2 \cdot \log(P(C_0))}{\log(P(C_1)) + \log(P(C_2))}$ | Відстань lin [10] |
| 4. Перетин тексту | |
| перетин тексту (на основі WordNet) | Lesk [11] |
| extended gloss overlap – перетин текстів з врахуванням сусідніх концептів WordNet | Banerjee & Pedersen, 2003 [12] |
| $relate_{\text{gloss/text}}(T_1, T_2) = \tanh \frac{\text{overlap}(T_1, T_2)}{\text{length}(T_1) + \text{length}(T_2)}$ | Відстань $relate$ [4] |

Відстань Google – це міра семантичної зв'язності, яка обчислюється на основі числа сторінок, отриманих за допомогою пошукувача Google для заданого набору ключових слів. У табл. 1 наведена формула обчислення нормалізованої відстані Google (NGD) для двох

Резник [8] запропонував вважати, що два слова тим подібніші, чим інформативніший концепт, до якого співвідносяться ці два слова, тобто чим нижче в таксономії знаходиться спільний верхній концепт (синсет у WordNet). При побудові ймовірнісної функції $P(C)$, вважають, що ймовірність концепту повинна не зменшуватись під час руху нагору за ієрархією:

$$res(C_1, C_2) = \max_{C \in S(C_1, C_2)} [-\log(P(C))].$$

Тоді абстрактніші концепти будуть менше інформативними. Резник запропонував оцінювати ймовірність через частоту синонімів концепту в текстовому документі (ТД) таким чином:

$$P(C) = \frac{\text{freq}(C)}{N},$$

$$\text{freq}(C) = \sum_{\text{newords}(C)} \text{count}(n),$$

де $\text{words}(C)$ – це іменники, що мають значення C ; при цьому N – загальне число іменників у ТД.

У роботі [9] метрика Резника res була адаптована до вікіпедійної інформативності категорії $P(C)$ обчислювалася як функція від числа гіпонімів (категорій у вікіпедії), а не статистично:

$$res_{\text{hypo}}(C_1, C_2) = 1 - \frac{\log(\text{hypo}(lcs(C_1, C_2)) + 1)}{\log(C)},$$

де lcs – найближчий спільний батько концептів C_1 і C_2 , $huro$ – число гіпонімів цього батька, а C – загальне число концептів в ієрархії.

У [10] lin визначає подібність об'єктів A і B як відношення кількості інформації, необхідної для опису подібності A і B , до кількості інформації, що цілком описує A і B . Для виміру подібності між словами lin враховує частотний розподіл слів у тексті (аналогічно до міри Резника):

$$lin(C_1, C_2) = \frac{2 \cdot \log(P(C_0))}{\log(P(C_1)) + \log(P(C_2))},$$

де C_0 – найближчий загальний супер-клас в ієрархії для обох концептів C_1 і C_2 , P – імовірність концепту, що обчислюється на основі частоти появи концепту в ТД. Відрізняється від формули res способом нормалізації, коректним обчисленням $lin(x, x)$ (не залежить від положення концепту x в ієрархії), враховує наявність спільних та відмінних властивостей в об'єктах.

У роботі [4] подібність двох текстів T_1 і T_2 обчислюється з подвійною нормалізацією (по довжині тексту і за допомогою гіперболічного тангенса) як:

$$relate_{gloss/text}(T_1, T_2) = \tanh \frac{overlap(T_1, T_2)}{length(T_1) + length(T_2)},$$

$$overlap(T_1, T_2) = \sum_n m^2,$$

де перетинаються n фраз та m слів.

Отже проведений аналіз показав, що жодна семантична метрика не базується на тезаурусах, лише окремі з них враховують таксономію понять. Тому актуальною є задача розробки методу контекстного пошуку на основі тезаурусу предметної області.

3. Семантична метрика на основі тезаурусу

Тезаурусний словник є переліком логіко-семантичних відношень між лінгвістичними термінами. Цей тезаурус охоплює не тільки множину окремих термінів, поданих у вигляді алфавітного списку з їхніми тлумаченнями, а й самі моделі показу зв'язків між термінами.

На основі досягнень сучасного мовознавства у компактній та доступній формі подано тлумачення термінологічних одиниць з авторитетних джерел (близько тридцяти): термінологічних словників, граматик, монографій. До словника вміщено загальнолінгвістичні терміни, переважно іменники або іменникові словосполучення, з усіх розділів граматики, лексикології, прикладної та комп'ютерної лінгвістики. Словникові одиниці до бази даних відбирали на евристичних заходах (знання укладачів тезаурусу, експертів-лінгвістів).

Словниковими одиницями є вживані загальнолінгвістичні терміни з морфеміки, словотвору, парадигматики, синтаксису, лексикології, семантики; терміни окремих прикладних напрямів, відомих в українській, російській, зарубіжній лінгвістиці; терміни з комп'ютерної лінгвістики, пов'язані з автоматизацією лінгвістичних процесів.

Побудова тезаурусу (ТЗ) передбачає розкриття всіх типів відношень між поняттями, основними з яких є гіпонімія (рід-вид), супідрядність на одному рівні – парціація (частина-ціле), синонімія, кореляція, асоціація, локалізація об'єкта, його призначення, функція, способи вираження функції, відношення тощо. Зміст відношень розширено настільки, щоб можна було охопити максимально широкий пласт термінів, з якими зв'язаний аналізований термін як реєстровий. Оскільки зміст тлумачення був недостатнім для здобуття всіх істотних для термінів відношень, ми орієнтувалися на енциклопедичні словники, наукові праці з конкретної проблематики, знання власні та лінгвістів-фахівців.

Назва відношення є двомісним предикатом $R(A, B)$, який зв'язує заголовне слово статті (A) і введений цим предикатом термін (B) [14].

Множину відношень R поділимо на типи (рід-вид, синонімія, частина-ціле, кореляти тощо) – $R = \{R_1, R_2, \dots, R_k\}$. Позначимо через n_i кількість відношень типу R_i у тезаурусі. Тоді загальна кількість відношень складає $N = \sum_{i=1}^k n_i$. Вважаємо, що тип відношення має більшу вагу, якщо цей тип частіше зустрічається у тезаурусі. Вагу типу відношення визначимо як $L_i = \frac{n_i}{N}$.

Зважимо нашу семантичну мережу, яка задає тезаурус. З цією метою задамо вагу зв'язку між термінами тезаурусу. Чим менша вага, тим терміни є більш подібними. Тому вагу дуг семантичної мережі визначимо як обернено-пропорційну величину до ваги типу відношення, яке задає цю дугу:

$$l_j = \frac{K}{L_i} = \frac{K \cdot N}{n_i}, \tag{1}$$

де K – деяка константа, яка задає величину виміру ваг дуг семантичної мережі [15 – 17]. Використаємо таким чином зважену семантичну мережу для пошуку потенційних партнерів, які займаються подібними науковими проблемами в предметній області для якої побудовано тезаурус.

Для цього визначимо множину ключових термінів $S = \{C_1, C_2, \dots, C_n\}$ взятих із тезаурусу, які на нашу думку найкраще задають проблематику певних наукових досліджень. Пошукова система знаходить множину документів, в яких зустрічаються терміни, що належать до цього тезаурусу. Для кожного такого документу T_s побудуємо множину потужності m , яка складається з термінів, які входять у тезаурус ПО й частіше зустрічаються у документі T_s : $\hat{C}^s = \{\hat{C}_1^s, \hat{C}_2^s, \dots, \hat{C}_m^s\}$.

Методом Флойда-Уоршалла або Дейкстри [18] знайдемо $n \times m$ найкоротших відстаней $d_{ij}^s = d(C_i, C_j^s)$ між термінами з множини S та \hat{C}^s . Тоді відстань до знайденого документу T_s обчислюється згідно до формули: $d^s = \sum_{i=1}^n \sum_{j=1}^m d_{ij}^s$. Рангуємо знайдені документи згідно до зростання величини d^s . Автори документів з вищим рангом можуть бути нашими потенційними клієнтами [19 – 21].

Розглянемо приклад тезаурусу для предметної області – ЛІНГВІСТИКА ТЕКСТУ.

Таблиця відстаней між термінами концептуального графа

| | | | | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 1 | | 1,2 | 2,2 | 3,8 | 2,6 | 4 | 5 | 5,2 | 5,2 | 6,4 | 5,4 | 4 | 7,4 | 6,2 |
| 2 | 1,2 | | 1 | 2,6 | 1,4 | 2,8 | 3,8 | 4 | 4 | 5,2 | 4,2 | 2,8 | 6,2 | 5 |
| 3 | 2,2 | 1 | | 3,6 | 2,4 | 3,8 | 4,8 | 5 | 5 | 6,2 | 5,2 | 3,8 | 7,2 | 6 |
| 4 | 3,8 | 2,6 | 3,6 | | 1,2 | 2,6 | 3,6 | 3,8 | 1,4 | 2,6 | 4 | 2,6 | 3,6 | 2,4 |
| 5 | 2,6 | 1,4 | 2,4 | 1,2 | | 1,4 | 2,4 | 2,6 | 2,6 | 3,8 | 2,8 | 1,4 | 4,8 | 3,6 |
| 6 | 4 | 2,8 | 3,8 | 2,6 | 1,4 | | 1 | 1,2 | 4 | 5,2 | 1,4 | 1,2 | 6,2 | 5 |
| 7 | 5 | 3,8 | 4,8 | 3,6 | 2,4 | 1 | | 2,2 | 5 | 6,2 | 1,5 | 1,2 | 7,2 | 6 |
| 8 | 5,2 | 4 | 5 | 3,8 | 2,6 | 1,2 | 2,2 | | 5,2 | 6,4 | 1,5 | 1,4 | 7,4 | 6,2 |
| 9 | 5,2 | 4 | 5 | 1,4 | 2,6 | 4 | 5 | 5,2 | | 1,2 | 5,4 | 4 | 2,2 | 1 |
| 10 | 6,4 | 5,2 | 6,2 | 2,6 | 3,8 | 5,2 | 6,2 | 6,4 | 1,2 | | 6,6 | 5,2 | 1 | 2,2 |
| 11 | 5,4 | 4,2 | 5,2 | 4 | 2,8 | 1,4 | 1,5 | 1,5 | 5,4 | 6,6 | | 1,4 | 7,6 | 6,4 |
| 12 | 4 | 2,8 | 3,8 | 2,6 | 1,4 | 1,2 | 1,2 | 1,4 | 4 | 5,2 | 1,4 | | 6,2 | 5 |
| 13 | 7,4 | 6,2 | 7,2 | 3,6 | 4,8 | 6,2 | 7,2 | 7,4 | 2,2 | 1 | 7,6 | 6,2 | | 3,2 |
| 14 | 6,2 | 5 | 6 | 2,4 | 3,6 | 5 | 6 | 6,2 | 1 | 2,2 | 6,4 | 5 | 3,2 | |

Вигляд семантичної мережі наведено на рис. 2. У вузлах мережі знаходяться терміни.

Зв'язки між термінами позначені: „С” – синоніми, „К” – корелят, „Г” – голонім, „М” – меронім, „Гіп” – гіперонім.

Використовуючи формулу (1) отримаємо, що вага дуг графа для синонімічного зв'язку рівна 1; кореляту – 1,2; голоніму та мероніму – 1,4; гіпероніму – 1,5. Виходячи з цих ваг, отримаємо зважений концептуальний граф, який наведено на рис. 3.

Знайдемо найкоротші відстані між термінами цього графу, використовуючи алгоритм Флойда-Уоршалла [18]. У результаті отримаємо таку таблицю відстаней (табл. 2).

Приклад. Нехай множина С складається із двох термінів $S = \{ \text{'рубрика', 'текст'} \}$, а деякий текстовий документ Т описується множиною \hat{S} , яка складається із трьох термінів $\hat{S} = \{ \text{'назва', 'когезія', 'нарратив'} \}$, то відстань до такого тексту рівна: $d = 2,2 + 6,4 + 4 + 2,4 + 3,8 + 1,4 = 20,2$.

З метою перевірки ефективності запропонованої у роботі семантичної метрики проведено серію експериментів з оцінювання релевантності знайдених в мережі Інтернет наукових праць (анотацій, документів) до запиту, у ролі якого виступала множина ключових слів розробленого тезаурусу.

Для порівняння виконано чисельні розрахунки подібності знайдених текстів до запиту також за іншими відомими метриками – Джаккарда та WUP. Проведено 10 експериментів за кожним з методів. Результати застосування методів оцінював експерт. З множини документів, визнаних релевантними кожним із методів, він відбирав дійсно релевантні на його думку з метою аналізу точності методів, що порівнювались:

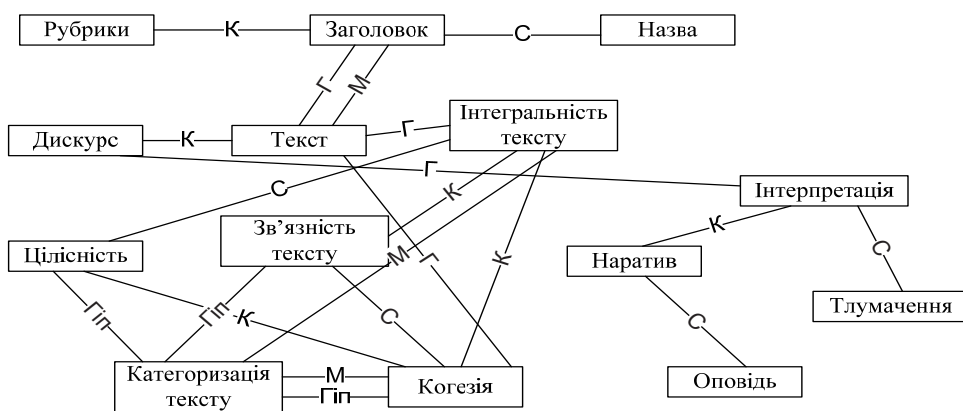


Рис. 2. Семантична мережа тезаурусу розділу „ЛІНГВІСТИКА ТЕКСТУ”

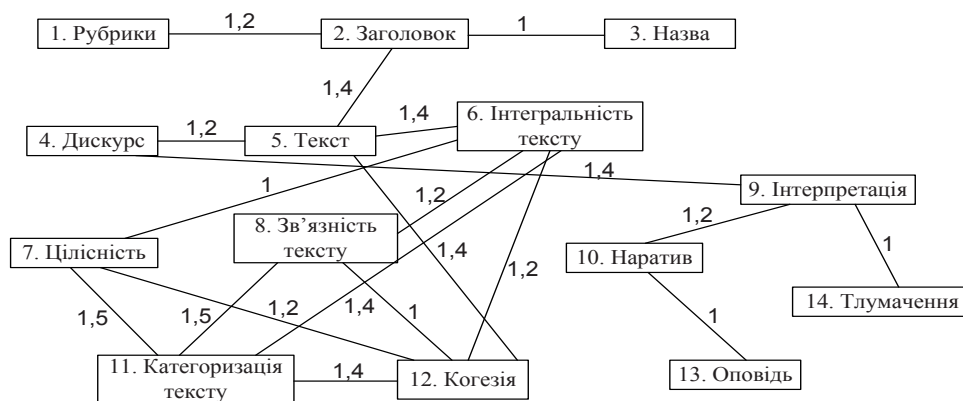


Рис. 3. Зважений концептуальний граф

$$\text{точність} = \frac{\text{число_знайдених_релевантних}}{\text{число_усіх_знайдених}}$$

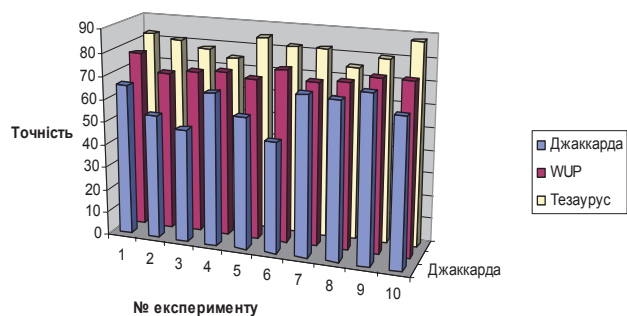


Рис. 4. Результати експериментального дослідження точності методів пошуку релевантних текстових документів

Таким чином, ефективність семантичних метрик оцінювалась за параметром точність τ : $\tau = \frac{r_e}{r_m} \cdot 100$, де $r_e = |R_e|$ – потужність множини R_e релевантних документів, знайдених за відповідним методом (згідно оцінки експерта з даної ПО); $r_m = |R_m|$ – потужність множини R_m усіх документів, знайдених за відповідним методом, $R_e \subseteq R_m$.

Результати десяти експериментів наведено на рис. 4, з якого видно, що точність пошуку системою релевантних документів за розробленим у цьому дослідженні методом оцінювання подібності (темні стовпчики) є вищою, порівняно з методом Джаккарда (штриховані) та WUP (світлі стовпчики).

Загалом ефективність такого підходу, що ґрунтується на використанні семантичної метрики на основі

тезаурусу, визначена за параметром τ , в середньому на 10...20% вища, порівняно з іншими метриками. Поліпшення досягається завдяки зважуванню концептуального графа, який задає тезаурус.

4. Висновки

Отже у роботі запропоновано підхід до побудови семантичної метрики на основі тезаурусу предметної області. Детально описано процес побудови такого тезаурусу як семантичної мережі. Запропоновано задати ваги дуг такої мережі як обернено-пропорційні величини до кількості відношень певного типу. Побудовано семантичну метрику на основі зваженої таким чином семантичної мережі. Запропоновано використати цю метрику для пошуку потенційних партнерів, які займаються подібними науковими проблемами в предметній області, для якої побудовано тезаурус.

Встановлено, що метод порівняння за семантичною метрикою Джаккарда у 40% знайдених текстів визначав найбільш подібними до запиту ті документи, що мали найбільшу кількість спільних слів. Метод за семантичною метрикою WUP, який враховує лише кількість спільних зв'язків, також не дав задовільний результат. У той же час врахування апріорної інформації про предметну область, завдяки використанню тезаурусу та зважування зв'язків концептуальних графів, дає змогу виділяти документи, зміст яких найбільше відповідає множині ключових слів користувача.

Таким чином, використання семантичної метрики на основі тезаурусу, запропонованої в цьому дослідженні, дає змогу підвищити ефективність автоматизованого пошуку релевантних документів.

Література

1. Гладун, А. Я. Формирование тезауруса предметной области как средства моделирования информационных потребностей пользователя при поиске в Интернете [Текст] / А. Я. Гладун, Ю. В. Рогушина // Вестник компьютер. и информ. технологий. – 2007. – № 1. – С. 26-33.
2. Gruber, T. A translation approach to portable ontologies [Текст] / T. Gruber // Knowledge Acquisition. – 1993. – № 5 (2). – P. 199–220.
3. Гаврилова, Т. А. Базы знаний интеллектуальных систем [Текст] / Т. А. Гаврилова, В. Ф. Хорошевский. – СПб.: Питер, 2001. – 384 с.
4. Strube M. WikiRelate! Computing semantic relatedness using Wikipedia. In Proceedings of the 21st National Conference on Artificial Intelligence [Електронний ресурс] / M. Strube, S. Ponzetto. // (AAAI 06). Boston, Mass., July 16-20, 2006. – Режим доступу: <http://www.eml-research.de/english/research/nlp/public>
5. Jarmasz M. Roget's Thesaurus and semantic similarity [Текст] / M. Jarmasz, S. Szpakowicz // In Proceedings of Conference on Recent Advances in Natural Language Processing (RANLP 2003). – Borovets, Bulgaria, September, 2003. – P. 212-219.
6. Fellbaum, C. WordNet: an electronic lexical database [Текст] / C. Fellbaum. – MIT Press, Cambridge, Massachusetts, 1998. – 423 p.
7. Wu, Z. Verb semantics and lexical selection [Текст] / Z. Wu, M. Palmer // In Proc. of ACL-94, 1994. – P. 133-138.
8. Resnik, P. Disambiguating noun groupings with respect to WordNet senses [Електронний ресурс] / P. Resnik // In Proceedings of the 3rd Workshop on Very Large Corpora. MIT, June, 1995. – Режим доступу: <http://xxx.lanl.gov/abs/cmp-lg/9511006>.
9. Resnik, P. Semantic similarity in a taxonomy: an information-based measure and its application to problems of ambiguity in natural language [Текст] / P. Resnik // Journal of Artificial Intelligence Research (JAIR), 1999. – Vol. 11. – P. 95-130.
10. Lin, D. An information-theoretic definition of similarity [Електронний ресурс] / D. Lin // In Proceedings of International Conference on Machine Learning, Madison, Wisconsin, July, 1998. – Режим доступу: <http://www.cs.ualberta.ca/~lindek/papers.htm>.
11. Смирнов, А. В. Онтологии в системах искусственного интеллекта: способы построения и организации [Текст] / А. В. Смирнов, М. П. Пашкин, Н. Г. Шилов, Т. В. Левашова // Новости искусственного интеллекта. – М.: Изд-во РАИИ, 2002. – № 2. – С. 3–9.

12. Совпель, И. В. Система автоматического извлечения знаний из текста и ее приложения [Текст] / И. В. Совпель // Науч.-теорет. журнал "Искусственный интеллект", ПИШ "Наука і освіта". – 2004. – Вып. 3 – С. 668–677.
13. Wu, Z. Verb semantics and lexical selection [Текст] / Z. Wu, M. Palmer // In Proc. of ACL-94, 1994. – P. 133-138.
14. Никитина, С. Е. Тезаурус по теоретической и прикладной лингвистике [Текст] / С. Е. Никитина. – М.: Наука, 1978. – 220 с.
15. Lytvyn, V. Searching the Relevant Precedents in Dataspaces Based on Adaptive Ontology [Текст] / V. Lytvyn, N. Shakhovska, V. Pasichnyk, D. Dosyn // Computational Problems of Electrical Engineering. – 2012. – V. 2, N. 1. – Lviv. – P. 75-81.
16. Dosyn, D. Planning of Intelligent Diagnostics Systems Based Domain Ontology [Текст] / V. Lytvyn, D. Dosyn // The VIIIth International Conference Perspective Technologies and Methods in MEMS Design. – 2012. – Polyana. – P. 103.
17. Lytvyn, V. Intelligent agent on the basis of adaptive ontologies construction [Электронный ресурс] / V. Lytvyn, D. Dosyn, M. Medykovskyj, N. Shakhovska // Signal Modelling Control. – 2011. – Lodz.
18. Свами, М. Графы, сети и алгоритмы [Текст] / М. Свами, К. Тхуласираман. – М.: Наука, 1984. – 256с.
19. Montes-y-Gómez M. Comparison of Conceptual Graphs [Электронный ресурс] / M. Montes-y-Gómez, A. Gelbukh, A. López-López // Lecture Notes in Artificial Intelligence. – 2000. – Vol. 1793. – Springer-Verlag: <http://ccc.inaoep.mx/~mmontesg/publicaciones/2000/ComparisonCG>.
20. Кнаппе, Р. Perspectives on Ontology-based Querying [Электронный ресурс] / R. Кнаппе, Н. Bulskov, Т. Andreasen // International Journal of Intelligent Systems. – 2004. <http://akira.ruc.dk/~knappe/publications/ijis2004.pdf>.
21. Lytvyn, V. Design of intelligent decision support systems using ontological approach [Текст] / V. Lytvyn // An international quarterly journal on economics in technology, new technologies and modelling processes. – 2013. – Vol. II. – No 1. – P. 31-38.

УДК 666.97.07+004

ОСОБЕННОСТИ СОЗДАНИЯ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ПРОЕКТИРОВАНИЯ СОСТАВА БЕТОНА

Розглядаються методи розв'язання задачі проектування складу бетону різних науково-дослідних шкіл матеріалознавства в Україні. Підкреслено доцільність застосування інформаційних технологій для формалізації різних методів проектування складу бетону. Викладено функціональні вимоги до програмного забезпечення, які розв'язують подібні задачі. Представлена логічна структура та інтерфейс користувача програмного забезпечення розробленої інформаційної системи для розв'язку задачі проектування складу бетону

Ключові слова: технологія бетону, проектування складу бетону, інформаційні технології, інтелектуальні системи

Рассматриваются методы решения задачи проектирования состава бетона различных научно-исследовательских школ материаловедения в Украине. Отмечена целесообразность применения информационных технологий для формализации различных методов проектирования состава бетона. Изложены функциональные требования к программному обеспечению, решающему подобные задачи. Представлена логическая структура и интерфейс пользователя программного обеспечения разработанной информационной системы для решения задачи проектирования состава бетона.

Ключевые слова: технология бетона, проектирование состава бетона, информационные технологи, интеллектуальные системы

Н. Д. Сизова

Доктор физико-математических наук,
профессор*

E-mail: sizova@ukr.net

И. А. Михеев

Кандидат технических наук*

*Кафедра экономической кибернетики и
информационных технологий

Харьковский национальный университет
строительства и архитектуры

ул. Сумская, 40, г. Харьков, Украина, 61002
E-mail: i.a.mikheev@gmail.com

1. Введение

Для получения качественного продукта или изделия необходимы качественные материалы (комплектующие), из которых изготавливают конечный продукт. Производство бетона, кроме этого, выдвигает

требование соблюдения рецептуры и технологии производства. Даже использование высококачественных составляющих материалов (цемента высокой марки, щебня необходимой формы и песка с минимальным содержанием пылевидных частиц, модифицирующих высокоэффективных добавок) не гарантирует высокое