

УДК



## ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПІДТРИМКИ КОНСУЛЬТУВАННЯ СТУДЕНТІВ ДИСТАНЦІЙНОЇ ФОРМИ НАВЧАННЯ

Н.С. Лєсна<sup>1</sup>, А.В. Найденко<sup>2</sup>, В.В. Сокол<sup>3</sup>

<sup>1</sup> ХНУРЕ, м. Харків, Україна, lmd@kture.kharkov.ua;

<sup>2</sup> ХНУРЕ, м. Харків, Україна, anaydenko@hotmail.com

<sup>3</sup> ХНУРЕ, м. Харків, Україна, vitalysokol@kture.kharkov.ua

У рамках роботи було досліджено методи покращення консультування студентів дистанційної форми навчання, розглянуто етапи та моделі пошуку інформації, розглянуто та порівняно методи імовірнісного латентно-семантичного аналізу, добування ключових понять із тексту, класифікації та кластеризації текстових документів, автоматичного анотування текстів. В результаті було спроектовано та розроблено інформаційну технологію підтримки консультування студентів дистанційної форми навчання у вигляді веб-застосування з використанням технології ASP.NET. Інформаційна технологія оптимізує процес консультування студентів, зменшуючи об'єм роботи викладача та підвищуючи швидкість отримання студентом відповіді на поставлене запитання, що реалізовано за допомогою імовірнісного латентно-семантичного аналізу, який виявляє приховані фактори та дає змогу добути контекстно-залежні значення слів за допомогою статистичної обробки великих наборів текстових даних.

ДИСТАНЦІЙНЕ НАВЧАННЯ, ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, ІМОВІРНІСНИЙ ЛАТЕНТНО-СЕМАНТИЧНИЙ АНАЛІЗ, TEXT MINING, TFIDF, NLP

### Вступ

Сьогодні дистанційне навчання через Інтернет усе частіше розглядається не просто як зручна форма підвищення кваліфікації, а як альтернатива традиційній освіті, що дозволяє студентові одержати глибокі знання.

Термін «дистанційне навчання» означає таку організацію навчального процесу, при якій викладач розробляє навчальну програму, головним чином основу на самостійному навчанні студента. Основною відмінною рисою дистанційного навчання від інших типів освіти є широке використання телекомунікаційних технологій у навчальному процесі.

Особливу роль у дистанційному навчанні відіграє інтерактивна взаємодія викладача та студентів. У вузькому розумінні слова (стосовно роботи користувача програмного забезпечення взагалі) інтерактивна взаємодія – це діалог користувача з програмою, тобто обмін командами, запитаннями й відповідями. *При дистанційному навчанні засобом здійснення інтерактивної взаємодії є розроблена інформаційна технологія підтримки консультування студентів дистанційної форми навчання. Ця технологія виконує інтелектуальний пошук відповіді перед відправкою питання викладачеві, що дозволяє зменшити об'єм роботи викладача, заощадити час студента завдяки автоматичному швидкому отриманню відповіді на поставлене запитання.*

**Постановка задачі.** В рамках роботи було поставлено наступні задачі:

- ознайомитися з дистанційною формою навчання з використанням дистанційних освітніх технологій, форм, методів, засобів навчання й освітніх масивів мережі Інтернет;

- дослідити способи автоматизації процесу консультування студентів дистанційної форми навчання;

- дослідити методи інтелектуального пошуку у глобальних і локальних обчислювальних мережах і базах даних;

- ознайомитися із засобами інтелектуального пошуку текстової інформації;

- розглянути та вивчити окремі етапи використання Text Mining – глибинного аналізу текстової інформації;

- порівняти методи інтелектуального аналізу тексту для вирішення поставлених задач;

- спроектувати інформаційну технологію підтримки консультування студентів дистанційної форми навчання з використанням інтелектуального аналізу тексту;

- розробити програмне рішення інформаційної технології підтримки консультування студентів дистанційної форми навчання з використанням інтелектуального пошуку інформації.

**Актуальність.** Робота є досить актуальною, оскільки за допомогою інформаційної технології підтримки консультування студентів дистанційної форми навчання:

- множить педагогічні можливості викладачів навчального закладу;

- робить процес навчання більш предметним;

- створює додаткову мотивацію у студентів до вивчення матеріалів;

- заощаджує час викладача і студента за рахунок автоматичного інтелектуального пошуку відповіді на запитання студента до того, як воно буде залишене на розгляд викладача.

### 1. Автоматизація процесу консультування студентів дистанційної форми навчання

Інтерактивна взаємодія викладача та студентів відіграє особливу роль у дистанційному навчанні. Чим більше існує можливостей управляти про-

грамою, чим активніше користувач бере участь у діалозі, тим вище інтерактивність. У широкому розумінні інтерактивна взаємодія припускає діалог будь-яких суб'єктів один з одним з використанням доступних їм засобів і методів. При цьому передбачається активна участь у діалозі обох сторін – обмін запитаннями та відповідями, керування ходом діалогу, контроль над виконанням ухвалених рішень тощо. При дистанційному навчанні суб'єктами в інтерактивній взаємодії будуть виступати викладачі та студенти, а засобом здійснення подібної взаємодії – розроблена інформаційна технологія підтримки консультування студентів дистанційної форми навчання. Ця технологія виконує інтелектуальний пошук відповіді перед відправленням запитання викладачеві, що дозволяє зменшити об'єм роботи викладача, заощадити час студента завдяки автоматичному швидкому отриманню відповіді на поставлене запитання.

Існує широкий спектр пропонуваніх рішень і перспективних напрямків досліджень в області інформаційного пошуку, починаючи від побудови глобальних розподілених інформаційних структур і пошукових систем, закінчуючи елементарними на перший погляд питаннями аналізу документів. Усі вони, безумовно, важливі та корисні при розв'язку своїх специфічних задач. Проте, саме від методів аналізу багато в чому залежить ефективність існуючих пошукових систем, тому що вони є основою будь-якої пошукової системи та багато в чому визначають можливості й обмеження цих систем.

Крім цього існує ще один важливий фактор, що визначає ефективність будь-якого інформаційного пошуку – це людський фактор. Найчастіше у більшості досліджень в області інформаційного пошуку цей фактор або ігнорується, або його значення багато в чому недооцінюється. Але саме люди в остаточному підсумку користуються розробленими інформаційно-пошуковими системами. Облік людського фактора, специфіки його роботи, переваг і очікувань є перспективним і багатообіцяючим напрямком досліджень.

Представлені на сьогоднішній день у більшості популярних пошукових систем способи організації повнотекстового пошуку та методи аналізу документів не враховують у достатній мірі саме людський фактор. Тобто, не враховується той факт, що багато в чому пошук визначається слабко формалізованими та нечіткими умовами, у значній мірі залежними від досвіду та переваг самої людини. Далеко не завжди користувач інформаційно-пошукової системи може чітко й однозначно сформулювати саме той набір ключових слів, який і приведе його до бажаного результату. Мова йде про варіант пошуку на основі формування інформаційних запитів, що складаються із набору ключових слів. Цей варіант пошуку найпоширеніший і методологічно пророблений на сьогоднішній день.

## 2. Методи інтелектуального пошуку інформації

В рамках даної роботи було досліджено наступні методи інтелектуального аналізу текстових даних з метою пошуку релевантної інформації:

а) метод добування ключових понять із тексту. Первинною метою добування ключових понять (feature extraction) є ідентифікація фактів і відносин у тексті. У більшості випадків такими поняттями є іменники власні та загальні: імена й прізвища людей, назви організацій тощо. Алгоритми добування понять можуть використовувати словники, щоб ідентифікувати деякі терміни та лінгвістичні шаблони для визначення інших;

б) метод класифікації текстових документів. Класифікація (classification) – стандартна задача з області Data Mining. Її метою є визначення для кожного документа однієї або декількох заздалегідь заданих категорій, до яких цей документ відноситься. Особливістю задачі класифікації є припущення, що множина документів, що класифікуються, не містить “сміття”, тобто кожний з документів відповідає якій-небудь заданій категорії. Окремим випадком задачі класифікації є задача визначення тематики документа;

в) метод кластеризації текстових документів. Метою кластеризації (clustering) документів є автоматичне виявлення груп схожих документів серед заданої фіксованої множини. Відзначимо, що групи формуються тільки на основі попарної схожості описів документів, і ніякі характеристики цих груп не задаються заздалегідь;

г) метод автоматичного анотування текстів. Автоматичне анотування (summarization) дозволяє скоротити текст, зберігаючи його зміст. Розв'язок цієї задачі звичайно регулюється користувачем за допомогою визначення кількості добутих речень, або відсотком тексту, що добувається, стосовно усього тексту. Результат містить у собі найбільш значимі речення в тексті;

д) метод імовірнісного латентно-семантичного аналізу. Імовірнісний латентно-семантичний аналіз базується на статистичному аналізі даних, дозволяє зробити мовнезалежний семантичний пошук інформації з можливістю виявлення схованих факторів (тем), присутніх в інформаційному масиві, з урахуванням семантичної близькості слів.

Найбільш підходящим методом, який би дозволив максимально ефективно вирішити нашу задачу інтелектуального пошуку текстової інформації, є метод імовірнісного латентно-семантичного аналізу.

## 3. Семантичний інформаційний пошук

Окремою задачею інформаційного пошуку є латентно-семантичний аналіз тексту. Латентно-семантичний аналіз (LSA – Latent Semantic Analysis)

– це теорія й метод для добування контекстно-залежних значень слів за допомогою статистичної обробки великих наборів текстових даних [1]. Даний метод аналізу використовується не тільки в області пошуку інформації [2], але й у задачах фільтрації й класифікації [3].

Основна ідея латентно-семантичного аналізу полягає в тому, що сукупність усіх контекстів, у яких зустрічається й не зустрічається дане слово, задає множину взаємних обмежень, які дозволяють визначити подібність змістовних слів між собою.

Вихідною інформацією для LSA є матриця термів на документи, яка описує використовуваний для навчання системи набір даних. Елементи цієї матриці містять частоти використання кожного терму в кожному документі.

Один з найпоширеніших варіантів LSA заснований на використанні розкладання вихідної матриці за сингулярними значеннями (SVD – Singular-Value Decomposition). Використовуючи SVD, велика вихідна матриця розкладається в множини із  $k$  ортогональних матриць, лінійна комбінація яких є достатнім наближенням вихідної матриці.

Згідно з теоремою про сингулярне розкладання, будь-яка речовинна прямокутна матриця  $X$  може бути розкладена в добуток трьох матриць:

$$X = U \Sigma V^T,$$

де матриці  $U$  і  $V$  – ортогональні, а  $\Sigma$  – діагональна матриця, значення на діагоналі якої називаються сингулярними значеннями матриці  $X$ .

Особливість такого розкладання в тому, що якщо в  $\Sigma$  залишити тільки  $k$  найбільших сингулярних значень, а в матрицях  $U$  і  $V$  тільки відповідні до цих значень стовпці, то добуток матриць, що вийшли,  $U_{lsa}$ ,  $\Sigma_{lsa}$  і  $V_{lsa}$  буде найкращим наближенням вихідної матриці  $X$  матрицею рангу  $k$ .

$$M = \hat{X} = U_{lsa} \Sigma_{lsa} V_{lsa}.$$

Ідея такого розкладання та суть латентно-семантичного аналізу полягає в тому, що якщо в якості  $X$  використовувалася матриця термів на документ, то матриця  $\hat{X}$ , що містить тільки  $k$  перших лінійно незалежних компонентів  $X$ , відображає основну структуру асоціативних залежностей, присутніх у вихідній матриці, і в той же час не містить шуму.

Таким чином, кожний терм і документ представляються за допомогою векторів у загальному просторі розмірності  $k$  (так званім просторі гіпотез). Близькість між будь-якою комбінацією термів або документів може бути легко обчислена за допомогою скалярного добутку векторів.

Найбільший інтерес у рамках даної роботи викликає метод імовірнісного латентно-семантичного аналізу тексту. Розглянемо його докладніше.

Метод імовірнісного латентно-семантичного аналізу (PLSA – Probabilistic Latent Semantic

Analysis) ставить своєю задачею виявлення латентних, схованих, факторів (тем), присутніх у колекції й пов'язаних з її документами й словами. Саме, фіксувавши число схованих факторів  $r$ , за допомогою методу PLSA можна оцінити наступні величини:

–  $P(z_i)$  – імовірність того, що випадково обраний з колекції документ найбільш тісно пов'язаний з фактором (найбільшою мірою відповідає темі)  $z_i$ ;

–  $P(d_j | z_i)$  – імовірність того, що найбільш тісно пов'язаний з даним фактором  $z_i$  документ – це  $d_j$ ;

–  $P(w_j | z_i)$  – імовірність того, що для даного фактора  $z_i$  найбільш тісно пов'язане з ним слово – це  $w_j$ .

Тут  $D = \{d_1, \dots, d_n\}$  – множина усіх проіндексованих документів,  $W = \{w_1, \dots, w_n\}$  – множина усіх різних слів, що зустрілися в проіндексованих документах,  $Z = \{z_1, \dots, z_r\}$  – множина латентних факторів.

Позначимо за  $P(d, w)$  ймовірність спільного спостереження документа  $d$  й слова  $w$ , тобто  $P(d, w) = P(d)P(w|d)$ . У рамках підходу, використовуваного при імовірнісному латентно-семантичному аналізі, величина  $P(d, w)$  оцінюється за наступною формулою:

$$P(d, w) = \sum_{z \in Z} P(z)P(d|z)P(w|z).$$

Відповідно до принципу максимальної правдоподібності функції  $P(z)$ ,  $P(d|z)$  й  $P(w|z)$  визначаються шляхом максимізації функції правдоподібності  $L = \sum_{d \in D} \sum_{w \in W} n(d, w) \log P(d, w)$ , де  $n(d, w)$  є число входжень слова  $w$  у документ  $d$ .

На етапі максимізації функції  $L$  використовується метод оцінювання-максимізації. На кожній ітерації виконується крок оцінювання

$$P(z|d, w) = \frac{P(z)P(d|z)P(w|z)}{\sum_{z'} P(z')P(d|z')P(w|z')}$$

і потім крок максимізації

$$P(w|z) = \frac{\sum_d n(d, w)P(z|d, w)}{\sum_{d, w'} n(d, w')P(z|d, w')},$$

$$P(d|z) = \frac{\sum_w n(d, w)P(z|d, w)}{\sum_{d', w} n(d', w)P(z|d', w)},$$

$$P(z) = \frac{1}{R} \sum_{d, w} n(d, w)P(z|d, w),$$

$$R = \sum_{d, w} n(d, w).$$

Будемодалізаувати вектор  $(P(d|z_1), \dots, P(d|z_r))$  образом документа  $d$  в просторі факторів  $Z$  і, аналогічно, вектор  $(P(w|z_1), \dots, P(w|z_r))$  – образом слова  $w$ . Позначимо через  $D'$  деяку множину документів, випадково відібраних з множини  $D$

всіх документів, у якій виконується пошук, а через  $W'$  – множину усіх слів з документів, що ввійшли до  $D'$ . Нехай також  $d \in D - D'$ .

Далі припускаємо, що множина  $D'$  може розглядатися у якості представницької вибірки документів з колекції  $D$ , і в ній порушені всі теми, відбиті в повній колекції. Іншими словами, довільний документ  $d$  містить значну кількість слів з  $W'$ , при цьому підмножина  $W'$  достатньо повно відображає тематичну спрямованість  $d$ .

Розглянемо систему лінійних алгебраїчних рівнянь

$$P(d, w) = \sum_{z \in Z} P(z)P(w|z)P(d|z),$$

де  $w \in W'$ ,  $n(d, w) > 0$ .

У якості невідомих розглядаємо величини  $P(d|z)$ ,  $z \in Z$ . Значення величин  $P(z)$ ,  $P(w|z)$ ,  $z \in Z$ ,  $w \in W'$  отримані в результаті застосування PLSA до множини документів  $D'$ . Величина  $P(d|w)$  апроксимується по формулі  $P(d, w) \approx \frac{1}{|D'|} \frac{n(d, w)}{\text{length}(d)}$ , де  $\text{length}(d)$  – кількість слів із  $W'$ , присутніх у документі  $d$ .

Для наближеного розв'язку даної системи ми будемо псевдообернену матрицю для матриці коефіцієнтів системи методом Гревілья. Отримане наближення є найкращим (за методом найменших квадратів) і може розглядатися як досить точна апроксимація образу документа  $d$  у просторі факторів  $Z$ .

Критерієм оцінки якості пошуку інформації є побудова усередненої за запитами кривої повноточності. Однак цей підхід неприйнятний у зв'язку із широким розкидом числа релевантних документів для різних запитів. У зв'язку із цим для порівняння результатів пошуку за вихідним й розширеним запитами використовуємо наступний підхід. Позначимо через  $R_q$  множину рангів (порядкових номерів) релевантних документів для запиту  $q$  серед перших 200 документів, отриманих у результаті пошуку. Нехай далі  $Q$  є множина запитів, а  $rel(q)$  – загальне число документів в  $D$ , релевантних запиту  $q$ . Нехай

$$quality = \frac{1}{|Q|} \sum_{q \in Q} \frac{1}{(1 + rel(q))} \sum_{i \in R_q} \frac{1}{i}.$$

Саме величина  $quality$  буде використовувати для оцінки ефективності пропонованого методу розширення запиту, обчислюючи її для вихідних і розширених запитів. Ця інтегральна оцінка враховує як ранги всієї сукупності релевантних документів, отриманих у відповідь на запит, так і загальне число релевантних документів у колекції. Ранги отриманих релевантних документів для запитів з невеликим загальним числом релевантних документів урахуються з більшими вагами, що відповідає інтересам користувача, що бажає бачи-

ти релевантні документи серед перших повернутих системою результатів, особливо у випадку їх мало-го числа [4].

#### 4. Розробка інформаційної технології підтримки консультування студентів

Усю сукупність представлених на сьогоднішній день методів аналізу тексту, щодо задачі аналізу його змісту, можна розділити на дві великі групи:

а) лінгвістичний аналіз, орієнтований на добування змісту тексту за його семантичною структурою;

б) статистичний аналіз, орієнтований на добування змісту тексту за частотним розподілом слів.

Авторами було ухвалене рішення використовувати методи статистичного аналізу в силу їх зручності використання і мовної незалежності. Методи лінгвістичного аналізу хоча й дозволяють точніше аналізувати текст, виділяючи його структурні особливості, але є більш складними у використанні. Пов'язане це, насамперед, з багатством семантики й морфології природних мов. Формальний опис правил природної мови і їх реалізація – досить трудомісткий процес, що вимагає залучення фахівців з області лінгвістики. Крім того, лінгвістичний аналіз припускає орієнтацію на конкретну мову з її конкретними семантичними особливостями, це обумовлює погану міжмовну переносимість. Роботи в даному напрямку йдуть, і існують множини практичних реалізацій, але на сьогоднішній день лінгвістичний аналіз у галузі аналізу семантики досить проблематичний.

Усе це обумовило доцільність застосування статистичних методів для розв'язку задач, що були поставлені в даній статті, тому в рамках даної роботи тематичний аналіз тексту припускає розв'язання задачі семантичного пошуку інформації з використанням статистичного аналізу в силу його мовної незалежності.

#### 5. Попередня обробка тексту

Однією з головних проблем аналізу текстів є велика кількість слів у документі. Якщо кожне із цих слів піддавати аналізу, то час пошуку нових знань різко зростає та навряд чи буде задовольняти вимогам користувачів. У той же час очевидно, що не всі слова в тексті несуть корисну інформацію. Крім того, у силу гнучкості природних мов формально різні слова насправді означають однакові поняття. Таким чином, видалення неінформативних слів, а також приведення близьких за змістом слів до єдиної форми значно скорочують час аналізу текстів. Усунення описаних проблем виконується на етапі попередньої обробки тексту.

Звичайно використовують наступні способи видалення неінформативних слів і підвищення строгості текстів:

– Видалення стоп-слів. Стоп-словами називаються слова, які є допоміжними та несуть мало інформації про зміст документа. Звичайно заздалегідь складаються списки таких слів, і в процесі попередньої обробки вони видаляються з тексту. Типовим прикладом таких слів є допоміжні слова й артиклі, наприклад: “тому що”, “крім того” тощо.

– Стеммінг – морфологічний пошук. Він полягає в приведенні кожного слова до його нормальної форми, виділення основи слова. Нормальна форма виключає відмінювання слова, множинної форми, особливості усної мови тощо. Наприклад, слова “стиск” і “стислий” повинні бути перетворені в нормальну форму слова “стискати”. Алгоритми морфологічного розбору враховують мовні особливості та внаслідок цього є мовно залежними алгоритмами.

– Використання TV-грам. TV-грами – це альтернатива морфологічному розбору та видаленню стоп-слів. TV-грама – це частина рядка, що полягає з TV символів. Наприклад, слово “дата” може бути презентовано 3-грамою “да”, “дат”, “ата”, “та” або 4-грамою “дат”, “дата”, “ата\_”, де символ підкреслення заміняє попередній або замикаючий слово пробіл. У порівнянні зі стеммінгом або видаленням стоп-слів, TV-грами менш чутливі до граматичних і типографічних помилок. Крім того, діаграми не вимагають лінгвістичного представлення слів, що робить даний спосіб більш незалежним від мови. Однак TV-грами, дозволяючи зробити текст більш строгим, не вирішують проблему зменшення кількості неінформативних слів.

– Приведення регістру. Цей спосіб полягає в перетворенні всіх символів до верхнього або нижнього регістру. Наприклад, усі слова “текст”, “Текст”, “ТЕКСТ” приводяться до нижнього регістру “текст”.

## 6. Архітектура системи

Система передбачає використання архітектури клієнт-сервер – обчислювальної або мережної архітектури, у якій завдання або мережне навантаження розподілене між постачальниками послуг (сервісів), що називаються серверами, і замовниками послуг, що називаються клієнтами. Нерідко клієнти та сервери взаємодіють через комп’ютерну мережу та можуть бути як різними фізичними обладнаннями, так і програмним забезпеченням.

У якості архітектури пошукової системи була обрана трьохрівнева архітектура клієнт-сервер, яка передбачає наявність клієнтського застосування (звичайно говорять «тонкий клієнт» або термінал), підключене до сервера застосувань, який у свою чергу підключений до сервера бази даних. Така архітектура забезпечує масштабованість, конфігурованість (завдяки ізольованості рівнів один від одного), високу безпеку та надійність, низькі вимоги

до продуктивності та до технічних характеристик терміналів.

На рис.1 представлена архітектура пошукової системи у вигляді діаграми взаємодії компонентів системи.

Опис представлених компонентів на діаграмі:

– SearchManager – двигок пошукової системи, що забезпечує взаємодію всіх компонентів системи;

– ProbabilisticLatentSemanticAnalyzer – компонент, який виконує імовірнісний латентно-семантичний аналіз тексту;

– StopWordsRemover – компонент, який відповідає за видалення словникових стоп-слів;

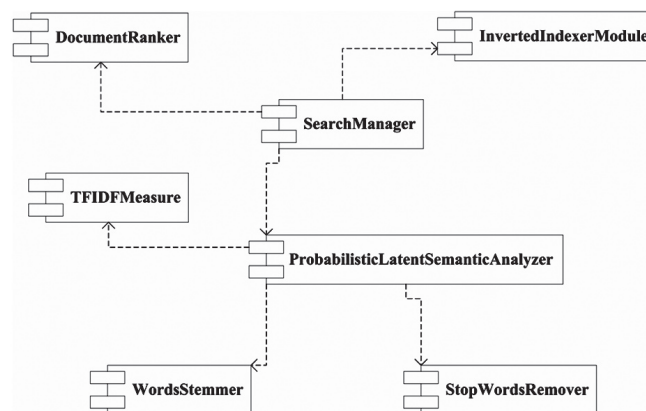


Рис. 1. Діаграма взаємодії компонентів системи

– WordsStemmer – компонент, який виконує стеммінг слів (знаходження основи слова), і використовується для знаходження найбільш інформативних слів;

– TFIDFMeasure – компонент, який відповідає за обчислення ступеня важливості слова в контексті документів по алгоритму TF-IDF;

– DocumentRanker – компонент, який виконує ранжирування пошукових образів документів, відповідно до кількості та важливості слів, що в ньому зустрічаються;

– InvertedIndexerModule – модуль, який виконує відкладене зворотне індексування документів, забезпечуючи тим самим формування бази існуючих відповідей.

Етапи інтелектуального аналізу текстової інформації з використанням імовірнісного латентно-семантичного аналізу представлені на рис. 2 у вигляді діаграми послідовності, яка зображує процес взаємодії 5 основних об’єктів пошукової системи:

- користувача системи (User);
- двигка пошукової системи (Search Manager);
- попереднього оброблювача слів (Words Processor);
- імовірнісного латентно-семантичного аналізатора тексту (Probabilistic Latent Semantic Analyzer);
- ранжиру документів (Document Ranker).

Інтелектуальний пошук текстової інформації можна представити у вигляді декількох етапів: пошук документів, їх попередня обробка, добування інформації, застосування методу імовірнісного латентно-семантичного аналізу, інтерпретація результатів.

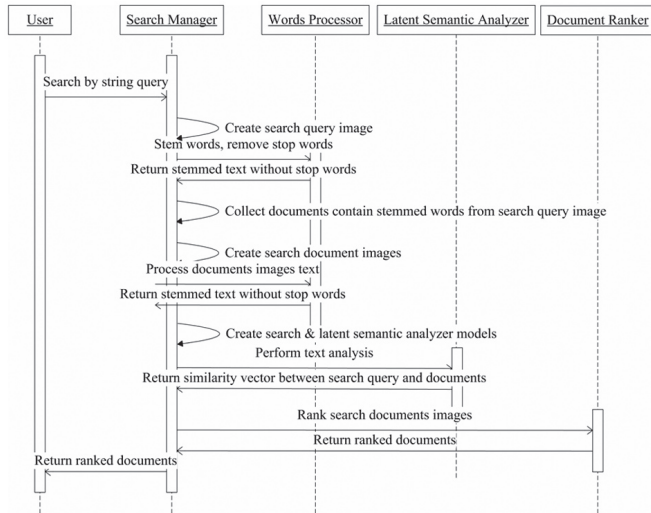


Рис. 2. Діаграма послідовності

### Висновки

Дистанційне навчання ще недостатньо впроваджене в систему освіти України, хоча, судячи з останніх тенденцій, має великі шанси міцно увійти в життя як студентів, так і фахівців старших вікових категорій, надаючи їм можливість реалізувати концепцію постійного підвищення кваліфікації протягом всього життя.

Об'єктом дослідження у рамках роботи був процес консультування студентів дистанційної форми навчання. Методом дослідження виступало поліпшення процесу консультування для зменшення об'єму роботи викладача та підвищення швидкості отримання студентом відповіді на поставлене запитання.

У ході аналізу предметної області було зроблено загальний огляд системи дистанційного навчання, дистанційних навчальних технологій, процесу консультування студентів та його можливої автоматизація.

В результаті було спроектовано та розроблено інформаційну технологію підтримки консультування студентів дистанційної форми навчання у вигляді веб-застосування з використанням технології ASP.NET, яка оптимізує процес консультування студентів, зменшуючи об'єм роботи викладача та підвищуючи швидкість отримання студентом відповіді на поставлене запитання.

Отриманий програмний продукт може бути використано у закладах, що забезпечують навчання

та перепідготовку з використанням дистанційної освіти. Крім того, він може бути використаний на фірмах, що хотіли б збільшити швидкість навчання своїх нових співробітників.

Розроблена інформаційна технологія може бути розширена за наступними напрямками:

- застосування морфологічного аналізу при пошуку текстової інформації;
- інтегрування перевірки орфографічних помилок у пошуковому запиті;
- реалізація надання пошуку мультимедійної інформації, наприклад, формул у вигляді рисунків;
- розгляд можливості використання онтологій;
- розширення локалізації засобів попередньої обробки даних.

**Список літератури:** 1. *Deerwester, S.* Indexing by Latent Semantic Analysis [Текст] / S. Deerwester, S. T. Dumais, G. W. Furnas and other // Journal of the American Society for Information Science. – 1990. – Т. 41, №6. – PP. 391-407. 2. Latent semantic indexing (LSI) and TREC-2 [Текст]: In Proc. of the Second Text Retrieval Conference / D. Harman. – 1994. – PP. 105-116. 3. Using latent semantic indexing for information filtering [Текст] : In ACM Conference on Office Information Systems (COIS) / P. W. Foltz. – 1990. – PP. 40-47. 4. *Hofmann, T.* Probabilistic latent semantic indexing [Текст] / T. Hoffman. – In Proc. of the SIGIR'99. – 1999. – PP 50-57

Надійшла до редколегії 15.09.2010.

УДК 519.863.5

**Информационная технология поддержки консультирования студентов дистанционной формы обучения** / Н.С. Лесная, А.В. Найденко, В.В. Сокол // Бионика интеллекта: науч.-техн. журнал. – 2010. – № 3 (74). – С. 138–143.

Статья посвящена решению актуальной задачи – повышению качества консультирования в системах дистанционного обучения. В процессе решения были проанализированы существующие подходы и определен наилучший – метод вероятностного латентно-семантического анализа. В результате было получено программное средство, которое позволяет ускорить процесс консультирования и повысить качество поиска в базе существующих ответов.

Ил. 2. Библиогр.: 4 назв.

UDC 519.863.5

**Information technology of support for counseling students of distance learning** / Lesna N.S., Naydenko A.V., Sokol V.V. // Bionics of Intelligence: Sci. Mag. – 2010. – № 3 (74). – P. 138–143.

The main goal of the article is solving the actual problem - improving the quality of counseling in the systems of distance learning. A method of probabilistic latent-semantic analysis was chosen to solve the problem. As a result of the software tool which allows improving the performance of counseling and decreasing the searching time of correct answer.

Fig. 2 Ref.: 4 items