УДК 681.3.016

ТЕЛЕНИК С.Ф.,
АМОНС О.А.,
ШКАБУРА О.Ю.,
ПОДРИГАЙЛО Н.О.

ПОШУК І РЕФЕРУВАННЯ В СИСТЕМІ ЕЛЕКТРОННОГО ДОКУМЕНТООБІГУ

Робота присвячена проблемі пошуку документів у масиві за атрибутами та на основі повнотекстового пошуку. Представлено модифікований метод рубрикації та метод реферування на основі рубрикації. Показано переваги використання цього підходу на прикладі системи електронного документообігу SmartBase.SEDO.

This work deals with the problem of document search in arrays by attributes and uses full-text search technology. Modification of rubrication method is presented and abstracting rubrication-based method is developed. The advantages of this conception usage is demonstrated on the electronic documents circulation system SmartBase.SEDO.

Вступ

Сьогодні обсяги інформації зростають швидкими темпами, причому левова частка надходжень забезпечується за рахунок неструктурованої інформації. Склалася ситуація, за якої фахівці більше часу витрачають на пошук інформації, ніж на її використання за призначенням. Щоб зарадити цій ситуації в системах автоматичного та автоматизованого оброблення великих масивів неструктурованої інформації широко використовуються моделі та алгоритми рубрикації та реферування інформації. Такі ж проблеми притаманні системам електронного документообігу, які поступово стають незамінними у міністерствах, відомствах та великих корпораціях. До того ж у цих системах час пошуку документів, які можуть бути корисними для прийняття рішень, часто суттєво обмежений, причому апріорі відомою може бути лише тема документу.

У статті розглядається проблема пошуку інформації у системах електронного документообігу. У цих системах необхідно підтримувати пошук документів за атрибутами, на основі ключових слів та повнотекстовий пошук. Для прискорення процесу повнотекстового пошуку інформації в системах електронного документообігу пропонуються модифікації методу опорних векторів для рубрикації та методу реферування документів на основі рубрикації.

Експериментальні результати, одержані за допомогою реалізованої на основі запропонованих методів пошукової системи у складі системи електронного документообігу SmartBase.SEDO, підтвердили їх ефективність.

Системи електронного документообігу, реферування і рубрикація

Документ – це матеріальний носій із зафіксованою на ньому інформацією. Таким носієм може бути папір, диск комп'ютера, фотоплівка і т.п. Діловий документ служить для фіксування адміністративної (управлінської) інформації. У нашому випадку це електронний документ (ЕД). Вважатимемо, що ЕД – це сукупність вмісту і службової інформацію щодо документа у вигляді електронної реєстраційно-контрольної картки документа (ЕРКК). Інформація ЕРКК документа представлена в системі окремим програмним класом і зберігається в відповідній таблиці бази даних. Для ефективної автоматизації функцій опрацювання ЕД необхідно закріпити за документом чітку формалізовану структуру у вигляді шаблону. Надалі шаблоном будемо називати об’єкт, який описує сталі частини інформації документів, створених за даним шаблоном, та поля змінної інформації з описанням методів і механізмів їх заповнення.
Об’єкт шаблон містить загальну інформацію, вміст (так звана статична інформація), перелік усіх елементів (тобто поля змінної інформації) та набір методів і механізмів, які заповнюють елементи документа. Уся наведена інформація представлена відповідними класами і зберігається в організаційній базі даних. Шаблон документа можна розглядати

як заздалегідь підготовлений макет документа. Використання шаблонів дозволяє заощаджувати час на формування та оброблення документів і досягти певного рівня уніфікації і стандартизації документів.

Рубрикація документів – це розповсюджена технологія впорядкування інформації щодо вмісту документа. Процес рубрикації полягає в описанні змісту документа шляхом використання елементів визначеного скінченого переліку тем (рубрикатора).

Тема – це сукупність ключових слів чи словосполучень з вказаними для них деякими параметрами. Вага теми – це числове значення, яке присвоюється кожній темі в залежності від таких критеріїв:

- кількість входжень слів чи словосполучень в текст;
- входження слів чи словосполучень в запит користувача;
- входження слів чи словосполучень в тему документа;
- входження слів чи словосполучень в поле «Короткий зміст документа».

Система електронного документообігу оперує n·1000000 документів на різні теми, різного обсягу та складності. Для ознайомлення с таким обсягом інформації необхідно витратити велику кількість часу. Ось чому проблемам автоматичного реферування та рубрикації приділяється так багато уваги.

Огляд існуючих рішень

За останні роки з'явилося багато публікацій, в яких розглядаються проблеми автоматичного реферування. Навіть саме поняття «Автоматичного реферування» стало більш об'ємним і тепер стосується не лише текстової інформації а й мультимедіа у цілому. Поряд з традиційними задачами створення рефератів для окремих документів з'явилося нове поле діяльності, пов'язане з багатомовним реферуванням багатьох документів. В оглядовій статті [1] наведений детальний огляд стану автоматичного реферування. На сайті постійно діючої конференції DUC [2] можна побачити останні публікації в області автоматичного реферування, порівняльні характеристики та результати тестування різних систем. Методи, які використовуються для реферування текстової інформації, можна поділити на два напрямки: квазіреферування та генерування рефератів. Перший напрям включає в себе методи, які базуються на виділенні з тексту найбільш інформативних частин (речень), що передають головний зміст тексту документа. Другий напрямок націлений на виділення найбільш інформативної частини тексту з наступним генеруванням на її основі нового тексту. Практично усі сучасні системи реферування відносяться до напрямку квазіреферування, хоча останнім часом з'явилися цікаві роботи, в яких розвиваються технології генерування рефератів [3].

Більшість існуючих методів обох підходів до реферування базується на запропонованій Г.Луном концепції [4]. Сутність зазначеної концепції полягає у послідовному виділенні у тексті слів, яким певним чином приписується вага, визначенні ваги речень шляхом сумування ваги слів, що входять до його складу, та включенні в реферат речень з найбільшою вагою. Для сучасних методів реферування характерною ознакою є використання традиційного підходу з деякими модифікаціями. Наприклад, в якості значущих елементів вибираються не слова, а словосполучення [5], вводяться додаткові критерії відбору значущих слів, наприклад вага слова збільшується в залежності від його знаходження в заголовку, в першому чи останньому реченні або у запиті користувача [6] тощо.

У праці [7] пропонується концепція фрактального реферування для подачі контенту на мобільні прилади, які звичайно характеризуються малими розмірами екрану та обчислювальною потужністю, низькою швидкістю передачі даних. Запропонований підхід використовує традиційні методи виділення речень, але додатково враховує інформацію про ієрархічну структуру документу та потрібний «рівень абстракції» представленого документу.

Методи симетричного реферування, описані у працях [8], враховують, насамперед, зв'язки між реченнями, причому важливішими вважаються речення, які містять багато зв'язків. Зв'язки між реченнями виявляються за допомогою готового словника термінів предметної області, що дещо обмежує застосування цих методів.

Підхід на основі різноманіття запропонований у праці [9]. Ідея підходу полягає у тому, щоб спочатку знайти тематичні кластери документа (тобто групи речень, які належать

до однієї підтеми документа). Після цього важливі речення виділяються з кожного кластера шляхом застосування традиційних методик. Кластеризація речень виконується за допомогою модифікованого методу *k* – середніх. Цей метод відносить кожне навчальне спостереження до одного з *k* кластерів (де *k* – число радіальних елементів) таким чином, щоб кожен кластер був представлений центроїдом відповідних спостережень, а кожне спостереження знаходилося на меншій відстані від центроїду свого кластера, ніж відстань до центроїда всіх інших кластерів.

Відомі також підходи до реферування на основі попередньо виконаної тематичної кластеризації документа з подальшим виділенням ключових речень з кожного кластера [9]. У відповідності з попереднім розбиттям документа на частини з урахуванням структури документа, на основі кластеризації здійснюється побудова реферату для кожної з частин і відбір найбільш важливих з них [10].

Важливою складовою задачі реферування є оцінювання якості одержаного реферату. Цій проблемі присвячено багато праць, наприклад [11,12]. Окрім традиційних методик, пов'язаних з експертними оцінками якості рефератів, останніми роками розвиваються автоматичні методи оцінювання. Наприклад, у праці [12] пропонується оцінювати якість рефератів за наявністю в них частотних словосполучень з оригіналу і близькістю розподілів частот появи цих словосполучень в документі та рефераті.

Недоліком існуючих методів реферування, побудованих на застосуванні рубрикації, є рознесення цих процесів у окремі процедури. Оскільки процеси рубрикації і реферування містять у собі спільні процедури оброблення текстів, наприклад, морфологічний чи концептуальний аналіз, можна суттєво пришвидшити процес реферування за рахунок його тісної інтеграції з процесом рубрикації. Було б доцільним зберігати теми разом з документом і використовувати у подальшому для визначення тем в документах, створених на основі цього ж шаблону, оскільки кожний шаблон значно звужує тематику документів створених на його основі.

Постановка проблеми

Нехай заданий масив з *n* документів $D = \{d_i,\ i = 1,.....n\}$. Необхідно розробити підсистему пошуку документів у масиві за атрибутами, на основі ключових слів та на основі повнотекстового пошуку. Для вдосконалення пошуку необхідно розробити систему автоматичного реферування. Повнотекстовий пошук має здійснюватися на основі рефератів. Реферування нових документів повинно виконуватися паралельно з рубрикацією на основі існуючого рубрикатора. Результатом цього процесу повинні бути реферати нових документів і оновлений рубрикатор.

Загальний опис підходу

Підхід, який пропонується, базується на ідеї тісної інтеграції процесів реферування та рубрикації. Процес реферування, відповідно і процес пошуку, суттєво пришвидшується за рахунок раціонального використання спільних процедур оброблення текстів зазначених процесів, наприклад, морфологічного чи концептуального аналізу. Так, тему пропонується зберігати разом з документом і використовувати у подальшому для визначення тем в документах, створених на основі цього ж шаблону. Саме застосування шаблонів, традиційне для систем електронного документообігу, підвищує ефективність пошуку, значно звужуючи тематику документів, створених на їх основі.

Для виконання рубрикації документів пропонується використовувати модифікацію методу опорних векторів (SVM).

Після виконання процедури рубрикації масиву документів *D* кожному документу приписується множина пар, першим елементом яких є номер теми, а другим – її вага в документі. Кожна з тем множини *T* визначається множиною ключових слів та словосполучень, які містяться у тексті документу і в описі теми, і частотами появи цих слів та словосполучень і описується відповідним масивом двійок, першим елементом яких є слово або словосполучення з документу, яке визначає тему, другим елементом – частота появи в документі першого елементу.

При створенні нового документа, у ньому визначаються теми, що дозволяє одночасно

будувати реферат документу і поповнювати рубрикатор.

Модифікація метода опорних векторів

Реферування виконується паралельно з рубрикацією. Для виконання рубрикації документів, як зазначено вище, будемо використовувати модифікацію методу SVM.

Модифікацію виконуємо з метою підвищення точності рубрикації, оскільки звичайний метод SVM виконує класифікацію лише до того моменту поки не отримає позитивного класу ознак. Модифікація полягає у тому, щоб відділяти одну рубрику від усіх інших. Тобто для класифікації по 5 рубрикам буде проводитися навчання відразу 5 рубрикаторів. Це дозволить нам відносити тему не до першої позитивної рубрики, а до рубрики з найбільшою релевантністю.

Алгоритм модифікованого методу рубрикації:

1. Аналізується новий документ;
2. Для кожного з уже існуючих кластерів, перебираються теми, які до нього належать. Для кожної теми будується вектор ознак.
3. Виконується класифікація векторів ознак. У випадку, якщо вектор відноситься до позитивного класу, нова тема відноситься до того кластеру, до якого належить тема, для якої був отриманий вектор ознак. Якщо тема не була віднесена до жодного з існуючий кластерів, то на основі цієї теми будується новий кластер.

сервиса VoIP.

Після виконання процедури рубрикації масиву з n документів $D = \{d_i\}$, $i = 1,.....n$, кожному документу d_i приписується множина пар $TW_i = \{(t_1,w_1),\ldots,(t_j,w_j),...,(t_m,w_m)\}$, де t_i, w_i, $j = 1,...,m$, – відповідно номер теми і її вага в документі. Нехай T – набір тем документа d_i, W – набір ваг тем в документі d_i (сумарних частот появи в тексті словосполучень, які містяться у описі теми з набору T).

У свою чергу, кожна з тем $t_j \square T$ визначається множиною ключових слів та словосполучень, які містяться у тексті документу d_i і в описі теми t_j і частот появи цих слів та словосполучень. Кожна тема t_j описується масивом двійок $PF = \{(p^{(1)}_{ij},f^{(1)}_{ij}),...,(p^{(k)}_{ij},f^{(k)}_{ij}),\ldots,(p^{(l_j)}_{ij},f^{(l_j)}_{ij})\}$, де $p^{(k)}_{ij}$ – слово або словосполучення з документу d_i, яке визначає тему t_j; $f^{(k)}_{ij}$ – частота появи в документі d_i слова чи словосполучення $p^{(k)}_{ij}$; l_j – кількість слів чи словосполучень, які описують тему t_j.

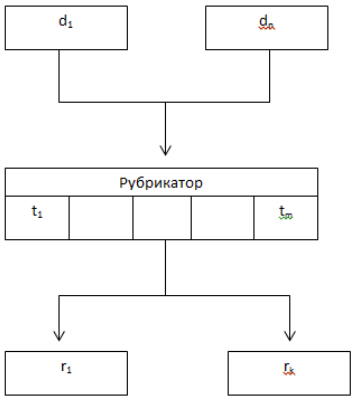

Рис. 1. Схематичне представлення роботи механізму

При створенні нового документа, у ньому визначаються теми. Якщо чергова визначена тема відсутня в рубрикаторі, то останній поповнюється, а тема з відповідною вагою приписується документу. Якщо ж чергова тема вже зафіксована у рубрикаторі, то останній не поповнюється, але тема з її вагою приписується документу.

Методи реферування на основі рубрикації

Вибір найбільш інформативних речень з тексту документа d_i пропонується здійснюється за допомогою такого модифікованого алгоритму:

Крок 1. Для кожної теми формується список речень, який її характеризує. Ця інформація береться з рубрикатора.

Крок 2. Усі речення списку оброблюються з метою вилучення стоп-слів з використанням словника стоп-слів, який містить службові частини мови, а також неінформативних слів та словосполучень.

Крок 3. Визначається вага кожного речення шляхом підсумовування частот появи у ньому слів та словосполучень, які визначають тему. Вага залежить від місця, де це речення вживається – на початку, кінці абзацу чи у ключових частинах, наприклад у висновках чи вступі.

Крок 4. У кожному з відібраних на кроці 1 речень видаляються примітки (слова у дужках) та деякі звороти за допомогою спеціального словника зворотів (різниця між словником зворотів та словником стоп-слів полягає у тому, що неінформативні слова

повністю задаються списком, а в словнику зворотів присутня лише початкова частина речення, наприклад “Як повідомляється…”). Після розпізнавання звороту з речення тексту вилучається не лише та частина, яка наведена в словнику, а й увесь зворот до наступного розділового знака.

Крок 5. Усі речення після оброблення на попередніх кроках (вилучення стоп-слів, приміток чи зворотами), які були кандидатами на включення до реферату, послідовно перевіряються на тотожність. Речення вважаються близькими, якщо вони мають 80% спільних слів. З усіх близьких речень в реферат включаються речення з найбільшою вагою.

Пошук

У системі електронного документообігу SmartBase.SEDO реалізовані такі типи пошуку:

1. атрибутивний пошук: пошук за різними комбінаціями відомих значень полів ЕРКК документу, який потрібно знайти, наприклад, дата створення документу, назва документу, відправник; автор документу тощо;
2. пошук по ключовим словам: дозволяє проводити пошук за відомими особі, яка здійснює пошук, ключовим словам у обсязі реферату;
3. повнотекстовий пошук: дозволяє проводити пошук за відомими особі, яка здійснює пошук, змістовними елементами документу (темами).

Метою цієї статі було розробити такий алгоритм пошуку, який би дозволяв здійснювати швидкий пошук, відштовхуючись від інформації, що вже міститься у системі. Саме ці особливості пошуку важливі для електронного документообігу. Оскільки повнотекстовий пошук є трудомістким, а тематика, за якою потрібно здійснювати пошук, може бути зовсім не освітленою в контенті конкретної системи, то звичайні методи пошуку вимагають не виправдано великих ресурсів і не забезпечують вимог систем електронного документообігу до оперативності. Пошукова підсистема для електронного документообігу, розроблена на основі попереднього реферування та рубрикації, дозволяє досить ефективно розв’язати ці проблеми.

Щоб перейти до реалізації запропонованого підходу розглянемо вимоги до функціональності пошукової підсистеми з боку користувача. Користувач, який здійснює пошук з використанням запропонованого підходу, повинен отримати у своє розпорядження такий функціонал:

- можливість вибору користувачем у рубракаторі тем, що його цікавлять;
- створення рефератів для документів, які попадають у систему електронного документообігу;
- відбір документів за вибраними користувачем темами і видача користувачеві рефератів усіх відібраних документів (у визначеному користувачем порядку);
- можливість надання користувачу повного тексту документа після ознайомлення з рефератом (на вимогу користувача).

Реалізація мехазму

Для програмної реалізації механізму пошуку документів у системі електронного документообігу на основі реферування та рубрикації документів розроблена узагальнена об’єктна модель, наведена на рис. 1. Вона не прив’язана до конкретного середовища розроблення або мови програмування. На основі цієї моделі створено відповідні таблиці в системі управління базами даних (СУБД) Oracle і програмні класи. Шаблони документів та самі документи зберігаються в базі даних (БД), що дозволяє працювати з ними територіально відокремленим групам користувачів.

Основу механізму роботи з шаблонами документів складають вісім класів, які забезпечують базову функціональність пошукової підсистеми: «Тип документа», «Організація», «Електронна реєстраційно-контрольна картка», «Шаблон документа», «Опис елемента», «Статичний вміст шаблону документа», «Елемент документа», «Джерело даних для заповнення елемента».

Клас «Тип документа» розрізняє всі документи і шаблони документів за призначенням, а клас «Організація» визначає організацію, до якої належить користувач. За допомогою цих двох класів визначаються шаблони документів, доступних користувачеві.

Клас «Шаблон документа» використовується для описання і оброблення шаблонів документів. Кожний шаблон має унікальний ідентифікатор, а також набір полів, наприклад тип, назва, додаткова інформація тощо.

Цей клас забезпечує збереження шаблону документа у БД і генерацію документа за вибраним шаблоном. Елементи шаблону, спільні для всіх документів, створюваних за цим шаблоном, зберігаються у БД. Усі дані шаблону, які будуть унікальними для конкретних документів, описуються відповідними графічно-програмними елементами. Ці елементи програмно закріплюються в шаблоні документа і зберігаються в його структурі за допомогою класу «Опис елемента». Кожний елемент шаблона пов'язаний з класом «Шаблон документа» за допомогою унікального ідентифікатора шаблона. У системі реалізовано такі типи елементів як текстове поле, дата, випадаючий список та ін. Тип елемента визначається класом «Опис елемента».

Кожний елемент шаблону може бути заповнений автоматично або залишений незаповненим. Заповнення елементів шаблону можна визначати за допомогою логічних правил. Наприклад, це дуже зручно для введення дати. Для реалізації цієї ідеї розроблений механізм автоматичного заповнення елементів за допомогою класу «Джерело даних для заповнення елемента». Автоматично перевіряються типи елементів та функцій.

Окрім готових функцій можна використовувати формули заповнення. Логічний апарат оброблення формул, виконання необхідних розрахунків і заповнення елементів результатами обчислень описано в окремому класі «Менеджер формул». Формули корисні у випадку необхідності виконати певні математичні операції над визначеними даними в документі. Типовим прикладом може бути сума усіх значень визначених полів таблиці.

Клас «Статичний вміст шаблона документа» відповідає за збереження шаблону документа в БД. Вся статична структура шаблону документа створюється в форматі MS Word, архівується й упаковується в байтовий формат. В упакованому вигляді весь шаблон зберігається в одному полі БД.

Після створення шаблону документа, на його базі можна створювати електронні документи за допомогою класу «Електронна реєстраційно-контрольна картка». Користувачу необхідно лише вибрати відповідний шаблон і визначити потрібні опції, наприклад електронний цифровий підпис, затвердження, тощо. Елементи електронного документа описуються класом «Елемент документа», який також відслідковує всі зміни елементів конкретного документа після його редагування і відповідає за заповнення елементів документа.

Для реалізації механізмів реферування та рубрикації в систему вводяться додаткові класи «Анотація» та «Кластер». Перший з них забезпечує аналіз змісту документа та його збереження у БД разом з самим документом. Другий клас забезпечує збереження всіх тем документів, створених на основі визначеного шаблону. До кожної теми також зберігаються слова та їх ваги.

Розглянемо роботу механізму більш детально. У процесі створення електронного документа спочатку формуються загальні відомості про документ, насамперед тип, назва й опис. На основі аналізу ряду показників, насамперед організації користувача і типу документа, формується відповідний список шаблонів документа. Після вибору шаблону документа користувачем він завантажується з БД, при цьому всі елементи, що мають активний метод заповнення, заповнюються відповідними даними. Генерований таким чином електронний документ зберігається у БД, над ним виконуються дії відповідно до визначених користувачем опцій. Варто зазначити, що користувач має можливість у будь-якому документі відповідно до своїх прав створювати нові елементи або вилучати непотрібні елементи, змінювати методи їх заповнення. Створення електронного документу завершується його автоматичним анотуванням (із застосуванням рубрикації) і реферуванням.

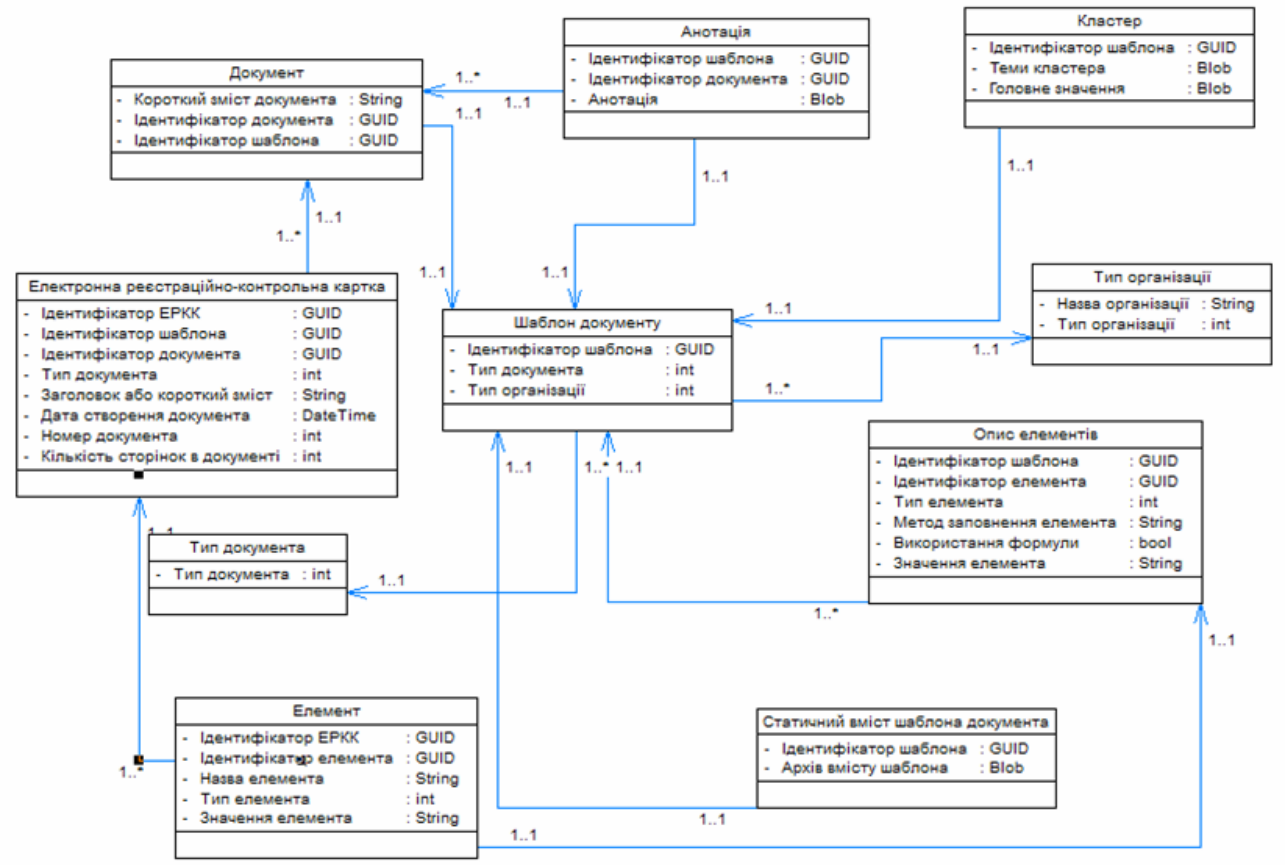

Рис.2. Об'єктна модель механізму анотування та реферування

Результати експеременту

Для визначення впливу механізму реферування на ефективність пошуку в системі електронного документоообігу були проведені його експериментальні дослідження. Вони дозволили одержати цікаві результати. Визначимо вихідні та результуючі параметри дослідження. Вихідними параметрами будуть основні компоненти електронного документообігу – множина документів і структура шаблону. В якості результуючих характеристик будемо відслідковувати час, витрачений на пошук з використанням реферування, та час реферування.

У цьому дослідженні використовувався пошук документів у системі електронного документообігу з використанням механізму попереднього реферування і без використання попереднього реферування. Результати наведені на рис. 3.

Експериментальні дослідження виконувалися на тестових документах, створених для тестування механізмів і підсистем системи електронного документообігу.

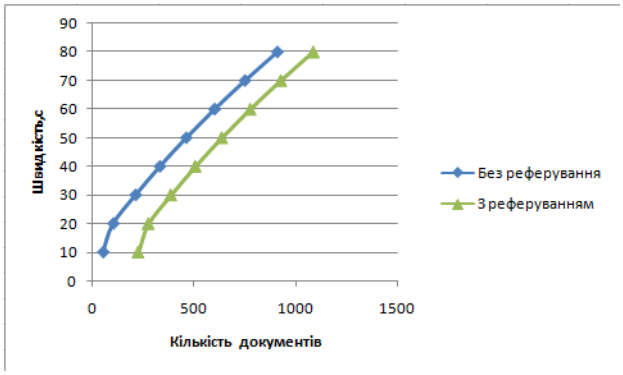

Рис. 3. Діаграма дослідження швидкості при пошуку документів

Наведені на рисунку результати дослідження дозволяють зробити висновок, що пошук з використанням реферування працює швидше – система за той самий час спроможна обробити більшу кількість документів. І чим більша кількість документів, що обробляються, тим більше часу при цьому економиться. Про це свідчить діаграма залежності часу реферування від кількості документів, наведена на рис. 4.

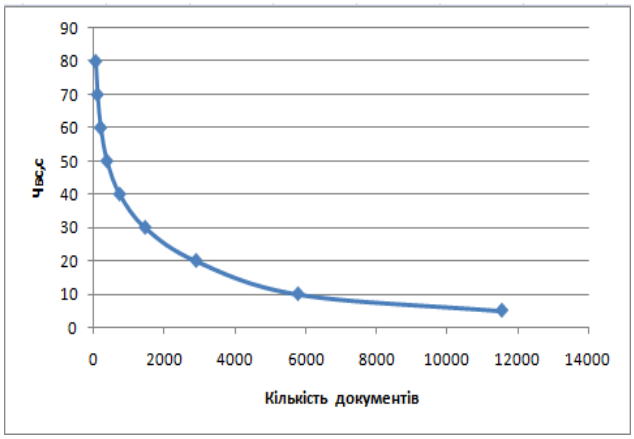

Рис. 4. Часова діаграма роботи механізму реферування

Це ще одне підтвердження ефективності використання механізму реферування. Досвід впровадження SmartBase.SEDO у системах електронного документообігу великих міністерств і відомств, де накопичуються сотні тисяч різноманітних документів, продемонстрував високу ефективність пошуку на основі реферування. Чим більше документів вже пройшли процес реферування, тим менше часу йтиме на реферування інших типових документів.

Висновки

Запропонований у статті метод пошуку може бути з успіхом застосований у сучасних інформаційних та інформаційно-аналітичних системах для оброблення великих масивів документів. Попереднє автоматичне реферування документів зменшує витрати у перерахунку на один документ і дозволяє значно скоротити час пошуку документів.

Список посилань

1. Hahn U., Mani I. The Challenges of Automatic Summarization // Computer.- 2000.- vol.33.- №11.- P. 29–36.
2. Document Understanding Conferences (DUC) Web site, 2008. http://duc.nist.gov.
3. Алыгулиев Р.М. Автоматическое реферирование документов с извлечением информативных предложений // Вычислительные технологии. – 2007. – Т. 12, № 5. – С. 5–15.
4. Luhn H. The automatic creation of literature abstracts // IBM Journal of Research and Development.- Vol. 2(2).- 1958.- P. 159–165.
5. Белоногов Г.Г., Калинин Ю.П., Хорошилов А.А. Компьютерная лингвистика и перспективnse информационные технологии. – М.: Русский мир, 2004. – 246 с.
6. Браславский П.И., Колычев И.С. Автоматическое реферирование веб-документов с учетом запроса // Интернет-математика-2005. – М.: Яндекс, 2005. – С. 485–501.
7. Yang, Ch. C. , Wang, F. L. Fractal Summarization for Mobile Devices to Access Large Documents on the Web // Proc. of the WWW2003, May 20-24, 2003, Budapest, Hungary. P. 26–31 .
8. Ступин В. С. Система автоматического реферирования методом симметричного реферирования // Компьютерная лингвистика и интеллектуальные технологии. Труды международной конференции «Диалог’2004». (Верхневолжский, 2 – 7 июня 2004 г.). – М.: Наука, 2004. – С. 579-591.
9. Nomoto T., Matsumoto Y. The diversity-based approach to open-domain text summarization // Information Processing & Management.- 2003.- №39.- P. 363–389.
10. Губин М.В., Меркулов А.И. Эффективный алгоритм формирования контекстно-зависимых аннотаций // Компьютерная лингвистика и интеллектуальные технологии : Труды международной конференции «Диалог’2005» (Звенигород, 1 – 6 июня 2005 г.). – М. : Наука, 2005. – С. 116–120.
11. Harman D., Over P. The effects of human variation in DUC summarization evaluation // Text summarization branches out workshop at ACL’2004.
12. Tait J. Making Better Summary Evaluations // Proc. of the Internatinal Workshop “Crossing Barriers in Text Summarisation Research” / Horacio Saggion and Jean-Luc Minnel (eds.).- Borovets, Bulgaria, Septemeber 2005, P. 26–31.

Поступила в редакцию 16.12.2009