

Модуль формування тестових запитань характеризується простотою додавання тестів та високою швидкістю обробки даних.

Розроблений модуль генерації тестових запитань у порівнянні з існуючими аналогами, наведеними у таблиці 3 характеризується вищою швидкістю генерації запитань та не поступається за функціональністю.

Список використаних джерел

1. Федорова Я.Б. Управління іт вищих навчальних закладів: як інформаційні технології допомагають зробити управління ефективним / Я.Б. Федорова – Херсон: Айлант, - 2006. -356 с.
2. Морган К. Методы синтаксического анализа объемных или сложных XML-документов [Электронный ресурс] – Режим доступа: <http://www.ibm.com/developerworks/ru/library/x-xmlphp2/>
3. Анисимов А.М. Работа в системе дистанционного навчання / А.М. Анисимов – Харків: ХНАГХ, - 2009.- 292с.

УДК 658.012

ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ СИСТЕМ ГРАФЕМАТИЧНОГО АНАЛІЗУ

Бобас О.І., Городиський Н.В.

Тернопільський національний економічний університет, магістранти

І. Постановка проблеми

В даний час дуже поширені інформаційні системи і бази даних, що містять в собі величезну кількість текстової інформації. Великі освітні центри організують в Інтернеті для студентів і співробітників бази наукових статей, авторефератів, багато організацій надають доступ до ресурсів електронних бібліотек, оргкомітети конференцій публікують тисячі повних текстів доповідей і т. п. Кількість електронної інформації зростає настільки, що людині просто не під силу проаналізувати її самостійно, хоча необхідність проведення такого аналізу цілком очевидна, адже в цих даних укладені знання, які можуть бути використані при прийнятті рішень. Для автоматичної обробки таких даних широко використовуються методи комп'ютерної лінгвістики.

Комп'ютерна лінгвістика (КЛ) – один з напрямів прикладної лінгвістики, який орієнтований на використання різних комп'ютерних інструментів (програми, комп'ютерні технології обробки даних та їх організації) для моделювання функцій мови в тих чи інших умовах, ситуаціях, сферах і т.д., а також вся сфера застосування комп'ютерних моделей мови в лінгвістиці та суміжних дисциплінах. Комп'ютерну лінгвістику так само часто називають автоматичною обробкою тексту або машинною лінгвістикою [1].

Комп'ютерна лінгвістика вирішує різні задачі автоматичної обробки текстів на природних мовах такі як:

- машинний переклад;
- розпізнавання мови;
- інформаційний пошук;
- автоматична класифікація і реферування документів;
- автоматична лінгвістична обробка і складання машинних словників;
- побудова лінгвістичних процесорів, що забезпечують спілкування користувачів з автоматизованими інтелектуальними системами;
- витяг фактографічної інформації з неформалізованих текстів [2].

У наші дні складно уявити існування таких пошукових гігантів як Google і Яндекс без використання методів комп'ютерної лінгвістики. Великі агентства ЗМІ, що мають свої Інтернет-портали, широко використовують комп'ютерну лінгвістику для вирішення великої кількості завдань, наприклад, для пошуку схожих статей.

Практично всі продукти автоматичної обробки тексту володіють функціональністю первинного аналізу тексту – графематичного. У деяких програмних продуктах дана функціональність зводиться до розбиття тексту на слова, інші в доповнення до цього здійснюють пошук текстових конструкцій за певними шаблонами, розбиття тексту на речення і т.д.

II. Мета роботи

У даній роботі будуть детально розглянуті завдання графематичного аналізу тексту, запропоновані методи їх вирішення, а також спроектовано та реалізовано ПЗ, що здійснює вирішення поставлених завдань.

III. Опис функціоналу системи графематичного аналізу

Як зазначалось раніше, графематичний аналіз – це початковий етап автоматичної обробки текстів природною мовою. На вхід для графематичного аналізу подається текстовий файл.

Модуль графематичного аналізу повинен виконувати такі завдання:

- розбиття вихідного тексту на слова, роздільники і т.д.;
- виділення в тексті скорочень і абrevіатур;
- пошук в тексті ПІБ в тих випадках, коли ім'я та по батькові представлені у вигляді ініціалів;
- виділення електронних адрес, URL – адрес і т.д.;
- визначення в початковому тексті меж речень.

Вихідні дані повинні бути представлені у форматі, придатному для передачі на наступний етап – морфологічний аналіз, і містити таку інформацію:

- список виділених у тексті речень;
- список знайдених графематичних дескрипторів.

У списку речень повинні міститися:

- порядковий номер речення в тексті;
- позиція його початку в тексті;
- його довжина (кількість символів);
- текст речення.

Список графематичних дескрипторів повинен містити в собі такі дані:

- тип дескриптора;
- позиція почала дескриптора в тексті;
- його довжина (кількість символів);
- текст дескриптора;
- інформацію про те, чи може даний дескриптор стояти в кінці речення;
- семантичну інформацію про дескрипторі (для скорочень і абrevіатур).

Для вирішення поставлених завдань необхідно провести аналіз тексту в декілька етапів:

- 1) пошук в тексті скорочень і абrevіатур;
- 2) пошук текстових конструкцій з використанням шаблонів;
- 3) виділення меж речень і прямої мови, остаточно розмітка вхідного тексту.

IV. Особливості організації даних системи

У процесі реалізації ПЗ для пошуку в тексті скорочень і абrevіатур був використаний словник, для роботи з яким необхідно було вирішити два завдання:

- мінімізувати час завантаження словника при старті програми;
- забезпечити мінімальний час пошуку за словником.

З метою скорочення часу, потрібного на завантаження словника при старті програми, було прийнято рішення реалізувати механізми експорту та імпорту словника в бінарний файл, що містить в собі серіалізовані структури програми, в яких зберігається словник. Завантаження словника з такого файлу займає значно менше часу.

Другий і найбільш важливим завданням є організація зберігання словника і пошуку в ньому елементів з мінімальними тимчасовими витратами. Через порівняно невеликого обсягу словника (20000 записів) питання про мінімізацію обсягу використовуваної для зберігання словника оперативної пам'яті відходить на другий план. Так як необхідно побудувати словник тільки один раз (після цього він серіалізується і зберігається в бінарному файлі), то швидкість вставки в нього елементів так само не має значення. Операція видалення елементів застосовуватися не буде.

В якості структури для зберігання та подальшого пошуку були розглянуті такі варіанти:

- префіксне дерево (*trie*, ліс, навантажене дерево). Складність пошуку в префіксному дереві становить $O(2N)$, де N – це довжина рядка, в якому ведеться пошук.

– кінцевий автомат, алгоритм пошуку – Ахо-Корасік. Складність такого пошуку складає $O(N+M+Z)$, де N – довжина рядка, в якому ведеться пошук, M – сумарна довжина всіх ключових слів, Z – кількість появ шаблонів.

Висновки

В роботі було спроектовано та реалізовано (мова C++) систему графематичного аналізу. Для експортування словника в бінарний і текстовий архіви використовувалися такі бібліотеки проекту Boost: «boost_serialization» і «boost_archive». Для пошуку текстових конструкцій з використанням шаблонів була застосована бібліотека «boost_regex», що надає функціональність пошуку за регулярними виразами. В якості СУБД для експортування результатів графематичного аналізу була використана MySQL.

Список використаних джерел

1. Компьютерная лингвистика | Энциклопедия Кругосвет. [Электронный ресурс] URL: <http://files.school-collection.edu.ru/dlrstore/0032e870-dadd-5ab2-a83e-85e032c459b2/1009220A.htm>
2. Задачи компьютерной лингвистики. [Электронный ресурс] URL: <http://www.kompling.narod.ru/index1.html>.

УДК 658.012

ВИМОГИ ДО ПРОЕКТУ СТВОРЕННЯ ВЕБ-САЙТУ БУДІВЕЛЬНОЇ ФІРМИ

Васильків Н.М.¹⁾, Седляр М.О.²⁾

Тернопільський національний економічний університет

¹⁾ к.т.н., ²⁾ магістрант

I. Постановка проблеми

Створення сайту є необхідним фактором ефективної діяльності будівельної фірми. При цьому необхідно поєднувати вимоги, які ставляться до сайту представниками фірми, та пропозиції від розробників. Для успішної реалізації проекту створення веб-сайту будівельної фірми необхідно використати і сучасні інформаційні технології, і методологію управління проектами.

II. Мета роботи

Метою дослідження є формування основних вимог до сайту будівельної фірми, що дасть змогу правильно сформулювати команду проекту та здійснити розподіл основних робіт згідно структури сайту з врахуванням часових ресурсів.

III. Основні вимоги до сайту

Сайт повинен бути не тільки своєрідною візитівкою будівельної фірми, він має забезпечувати супровід бізнесу, бути професійно виконаним, з якісним управлінням, легким та швидким в користуванні. Для задоволення цих вимог при створенні сайту необхідно ретельно вивчити та проаналізувати ринок будівельних фірм та послуг суміжних організацій (наприклад, з виробництва і збуту будівельних матеріалів чи здійснення ремонтних будівельних робіт), представлений та доступний користувачеві інтернет-ресурсів. Крім того, необхідно проаналізувати існуючі сайти будівельних фірм (організацій) різних форм власності та обсягів виконуваних робіт, вирізнити їх переваги та недоліки не тільки з професійної точки зору, а й з позиції звичайного споживача.

Варто зазначити, що особливістю такого веб-сайту має бути багатофункціональність, яка створюється за рахунок різних модулів, як от, наприклад: інформація про фірму, фотогалерея зданих об'єктів, пропозиції щодо планування житлових споруд, використовуваних будівельних матеріалів та технологій, гостьова книга, форма зворотного зв'язку, пошук по сайту, форма замовлень, форма поточного стану будівництва і багато інших.

Тому при створенні сайту будівельної фірми необхідно врахувати наступне:

- наявність споруджуваних будівельною фірмою чи вже зведених об'єктів, що в свою чергу змушує розробника винести їх в окремий розділ;
- плани будинків, поверхів, квартир мають бути доступні користувачеві для ознайомлення;