

Реферат

Актуальність

Зростання кількості інформаційних джерел збільшується з кожним днем, кількість накопичених даних все зростає. У лавині інформації людині працювати не просто складно, а практично неможливо (людина не встигає відреагувати на отримувану інформацію-дані). Так читаючи новини людина, відсіває цікаві для себе, а інші пропускає навіть не глянувши на текст, а орієнтуючись лише на рейтинг новини або її назву. Викладач не може прочитувати повністю роботу кожного свого студента. Стає неможливим орієнтуватися в масі книг, що випускаються.

Мета і завдання дослідження

Метою роботи є розгляд такого аспекту роботи інформаційно- пошукових систем на заданій множині сайтів, як технології збору інформації, що міститься на Веб-ресурсах.

Завдання, що вирішуються в роботі

Основними завданнями програмного забезпечення, що, призначеного для оцінки параметрів веб-сторінок, є:

1. огляд організаційної структури сайтів локальної мережі НТУУ «КПІ»;
2. розробка автоматизованого модуля збору інформації;
3. огляд існуючих моделей пошуку інформації;
4. огляд організації семантичного вебу;
5. огляд існуючих рішень для повнотекстового пошуку інформації;
6. створення прототипу системи пошуку на заданій множині сайтів, що включає всі етапи;
7. досягнення високої ефективності результатів.

Досягнуті результати

Результатом роботи є веб-орієнтований додаток, призначений для контролю інформації, розташованої на сторінках веб-сайтів порталів НТУУ «КПІ» та EGEE, з метою підвищення рейтингу НТУУ «КПІ» у світовому рейтингу університетів Webometrics.

Наукова новизна отриманих результатів

Наукова новизна роботи полягає в дослідженні можливості застосування веб-орієнтованих додатків для здійснення пошуку інформація відображена на веб-сайтах.

Практичне значення отриманих результатів

Практичним застосуванням розробленого додатку є можливість його використання для пошуку інформації, представленої на веб-сайтах НТУУ «КПІ», таким чином спрощуючи процес пошуку потрібної інформації.

Висновки

У даній магістерській роботі були досліджені різні системи пошуку інформації на множині заданих сайтів. Для реалізації пошуку даних у веб-просторі НТУУ «КПІ» був розроблений і реалізований модуль пошуку інформації для комплексної системи, створеної для контролю інформації, представленої на веб-сайтах.

У ході роботи були вирішені наступні завдання:

1. Вибір механізму індексації та пошуку, який задовольняє співвідношенню ефективності та швидкості пошуку і обчислювальних потужностей, які для цього необхідні.
2. Здійснення повнотекстового пошуку по інформації, зібраної з заданих веб-сайтів.
3. Досягнення високої швидкості обробки інформації.
4. Побудова інтерфейсу користувача для пошуку.

В результаті експериментів було показано, що якість системи пошуку є достатнім для використання результатів її роботи при вирішенні подальших прикладних задач.

Робота на 81 сторінках містить 4 таблиці, 3 ілюстрації. При підготовці роботи використовувалася література з 21 різних джерел.

Ключові слова: пошук, робот, павук, веб-сайт, веб-портал, індекс, моделі.