

Засоби багатокритеріальної оцінки якості веб-сайтів та пошук підробних сайтів

Виконав: студент групи ДА-52м,
Рубльов Дмитро Сергійович

Актуальність розглянутої проблеми

Оцінка веб-сайтів є дуже актуальною, оскільки при недостатній якості, вони не можуть в повній мірі виконувати представницькі, інформативні та інші свої функції.

Мета роботи

Метою цієї роботи являється дослідження можливостей оцінювання веб-сайтів та пошук шляхів для класифікації підробних сайтів, та розробка програмного забезпечення для пошуку плагіату.

Параметри, що надаються пошуковою системою

- Кількість сторінок у індексі
- Відсоток сторінок з індексу, що приймають участь у пошуці
- Цитованість, або кількість зворотніх посилань

Параметри, що визначаються додатково

- Унікальність, новизна інформації на сторінках сайту
- Відповідність сторінок сайту стандартам Консорціуму Всесвітньої Мережі (W3C)
- Оптимізація вмісту сторінок, для пришвидшення їх завантаження

Зв'язок між якістю сайтів, та підробними сайтами

Присутність підробних сайтів, які “позичають” інформацію з інших Інтернет-ресурсів може стати причиною заниження позицій сайтів, з яких ця інформація була запозичена. Це можливо із-за того, що індексація сторінок різних сайтів може відбуватись у різні періоди часу, і у індекс першими можуть потрапити матеріали з підробного сайту, тим самим завдавши шкоди сайту, що був першоджерелом, занижуючи його «корисність» для пошукової системи.

Загроза, якою являються підробні сайти

Для сайтів будь-яких серйозних організацій, наявність підробок, що або повністю дублюють зовнішній вигляд та вміст, або запозичують частину вмісту, може стати сильним ударом по іміджу, та довірі користувачів та погіршити ранжування у пошукових системах, якщо вчасно не буде прийнято заходів щодо вилучення їх підробок зі всесвітньої мережі.

Категорії підробних сайтів

У даній роботі розглянуто три основні категорії підробних сайтів:

- Сайти, що підробляють зовнішній вигляд інших (фішингові сайти)
- Сайти, що видають чуже за своє (плагіат)
- Сайти, що створені для наповнення пошукової видачі нерелевантними результатами (дорвеї)

Практична частина дипломної роботи

Оскільки плагіат є однією з найактуальніших проблем, пов'язаних з підробними сайтами, було розроблене програмне забезпечення для пошуку плагіату на сторінках Інтернет-сайтів. Розроблене ПЗ може використовуватися як для пошуку плагіату у глобальній мережі, так і для перевірки, чи були взагалі проіндексовані сторінки з текстом, що перевіряється.

Етапи роботи системи пошуку плагіату

- Розбиття введеного користувачем тексту на послідовності слів
- Пошук Інтернет-сторінок, що містять подібні фрагменти тексту
- Аналіз знайдених Інтернет-сторінок та підрахунок відсотку входження в сторінку тексту, що перевіряється

Особливості роботи системи

- Відсутність у результатах перевірки сторінки, з якої брався текст для пошуку, свідчить про відсутність її у індексі пошукової системи
- Система не може знаходити дублікати текстів, які є закритими від індексування пошуковими системами, та знаходяться в файлах таких форматів, вилучення тексту з яких дуже ускладнене.

В заключення можна відмітити, що дана робота має практичне значення для підвищення якості веб-сайтів та боротьби з негативним впливом від дублювання іншими сайтами вмісту сторінок як для сайтів НТУУ "КПІ", так і для інших установ та організацій.

Дякую за увагу