

УДК 004.896

Гураль І. – ст. гр. СНМ-52

*Тернопільський національний технічний університет імені Івана Пулюя*

## **ІНФОРМАЦІЙНИЙ ПОШУК**

Науковий керівник: к.т.н., доцент Козак Р.О.

Hural I.

*Ternopil Ivan Pul'uj National Technical University*

## **INFORMATION SEARCH**

Supervisor: Kozak R.

Ключові слова: пошукові системи, ранжування, інформація

Keywords: search engines, ranking, information

Пошукові системи можна порівняти з довідковою службою, агенти якої відвідують підприємства, збираючи інформацію в базу даних. При зверненні в службу інформація видається саме з цієї бази. Оскільки дані з часом старіють, агенти їх оновлюють. Деякі підприємства самі присилають інформацію про себе, і тому агентам немає потреби приходити до них. Довідкова служба має дві функції: створення і постійне оновлення даних в базі і пошук інформації в базі за запитом клієнта.

Аналогічно, пошукова система складається з двох частин: так званого робота (павука), який обходить сервери Мережі і формує базу даних пошукового механізму.

База робота в основному формується самим роботом (він сам знаходить посилання на нові ресурси) і меншою мірою – власниками ресурсів, які реєструють свої сайти в пошуковій машині. Крім робота (мережевого агента, павука, черв'яка), який формує базу даних існує і програма, яка визначає рейтинг знайдених посилань.

Принцип роботи пошукової системи зводиться до організації пошуку у власному внутрішньому каталозі (базі даних) за ключовими словами, які вказуються користувачем у полі запиту, і повертає список посилань, рангований за релевантністю.

Релевантність – відповідність знайденого в процесі пошуку документа до зробленого запиту. При цьому пошукові системи аналізують розташування знайдених ключових слів в документі (заголовку або тексті), число повторень ключових слів, їх взаємне розташування в документі.

Найбільш точно відібрані документи розташовуються на початку списку документів, що знайдені в процесі пошуку. Для ранжування знайдених документів використовують наступні показники:

- розташування ключового слова на сторінці (keyword prominence) – показник, що визначає, як близько від початку документа знаходиться задане ключове слово. Як правило, чим ближче до початку сторінки розташовується ключове слово, тим точніше документ відповідає запиту;
- частота ключового слова (keyword frequency) – показник, що враховує абсолютну частоту використання ключових слів (скільки разів зустрічається дане ключове слово на сторінці). Найбільшу «вагу» при ранжуванні документів мають слова, що розташовані в заголовку сторінки (так званий титул, він міститься між тегамі title). Саме цей заголовок відображається в рядку заголовка браузера при перегляді сторінки, і слова, що там вказані мають найбільшу цінність (вагу) для пошукової системи.

- деякі пошукові системи при ранжуванні враховують індекс цитування – кількість посилань в Інтернеті, що вказують на конкретний сайт. Чим їх є більше, тим вищим є індекс цитування.

Результати пошуку за допомогою різних пошукових систем будуть відрізнятися. Це пояснюється тим, що в кожній пошуковій системі застосовуються різні алгоритми індексації, різна періодичність оновлення інформації в базі даних, охоплюється різний простір навколишніх серверів, і відповідно індексується різне число документів.

Варто відмітити, що, опрацьовуючи певний запит, пошукова система оперує саме внутрішніми ресурсами (а не блукає по мережі, як часто вважають користувачі), а внутрішні ресурси обмежені. Незважаючи на постійне оновлення бази даних пошукової системи, пошукова машина не може проіндексувати всі веб-документи: їх кількість надто велика. Тому завжди існує ймовірність, що шуканий ресурс просто невідомий конкретній пошуковій системі.

Проблема неповного пошуку полягає не тільки в обмеженості внутрішніх ресурсів пошуковика, а і в тому, що швидкість робота обмежена, а кількість нових веб-документів постійно зростає. Збільшення внутрішніх ресурсів пошукової машини не може повністю вирішити проблему, оскільки швидкість обходу ресурсів роботом обмежена.

При цьому неправильним є думка, що пошукова система містить копію вихідних ресурсів Інтернету. Повна інформація (вихідні документи) зберігаються далеко не завжди, найчастіше зберігається лише їх частина – так званий індексований список, або індекс, який набагато компактніший тексту документів і дозволяє швидше відповідати на пошукові запити.

Для побудови індексу вихідні дані перетворюються таким чином, щоб об'єм бази був мінімальним, а пошук здійснювався дуже швидко і надавав максимум корисної інформації. Умовним паперовим аналогом індексованого списку може бути конкорданс, тобто словник, в якому в алфавітному порядку перелічені слова, які вживає певний автор, а також подані на них посилання і частота їх вживання. Очевидно, що конкорданс (словник) суттєво компактніший самого тексту і знайти необхідне слово у ньому набагато простіше, ніж переглядати всю книгу з надією знайти відповідне слово.

УДК 004.891; 004.942

Пасікова Т.- ст.гр. СНм-52

*Тернопільський національний технічний університет імені Івана Пулюя*

## **ПАРАЛЕЛЬНІ ОБЧИСЛЕННЯ**

Науковий керівник: ас. Маєвський О.В.

Pasikova T.

*Ternopil Ivan Pul'uj National Technical University*

## **PARALLEL COMPUTING**

Supervisor: Majeviskiy A.

Ключові слова: графічний процесор, обчислення.

Keywords: GPU, computing.

Паралельні обчислення – форма обчислень, в якій багато розрахунків виконуються одночасно, вона виконується за принципом: великі проблеми часто