

УДК 004.78

Комінко І. – ст. гр. СПмс – 61

*Тернопільський національний технічний університет імені Івана Пулюя*

## **АЛГОРИТМ РАНЖУВАННЯ ВЕБ-СТОРИНОК ПОШУКОВИМИ СИСТЕМАМИ**

Науковий керівник: к.т.н. доц. Литвиненко Я.В.

Kominko I.

*Ternopil Ivan Pul'uj National Technical University*

## **ALGORITHM FOR RANKING WEB PAGES SEARCH ENGINE**

Supervisor: Lytvynenko Y.V.

Ключові слова: АЛГОРИТМ, РАНЖУВАННЯ, ПОШУК

Keywords: ALGORITHM, RANKING, SEARCH

На даний час облік тематики посилань використовується майже всіма пошуковими системами. Використовуючи певний алгоритм ранжування відбираються  $N$  сторінок, що відповідають пошуковому запиту. Новий алгоритм ранжування буде працювати тільки з цими  $N$  сторінок. Кожна сторінка в цій групі має деякий ранг OldScore. При розрахунку LocalScore для даної сторінки виділяються всі сторінки з  $N$ , які мають зовнішні посилання на дану сторінку. Позначається множина цих сторінок  $M$ . При цьому, в множину  $M$  не потраплять сторінки з того ж хоста (host, фільтрація відбудеться за IP адресою), а також сторінки, які є дзеркалами даної. Множина  $M$  розбивається на підмножини  $L_i$ . У ці підмножини потрапляють сторінки, об'єднані наступними ознаками:

- Приналежність одному (або подібним) хостам. Таким чином, в одну групу потраплять сторінки, у яких перші три октети IP адреси збігаються. Тобто, сторінки, IP адреса яких належить діапазону xxx.xxx.xxx.0 - xxx.xxx.xxx.255 вважаються приналежними одній групі.
- Сторінки, які мають однаковий або схожий зміст (дзеркала, mirrors).
- Сторінки одного сайту (домена).

Кожна сторінка в кожній множині  $L_i$  має деякий ранг (OldScore). З кожної множини вибирається по одній сторінці з найбільшим OldScore, решта виключаються з розгляду. Таким чином, ми отримуємо певну множину  $K$  сторінок, що посилаються на дану сторінку. Сторінки в наборі  $K$  сортуються згідно параметру OldScore, потім в множині  $K$  залишаються тільки  $k$  перших сторінок ( $k$  – деяке задане число). Інші сторінки виключаються з розгляду.

Розрахунок LocalScore. За рештою  $k$  сторінкам відбувається підсумовування їх значень OldScore. Це можна виразити формулою. Після того, як розрахунок LocalScore для кожної сторінки з множини  $N$  закінчений, відбувається розрахунок значень NewScore і пересортування сторінок згідно з новим критерієм.

Поетапно відбувається відбір деякої кількості сторінок відповідних запиту. Це робиться за алгоритмами, що не враховують тематику посилань (наприклад, за релевантністю і загальної посилальної популярності).