

АНАЛИТИЧЕСКАЯ РАЗРАБОТКА АЛГОРИТМОВ РАНЖИРОВАНИЯ СТРАНИЦ И ФУНКЦИИ РЕЛЕВАНТНОСТИ

О.А. Козина, С.В. Лаптенко

Национальный технический университет

"Харьковский политехнический институт", г. Харьков

Рассматриваются новейшие алгоритмы ранжирования страниц и функции релевантности на примере ведущих мировых информационно-поисковых систем (ИПС).

Аналитически разработан собственный алгоритм ранжирования страниц и функции релевантности. Формула расчета релевантности:

$$\text{score}(D, Q) = \sum_{i=1}^n \text{IDF}(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot (1 - b + b \cdot \frac{|D|}{\text{avgdl}})},$$

где Q – запрос содержащий слова q_1, \dots, q_n , (q_i, D) – частота слова q_i в документе D ; D – количество слов в документе, avgdl – средняя длина документа в коллекции; k_1 и b – нормировочные коэффициенты, $\text{IDF}(q_i)$ – обратная документная частота слова q_i .

Выбрана типовая структура ИПС, которая приведена на рис. 1.

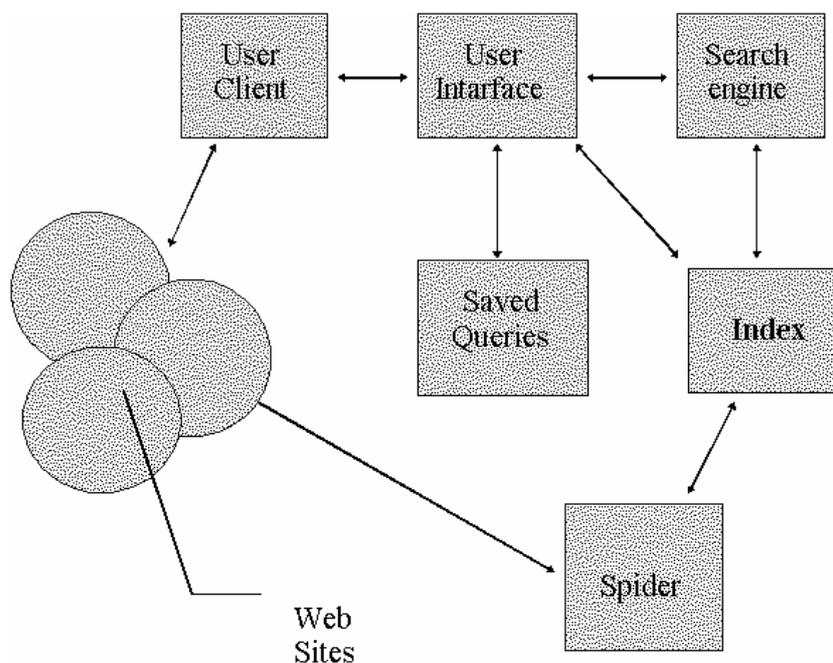


Рис. 1. Типовая схема информационно-поисковой системы

Архитектура блока Index (индекс) устроена таким образом, чтобы поиск происходил максимально быстро и при этом можно оценить ценность каждого из найденных информационных ресурсов сети.