

А. Ю. Чиркова, Л. А. Глухова
(БГУИР, Минск)
АЛГОРИТМЫ ССЫЛОЧНОГО РАНЖИРОВАНИЯ

Ранжирование – это определение ранга согласно релевантности информации. Также ранжированием называется процесс, результатом которого является ответ поисковой системы на запрос пользователя. Ссылочное ранжирование – это влияние текста ссылок на релевантность документа пользовательским запросам, то есть если ключевые слова на какой-либо странице встречаются в анкерах ссылок, идущих на неё с других сайтов, это делает её более значимой. При введении данного фактора разработчики поисковых систем считали, что если сайт ссылается на другую страницу каким-либо текстом, то можно быть уверенным в том, что данная информация присутствует на странице, и чем больше количество таких ссылок, тем выше эта вероятность. Понимание механизма ранжирования дает ответ на вопрос, почему два сайта, посвященных одной тематике, занимают в ответе системы разные места.

Ранг вершины сети, которая представляет собой граф, вычисляется согласно алгоритмам ссылочного ранжирования. Задачей таких алгоритмов является решение системы уравнений, состоящей из уравнений вычисления ранга для всех вершин графа. Наиболее известным алгоритмом является PageRank, предложенный С. Брином и Л. Пейджем в 1997 г. [1]. В этой модели ранг вершины равен вероятности случайного нахождения бродящим по сети пользователем. Эта вероятность складывается из некоторой минимальной вероятности и из суммы вероятностей перехода пользователя с ссылающегося документа, помноженной на некоторый коэффициент затухания.

Другим известным алгоритмом ссылочного ранжирования является алгоритм HITS. Согласно этому алгоритму все страницы делятся на источники (“authority”) и хабы (“hubs”) [2]. Источники являются наиболее популярными ресурсами данной тематики, а хабы – ресурсы, ссылающиеся на множество источников. Всем источникам присваивается начальный ранг a_0 , а каждому хабу – ранг h_0 . Ранг источника равен сумме рангов ссылающихся на него хабов, а ранг хаба – сумме рангов ссылающихся на него источников.

В настоящее время проводится большое количество исследований с целью дальнейшего улучшения и оптимизации поисковых систем.

Литература

- 1 Page, Larry. PageRank: Bringing Order to the Web. Stanford Digital Library Project / L. Page, S. Brin, R. Motwani, T. Winograd. – 1997.
- 2 Kleinberg, J. Authoritative sources in a hyperlinked environment / J. Kleinberg // Journal of the ACM. – № 46, 1999. – pp. 604-632.