

Современные коммерческие поисковые системы полагаются на науку поиска информации (IR). Эта наука существует с середины 20 века, когда поисковые системы использовались в библиотеках, научно-исследовательских центрах и государственных библиотеках. Когда поисковые системы находились на раннем этапе развития, ученые, занимавшиеся исследованием поиска информации, поняли, что два важнейших компонента составляют большинство функциональности поиска:

Релевантность – степень, которой соответствует содержание документов, выданных в результатах поиска, намерению и терминам запроса пользователя. Релевантность документа возрастает, если термины или фраза, использованные в запросе пользователя, встречается несколько раз или присутствует в названии работы.

Популярность – имеет относительное значение, измеряется путем числа цитирований (ссылка в одной работе на другую, как часто встречается в научных или деловых документах) данного документа, который соответствует запросу пользователя. Популярность данного документа возрастает с каждым его цитированием другим документом.

Эти два элемента были введены в сетевой поиск 40 лет спустя. Они являются основными понятиями в анализе документов и анализе ссылок.

В анализе документа, поисковая система определяет, присутствуют ли искомые термины в важных разделах документа – названии, метаинформации, оглавлении и основной части содержания текста. Они также пытаются автоматически определить качество документа (хотя сложные системы не включены в рамки этого руководства).

В анализе ссылок, поисковая система определяет не только кто ссылается на сайт или страницу, но и их мнение о том сайте/странице. Они также хорошо разбираются в том, кто с кем соединен (через архивные данные ссылок, записи регистрации сайта и другие ресурсы); кто достоин доверия (ссылки со страниц .edu и .gov больше ценятся по этой причине); и контекстные данные о сайте, на котором находится эта страница (кто ссылается на этот сайт, их мнение об этом сайте, и т.д.).

Определение релевантности и популярности

Автор: Administrator
21.09.2011 14:50

Анализ ссылок и документов совмещает и накладывает друг на друга сотни факторов, которые могли бы индивидуально определяться и фильтроваться через алгоритм поисковой системы (набор инструкций, которые определяют степень важности каждого фактора). Алгоритм оценивает документы и (в идеале) перечисляет результаты поиска в убывающем порядке степени важности (рейтинг).