

МЕТАПОИСКОВАЯ СИСТЕМА С ИСПОЛЬЗОВАНИЕМ ПОКАЗАТЕЛЯ ОБОБЩЕННОЙ РЕЛЕВАНТНОСТИ ДОКУМЕНТА ДЛЯ ФОРМИРОВАНИЯ РЕЗУЛЬТАТОВ ПОИСКА

Аборок Ш.К., Костин В.И.

Донецкий национальный технический университет

В этом докладе рассматривается механизм работы метапоисковой системы с использованием показателя обобщенной релевантности документа для формирования результатов поиска. Эта метапоисковая система решает проблему полноты и релевантности результатов поиска.

1. Метапоисковые системы

Метапоисковые машины, или метапоисковые системы сами поиск, по сути, не проводят [1]. Метапоисковики переадресовывают ваш запрос на несколько (чем больше - тем лучше) поисковиков разного типа (как ориентирующихся на каталоги, так и использующих различные механизмы поиска), а затем суммируют все полученные ссылки и удаляют из их списка дублирующие. При этом эффективность поиска определяется в основном качеством (и отчасти количеством) поисковых систем, используемых метапоисковиком, поскольку, как правило, он просто приводит в результатах поиска по 10 (или более) ссылок на наиболее релевантные (по мнению опрашиваемых поисковых машин) документы [2]. Все метапоисковые системы можно разделить на две категории - предназначенные для серьезного поиска и системы просто выводящие на одной странице результаты, полученные несколькими поисковиками, с минимальными возможностями по их обработке. К наиболее известным системам первой категории можно отнести метапоисковики SurfWax, Vivisimo и Ixquick. Ко второй - великое множество других машин [1]. Преимущество поиска с помощью подобных систем состоит в увеличении степени охвата ресурсов Internet, поскольку в базах данных различных поисковых машин хранится информация о разных сайтах. Недостаток их использования – невозможность формирования сложных запросов [1, 2], поскольку метапоисковые машины "не переводят" сложный запрос на язык тех ИПС, которые они опрашивают (исключение из этого правила - метапоисковик Ixquick), а также отсутствие механизма оценки релевантности документов непосредственно метапоисковыми системами.

2. Разработка метапоисковой системы с использованием показателя обобщенной релевантности документа для формирования результатов поиска

Для повышения качества результатов поиска предлагается использовать показатель обобщенной релевантности документов по результатам работы разных поисковых систем $R_a'(x)$. Обобщенная релевантность рассчитывается по следующей формуле:

$$R_a'(x) = \frac{\sum_{i=1}^n R_a^i(x)}{kol_ИПС} \cdot , \text{ для } a=\overline{1,m} , \quad (1)$$

где $R_a'(x)$ – показатель обобщенной релевантности документа a запросу пользователя x ;

$R_a^i(x)$ – релевантность i -ой ИПС документа a запросу x ;

m – количество документов, которые стали результатами поиска;

$kol_ИПС$ – количество ИПС, в которых встречается документ;

n – количество ИПС, которые участвовали в поиске.

Схема работы метапоисковой системы с применением показателя обобщенной релевантности документа представлена на рисунке 1.

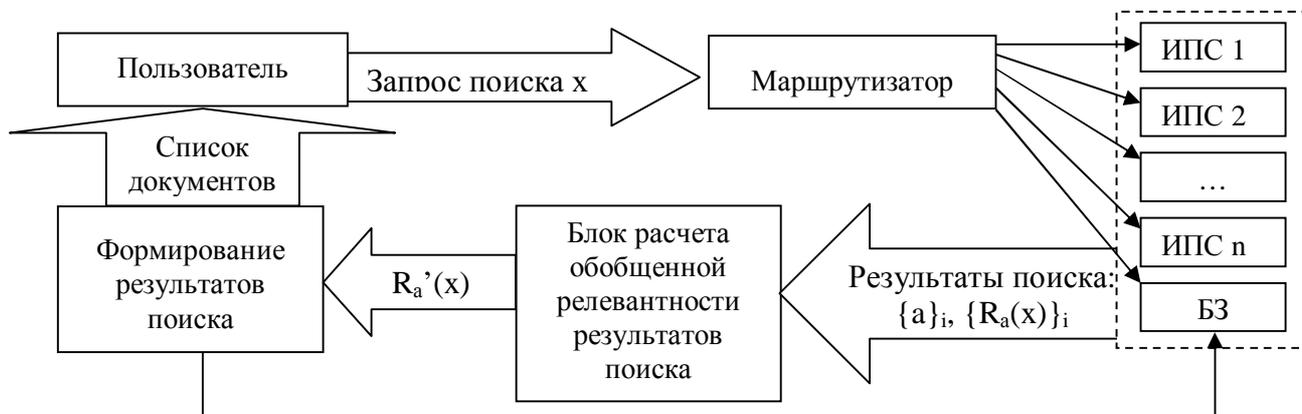


Рисунок 1. Схема работы метапоисковой системы с применением показателя обобщенной релевантности документа

Рассмотрим работу метапоисковой системы с применением показателя обобщенной релевантности документа. Пользователь формирует поисковый запрос x , данные запроса x через маршрутизатор кодируются относительно каждой поисковой системы, которая участвует в поиске, и поступают сразу к нескольким ИПС и в базу недавних запросов (БЗ). База запросов хранит результаты поиска запросов, которые уже были выполнены, в течение определенного времени. Результаты поиска каждой ИПС $_i$ и БЗ (множество документов первых страниц и их релевантность: $\{a\}_i, \{R_a(x)\}_i$) обрабатываются в блоке расчета обобщенной релевантности результатов поиска нашей поисковой системы. В этом блоке рассчитывается показатель обобщенной релевантности $R_a'(x)$ каждого документа-результата a запроса x , на основе $R_{ai}(x)$ всех ИПС. Показатель $R_a'(x)$ рассчитывается по формуле (1). Результаты поиска относительно $R_a'(x)$ формируются в документ с результатами поиска по запросу x .

Заключение

Разработанный механизм работы метапоисковой системы с использованием показателя обобщенной релевантности документа для формирования результатов поиска решает проблему полноты и релевантности результатов поиска. Использование показателя обобщенной релевантности документа повышает релевантность результатов поиска, так как основой для определения уровня релевантности $R_a'(x)$ каждого документа a являются результаты расчета релевантности $R_{ai}(x)$ разными поисковыми системами. Тем самым обеспечивается подтверждение высокого уровня релевантности $R_a'(x)$ документа с действительно высоким уровнем $R_{ai}(x)$ всеми ИПС, либо защита от документов, которые проходят в результаты поиска из-за специфики алгоритма ранжирования ИПС. Полнота поиска достигается за счет эффективного сотрудничества с различными поисковыми системами. Дальнейшие исследования направлены на автоматизацию данного алгоритма и реализацию в сети.

Литература

- [1] Гусев В.С. Поиск в Internet. Самоучитель.: - М.: Издательский дом «Вильямс», 2004.- 336 с.: ил.
- [2] Попов Владимир «Не запутайся в паутине или что нужно знать о поисковых системах» // <http://www.teenclub.ru/index.php?e=153>, 12.02.05...