

image not found or type unknown



С каждым годом объемы Интернета увеличиваются в разы, поэтому вероятность найти необходимую информацию резко возрастает. Интернет объединяет миллионы компьютеров, множество разных сетей, число пользователей увеличивается на 15-80% ежегодно. И, тем не менее, все чаще при обращении к Интернет основной проблемой оказывается не отсутствие искомой информации, а возможность ее найти. Поэтому особенно актуально правильно и грамотно научиться искать информацию.

Чтобы найти нужную информацию, необходимо найти её адрес. Для этого существуют специализированные поисковые сервера (роботы индексов (поисковые системы), тематические Интернет-каталоги, системы мета-поиска, службы поиска людей и т.д.).

Технология поиска - Web-технология World Wide Web (WWW) считается специальной технологией подготовки и размещения документов в сети Интернет. В состав WWW входят и web-страницы, и электронные библиотеки, каталоги, и даже виртуальные музеи. При таком обилии информации остро встает вопрос: «Как сориентироваться в столь огромном и масштабном информационном пространстве?» В решении данной проблемы на помощь приходят поисковые инструменты.

Поисковые инструменты - это особое программное обеспечение, основная цель которого - обеспечить наиболее оптимальный и качественный поиск информации для пользователей Интернета. Поисковые инструменты размещаются на специальных веб-серверах, каждый из которых выполняет определенную функцию:

- 1) Анализ веб-страниц и занесение результатов анализа на тот или иной уровень базы данных поискового сервера.
- 2) Поиск информации по запросу пользователя.
- 3) Обеспечение удобного интерфейса для поиска информации и просмотра результата поиска пользователем.

Приемы работы, используемые при работе с теми или другими поисковыми инструментами, практически одинаковы. Перед тем как перейти к их обсуждению,

рассмотрим следующие понятия:

1) Интерфейс поискового инструмента представлен в виде страницы с гиперссылками, строкой подачи запроса (строкой поиска) и инструментами активизации запроса.

2) Индекс поисковой системы – это информационная база, содержащая результат анализа веб - страниц, составленная по определенным правилам.

3) Запрос – это ключевое слово или фраза, которую вводит пользователь в строку поиска. Для формирования различных запросов используются специальные символы ("", ~), математические символы (*, +, ?)

Схема поиска информации проста. Пользователь набирает ключевую фразу и активизирует поиск, тем самым получает подборку документов по сформулированному (заданному) запросу. Этот список документов ранжируется по определенным критериям так, чтобы вверху списка оказались те документы, которые наиболее соответствуют запросу пользователя.

Большинство поисковых инструментов предлагают два способа поиска – simple search (простой поиск) и advanced search (расширенный поиск) с использованием специальной формы запроса и без нее.

Наиболее развитый сервис поиска русскоязычной информации предоставляет поисковый сервер Яндекс.

Поисковые системы (ПС) уже приличное время являются обязательной частью интернета. Сегодня они громадные и сложнейшие механизмы, которые представляют собой не только инструмент для нахождения любой необходимой информации, но и довольно увлекательные сферы для бизнеса. На сегодняшний день существует множество самых разнообразных поисковиков

Многие пользователи поиска никогда не думали о принципах их работы, о способах обработки пользовательских запросов, о том, как построены и функционируют данные системы. Данный материал поможет людям, которые занимаются оптимизацией и продвижением своих сайтов, понять устройство и основные функции поисковых машин.

Поисковая система – это аппаратно-программный комплекс, который предназначен для осуществления функции поиска в интернете, и реагирующий на пользовательский запрос который обычно задают в виде какой-либо текстовой

фразы (или точнее поискового запроса), выдачей ссылочного списка на информационные источники, осуществляющейся по релевантности. Самые распространенные и крупные системы поиска: Google, Bing, Yahoo, Baidu. В Рунете – Яндекс, Mail.Ru, Рамблер.

Рассмотрим поподробнее само значение запроса для поиска, взяв для примера систему Яндекс.

Запрос обязан быть сформулирован пользователем в полном соответствии с предметом его поиска, максимально просто и кратко. К примеру, мы желаем найти информацию в данном поисковике: «как выбрать автомобиль для себя». Чтобы сделать это, открываем главную страницу и вводим запрос для поиска «как выбрать авто». Потом наши функции сводятся к тому, чтобы зайти по предоставленным ссылкам на информационные источники в сети.

Но даже действуя таким образом, можно и не получить необходимую нам информацию. Если мы получили подобный отрицательный результат, нужно просто переформулировать свой запрос, или же в базе поиска действительно нет никакой полезной информации по данному виду запроса (такое вполне возможно при заданных «узких» параметрах запроса, как, к примеру, «как выбрать автомобиль в Анадыри»).

Самая основная задача каждой поисковой системы – доставить людям именно тот вид информации, который им нужен. А приучить пользователей создавать «правильный» вид запросов к поисковым системам, то есть фразы, которые будут соответствовать их принципам работы, практически, невозможно.

Именно поэтому специалисты-разработчики поисковиков делают такие принципы и алгоритмы их работы, которые бы давали пользователям находить интересующие их сведения. Это означает, что система, должна «думать» так же, как мыслит человек при поиске необходимой информации в интернете.

Когда он вводит свой запрос в поисковую машину, он желает найти то, что ему надо, как можно проще и быстрее. Получив результат, пользователь составляет свою оценку работе системы, руководствуясь несколькими критериями. Получилось ли у него найти нужную информацию? Если нет, то сколько раз ему пришлось переформатировать текст запроса, чтобы найти ее? Насколько актуальная информация была им получена? Как быстро поисковая система обработала его запрос? Насколько удобно были предоставлены поисковые результаты? Был ли нужный результат первым, или находился на 30-ом месте? Сколько «мусора»

(ненужной информации) было найдено вместе с полезными сведениями? Найдется ли актуальная для него информация, при использовании ПС, через неделю, либо через месяц?

В ходе ранжирования, поисковые системы должны точно угадать желание пользователя

Для того чтобы получить правильные ответы на подобные вопросы, разработчики поиска постоянно улучшают принципы ранжирования и его алгоритмы, добавляют им новые возможности и функции и любыми средствами пытаются сделать быстрее работу системы.

Обозначим главные характеристики поиска:

Полнота.

Полнота является одной из главнейших характеристик поиска, она представляет собой отношение цифры найденных по запросу информационных документов к их общему числу в интернете, относящихся к данному запросу. Например, в сети есть 100 страниц имеющих словосочетание «как выбрать авто», а по такому же запросу было отобрано всего 60 из общего количества, то в данном случае полнота поиска составит 0,6. Понятно, что чем полнее сам поиск, тем больше вероятность, что пользователь найдет именно тот документ, который ему необходим, конечно, если он вообще существует.

Точность.

Еще одна основная функция поисковой системы – точность. Она определяет степень соответствия запросу пользователя найденных страниц в Сети. К примеру, если по ключевой фразе «как выбрать автомобиль» найдется сотня документов, в половине из них содержится данное словосочетание, а в остальных просто есть в наличии такие слова (как грамотно выбрать автомагнитолу, и установить ее в автомобиль»), то поисковая точность равна $50/100 = 0,5$.

Чем поиск точнее, тем скорее пользователь найдет необходимую ему информацию, тем меньше разнообразного «мусора» будет встречаться среди результатов, тем меньше найденных документов будут не соответствовать смыслу запроса.

Актуальность.

Это значимая составляющая поиска, которую характеризует время, проходящее с момента опубликования информации в интернете до занесения ее в индексную базу поисковика.

К примеру, на следующий день после возникновения информации о выходе нового iPad, множество пользователей обратилась к поиску с соответствующими видами запросов. В большинстве случаев информация об этой новости уже доступна в поиске, хотя времени с момента ее появления прошло очень мало. Это происходит благодаря наличию у крупных поисковых систем «быстрой базы», которая обновляется несколько раз за день.

Скорость поиска.

Такая функция как скорость поиска теснейшим образом связана с так называемой «устойчивостью к нагрузкам». Ежесекундно к поиску обращается огромное количество людей, подобная загруженность требует значительного сокращения времени для обработки одного запроса. Тут интересы, как поисковой системы, так и пользователя целиком совпадают: посетитель хочет получить результаты как можно быстрее, а поисковая система должна отработать его запрос тоже максимально быстро, чтобы не притормозить обработку последующих запросов.

Наглядность.

Наглядное представление результатов является важнейшим элементом удобства поиска. По множеству запросов поисковая система находит тысячи, а в некоторых случаях и миллионы разных документов. Вследствие нечеткости составления ключевых фраз для поиска или его не точности, даже самые первые результаты запроса не всегда имеют только нужные сведения.

Это значит, что человеку часто приходится осуществлять собственный поиск среди предоставленных результатов. Разнообразные компоненты страниц выдачи ПС помогают ориентироваться в поисковых результатах.

Все эти элементы тесно связаны между собой и функционируют, взаимодействуя, образуя отчетливый, но достаточно непростой механизм функционирования ПС, требующий громадных затрат ресурсов.