

image not found or type unknown



Данная работа посвящена описанию поисковых систем, которые осуществляют поиск информации.

В информационно-поисковой системе должен храниться весь необходимый информационный массив, из которого по требованиям пользователей выдается нужная информация. Поиск информации по требованию пользователя осуществляется либо автоматически, либо вручную (как в библиотеках, когда с запросом к работнику справочного фонда обращается читатель, а работник пользуется системой каталогов).

Во втором случае используются ЭВМ, снабженные специальными программными средствами, анализирующими процессы запросов, поиска и выдачи нужных документов. Таким образом, информационно-поисковые системы (ИПС) реализуют вопросно-ответное отношение, что сближает задачи, стоящие перед создателями таких систем, с теми задачами, которые решают создатели человеко-машинных систем.

Поиск информации является одной из наиболее распространенных и одновременно наиболее сложных задач, с которыми приходится сталкиваться в Сети любому пользователю. Однако если для рядового члена сетевого сообщества знание методов эффективного информационного поиска является желательным, но далеко не обязательным качеством, то для работников высокоинтеллектуальной сферы умение быстро ориентироваться в ресурсах Интернет и находить требуемые источники сегодня относится уже к числу базовых квалификационных навыков.

Цель работы – описать и дать характеристику информационно-поисковым системам.

Данная цель решается с помощью раскрытия следующих основных задач:

- 1) описать принципы работы поисковых машин;
- 2) дать характеристику глобальным поисковым системам;
- 3) описать стратегию и методику профессионального поиска информации.

Сущность поисковых машин

Задача поисковых машин - обеспечивать детальное разыскание информации в электронной вселенной, что может быть достигнуто только за счет учета (индексирования) всего содержания максимально возможного числа web-страниц. В отличие от справочников, все они функционируют в автоматизированном режиме и имеют одинаковый принцип действия. Поисковые системы состоят из двух базовых компонентов. Первый компонент представляет собой программу-робот, задача которого путешествовать с сервера на сервер, находить там новые или изменившиеся документы и скачивать их на главный компьютер системы. При этом робот, просматривая содержимое документа, находит новые ссылки, как на другие документы данного сервера, так и на внешние сайты. Программа самостоятельно направляется по указанным ссылкам, находит новые документы и ссылки в них, после чего процесс повторяется вновь, напоминая хорошо известный в библиографии "метод снежного кома".

Выявленные документы обрабатываются (индексируются) вторым компонентом поисковой системы. При этом, как правило, учитывается все содержание страницы, включая текст, иллюстрации, аудио и видео файлы и пр. Индексации подвергаются все слова в документе, что как раз и дает возможность использовать поисковые системы для детального поиска по самой узкой тематике. Образующие гигантские индексные файлы, хранящие информацию о том какое слово, сколько раз, в каком документе и на каком сервере употребляется и составляют базу данных, к которой происходит обращение пользователей, вводящих в строку запроса сочетание ключевых слов.

Выдача результатов осуществляется с помощью специального модуля, который производит интеллектуальное ранжирование результатов. При этом берется в расчет местоположение термина в документе (название, заголовок, основной текст), частота его повторения, процентное соотношение искомого термина к остальному тексту страницы, а также число и авторитетность внешних ссылок на данную страницу с других сайтов.

Основные параметры поисковых машин

К основным параметрам поисковых систем относятся:

- объем индексных файлов (число проиндексированных серверов и отдельных документов);
- степень оперативности обновления базы данных за счет включения сведений о новых материалах и удаления устаревших;

- возможности для составления запроса;
- интеллектуальность системы ранжирования результатов поиска;
- наличие дополнительных сервисных функций, облегчающих работу пользователя.

Первая величина, являющаяся ключевой, устанавливает широту охвата материала и определяется числом проиндексированных документов. Сейчас эта цифра для лидеров мирового сетевого поиска колеблется в пределах от 1 до 3 с лишним миллиардов.

Учитывая тот факт, что в среднем интернетовский адрес сохраняет актуальность до полугода, после чего документ или меняет местоположение или убирается с сервера, большое значение имеет уровень оперативности обновления данных, характеризующий степень соответствия индексного файла поисковой системы реальному местоположению документов на сайтах. В настоящее время этот параметр колеблется от двух недель до полутора месяцев.

Возможности поискового механизма выразить запрос максимально точно в значительной степени определяют долю релевантных документов в перечне полученных результатов. Каждая машина имеет свою собственную лексику, которая по-разному позволяет детализировать поисковое предписание.

Все поисковые машины обладают модулем ранжирования результатов поиска. Создание таких модулей - целая область программирования, в которой конкурируют сложнейшие алгоритмы, созданные разными компаниями. Перечень факторов, принимаемых во внимание при определении места документа в перечне ссылок необычайно широк: от местоположения слова на странице до рейтинга (авторитета) страниц, имеющих ссылки на найденный документ.

Не последнюю роль играет и простота интерфейса, наличие дополнительных сервисных функций, как например, возможность перевода текста документа на иностранный язык, способность выделять все документы с определенного сайта, сужение критериев в ходе поиска, нахождение документов "по образцу" и т.д.

Практически все всемирно известные справочники и поисковые системы в настоящее время превратились во внушительные информационные корпорации с многомиллионными доходами. Заработав авторитет наиболее посещаемых мест в Сети, они предоставляют свои страницы для размещения рекламной информации, доходы от которой и составляют основу их бюджета. Постепенно поисковые

сервера превращаются в многофункциональные порталы, в которых поисковый сервис остается главной приманкой для пользователей, но далеко не единственной и даже не основной из предоставляемых услуг. Помимо разыскания информации, такие сервера обычно предоставляют пользователям бесплатную электронную почту, возможность бесплатно размещать собственные страницы, сведения о погоде, текущих новостях, биржевые котировки, карты местности и т.д.

Список использованной литературы:

1. Генкин А. Планета Web-денег, "Альпина Паблишер", 2003 г., 510 с.
2. Денисов А. И др. Интернет: Самоучитель. – СПб.: Питер, 2005.
3. Джерк Н. Разработка приложений для электронной коммерции., "ПИТЕР", 2007 г., 512 стр.
4. Ковалев А., Курдюмов И. и др. Управление проектом по созданию интернет-сайта, "Альпина Паблишер", 2005г., 337с.
5. Козье д. Электронная коммерция: Пер. с англ. -М.:Издательско-торговый дом "Русская редакция". 2009.-288с
6. Рейнхоулд А., Левин Дж.Р., Левин-Ян М. – Internet для «Чайников»: краткий справочник – Киев: Диалектика, 2006.
7. Соломенчук В. Интернет: краткий курс – СПб.: Питер, 2004.