Первые поисковые системы появились в сети Интернет более десяти лет назад. Тогда они выполняли лишь одну функцию – поиска ссылок к недавно созданным страницам.

На начальном этапе развития интернета, число пользователей сети было невелико и количество информации относительно небольшим. В подавляющем большинстве случаев пользователями Интернет были сотрудники различных университетов или научных организаций. В то время поиск нужной информации в сети был не столь актуален, как теперь. Сегодня же поисковые системы превратились в многофункциональный сервис. Они позволяют пользователям находить в сети Интернет самую разнообразную информацию, благодаря чему пользуются огромным успехом.

Internet является наиболее важным и наиболее часто используемым источником информации. Наибольшая полезность глобальной сети состоит в доступности информации любому пользователю и регулярной обновляемости ресурсов, что позволяет постоянно быть в курсе новых событий.

Поисковые системы обычно состоят из трех компонент:

• агент (паук или кроулер), который перемещается по Сети и собирает информацию;

•база данных, которая содержит всю информацию, собираемую пауками;

•поисковый механизм, который люди используют как интерфейс для взаимодействия с базой данных.

Во время путешествия по Интернету, вам обязательно понадобиться помощь поисковой машины. Очень часто приходится искать информацию в Сети не зная даже приблизительно адрес страницы, на которой она может располагаться. В таких случаях на помощь приходит поисковая машина.

Поисковые машины - это роботизированные системы. Специальная программа-робот, которую называют паук или ползун, постоянно обходит Сеть в поисках новой информации, которую она вносит в базу данных. База данных содержит URL-адреса

и проиндексированную информацию, связанную с этими адресами.

При поиске в Интернете важны две составляющие – полнота (ничего не потеряно) и точность (не найдено ничего лишнего). Обычно это все называют одним словом – релевантность, то есть соответствие ответа вопросу. Важными показателями являются охват и глубина поисковой машины (насколько велика база данных по документам), скоростью обхода и актуальностью ссылок (скорость обновления информации в этой базе данных), качеством поиска (чем ближе к началу списка оказывается нужный вам документ, тем лучше работает релевантность).

Кроме релевантности, существуют важные пользовательские характеристики: скорость поиска (медленная поисковая машина неэффективна в работе), поисковые возможности (как именно происходит индексация: только по ключевым словам web-страницы или по всему тексту, с учетом морфологии или без него, с поиском по тэгам HTML - заголовкам, ссылкам, подписям к изображениям и др.), а также дополнительные удобства (удобный интерфейс, наличие специальных функций, например, поиск по датам и серверам). Здесь все зависит от того, что вы предпочитаете.

Среди ведущих поисковых машин на данный момент - Яndex, Google, Rambler, Апорт! и др.

Основная часть

Одной из первых попыток организации доступа к информационным ресурсам сети стало создание тематических каталогов сайтов. Первым, открывшимся в апреле 1994 г, стал Yahoo. Это еще не было поисковой системой, в современном понимании, т.к. возможность поиска информации ограничивалась ресурсами, зарегистрированными в каталоге Yahoo. Каталоги ссылок ранее использовались довольно широко, но в настоящее время практически утратили свою популярность. Объяснение этому очень простое – даже современные, содержащие огромное количество ресурсов каталоги, представляют информацию лишь о довольно незначительной части сети. Для сравнения - самый полный каталог сети интернет – DMOZ содержит информацию примерно о 12.000.000 ресурсов, в то время как база данных самой полной поисковой системы Google состоит более чем из 28.000.000.000 документов.

Первой полноценной поисковой системой в 1994г. стал проект WebCrawler. Далее в 1995 году появились поисковые системы AltaVista и Lycos. В 1997 году в Стэнфордском университете, в рамках исследовательского проекта, была создана

Google - самая популярная поисковая система на данный момент в мире. В 1997 году появилась поисковая система - Yandex, лидер в русскоязычной части Интернета. На данный момент основными поисковыми системами являются три международных – Google, Yahoo и MSN Search. Остальные, коих не мало, используют целиком или частично базы и (или) алгоритмы выше приведенных систем. В Рунете основной поисковой системой является Яндекс, далее по популярности идут Rambler, Google.ru, Mail.ru и Aport.

Поисковая система – это сумма следующих компонентов:

Web server (веб-сервер) – сервер поисковой машины, который осуществляет взаимодействие между пользователем и остальными компонентами системы.

Spider (паук) – программа написанная по принципу браузера, предназначена для скачивания веб-страниц. Браузер предназначен для визуального использования страниц, а паук работает с HTML кодом напрямую. Чтобы посмотреть "сырой" исходник нажмите в меню браузера: Вид- Просмотр HTML кода.

Crawler («путешествующий» паук) – программа, которая автоматически уходит по всем внешним ссылкам страницы. Ее задача - поиск не известных (или измененных) документов и в расстановке приоритетов, куда дальше должен идти Spider.

Indexer (индексатор) – программа-анализатор скаченных пауками веб-страниц. Она "разбирает" на части скачанную страницу и анализирует ее элементы, такие как текст, служебные html-теги, заголовки, особенности стилистики и структурные формы.

Database (база данных) – хранилище для скачанных и обработанных страниц - общая база данных поисковой машины.

Search engine results engine (система выдачи результатов) – извлекает результаты поиска из базы данных поисковой системы. Именно она решает, какие страницы более соответствуют запросу пользователя и отсортировывает их в нужном порядке. Модуль работает согласно заданным поисковой системой алгоритмам ранжирования.

Заключение

Лидерами по показателям качества представленной информации оказались Yandex, Google и Aport.

Yandex оказался одной из наиболее эффективных систем с точки зрения ее релевантности и соответствия выданных результатов заданному запросу. Хоть страниц было много, но нужная информация находилась на самых первых из них. Мало затраченного времени – необходимые результаты. При этом немаловажную роль сыграла также относительная новизна представленной информации.

Google выдавал результаты страниц, на которых не всегда первое место занимали релевантные документы. Зато жалоб на разнообразие просто не было, т.к. в представленном количестве материала можно было найти что угодно (при этом было потрачено времени в два раза больше, по сравнению с поисками в других поисковых системах).

Система Aport оказалась менее эффективной, чем вышеназванные из-за ее чрезмерной ориентации на частные случаи, но результаты, которые она выдавала, значительно отличались от результатов других поисковых систем. Они были единственные в своем роде, не всегда релевантны, но неповторимы.

Rambler, несмотря на прочно занимаемое четвертое место в количественном рейтинге, оказался намного менее эффективной по релевантности системой. Здесь преобладает ориентация на российские источники информации, что снижает ее адекватность в оценке ситуации в других странах. К этой же категории по степени релевантности можно отнести и поисковую систему Google.

Поисковую систему Yahoo можно рассматривать как наиболее эффективную наравне с Yandex, но только в англоязычном поиске. На русском языке в данной системе имеется незначительное количество сайтов и их релевантность минимальна.

Степень актуальности того или иного предмета исследования определяется, главным образом, исходя из объема существующей по данному вопросу литературы. В ходе осуществленного поиска в Internet мною было найдено большое количество информации, касающейся классификации, обзора и анализа современных поисковых систем. Исходя из объема представленной литературы как на английском, так и на русском языках, можно сделать вывод, что к настоящему времени поисковые системы пользуются огромным спросом среди пользователей сети Internet.