

image not found or type unknown



Первые поисковые системы появились в сети Интернет более десяти лет назад. Тогда они выполняли лишь одну функцию – поиска ссылок к недавно созданным страницам. 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 скаченных 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. Поисковики – это своего рода монополисты. Они ведут скрытую политику, диктуют свои правила, не разглашая их при этом. А подчас, даже просто беспредельничают и показывают поразительную необъективность по отношению к сайтам. Возникает ситуация, что оптимизаторы не могут ясно понять, каким он должен быть, «хороший» сайт в понимании поисковика. Как сделать его таким, чтобы поисковик считал его наиболее релевантным по запросам. И именно из-за этой неопределённости и возникают у вэб-мастеров желание обхитрить систем.

Выход из этой проблемы – создание полностью ОТКРЫТОЙ поисковой системы, в которой бы ранжированием занимался не только поисковик но и сами веб-мастера. Чтобы в ней возможно было открыть все механизмы и алгоритмы ранжирования и при этом объективно оценивать сайты между собой.

Главный недостаток современных поисковых систем – это их централизация. А централизация означает, что вся информация хранится в одном месте, все работы и расчёты производятся в одном месте, все решения (результаты выдачи) принимаются в одном месте.