



# ПОИСКОВЫЙ СЕРВЕР

Авторы: Д. В. Барашев, Н. С. Васильева, Б. А. Новиков

---

ПОИСКОВЫЙ СЕРВЕР (поисковая машина), программно-аппаратный комплекс, предназначенный для формирования поисковой выдачи (поиска, хранения и предоставления информации пользователям), состоящей из документов справочно-информационного фонда (СИФ) релевантных (адекватных) поисковому запросу, формулирующему информационную потребность конечного пользователя. Обычно является частью [ПОИСКОВОЙ СИСТЕМЫ](#).

## Историческая справка

Первые прототипы П. с. (1960-е гг.) были способны выполнять поиск в коллекциях довольно скромного размера (порядка тысячи документов). В нач. 1970-х гг. информационно-поисковые системы (см. [Информационная система](#)) предоставляли возможность поиска библиографич. данных по ключевым словам или словосочетаниям (дискрипторам). Задача составления образов поисковых документов в таких информац. системах решалась вручную, спец. команды вводились с клавиатуры, а время составления поисковой выдачи по запросу измерялось минутами и часами. По мере того как количество присоединённых к [Интернету](#) компьютеров увеличивалось, назрел вопрос о необходимости создания инструментов, позволяющих легко находить текстовые и др. файлы на удалённом компьютере, в идеале – на любом, где бы он ни располагался в компьютерной сети. В кон. 1980-х гг. стали появляться специальные FTP-серверы (File Transfer Protocol – протокол передачи файлов), которые используют разные сетевые соединения для передачи команд и данных между клиентом и сервером. Первый работоспособный, доступный широкой публике П. с., индексирующий имена файлов (процесс описания документов и запросов, в результате чего каждому документу назначается набор ключевых слов, отражающих его смысловое содержание) на FTP-серверах, – «Арчи» («Archie»),

создан в 1990 в Монреале; был способен индексировать документы автоматически и оценивать релевантность нескольких сотен документов в секунду.

Развитие Интернета и языка HTML (HyperText Markup Language – язык разметки гипертекстовых документов во [Всемирной паутине](#); разработан брит. учёным Т. Бернерсом-Ли в 1989 в ЦЕРНе) положило начало П. с., осуществляющим полнотекстовый поиск, а изобретение и пром. реализация в кон. 1990-х гг. индекса цитирования HTML документов привели к существенному увеличению качества поиска. В это же время достижения в технологиях хранения и обработки данных, в частности использование недорогого широкодоступного аппаратного обеспечения, позволили решить мн. проблемы горизонтального масштабирования П. с. (добавление новых серверов с перераспределением нагрузки между ними).

## **Особенности технологий и тенденции развития**

П. с., предназначенный для поиска в Интернете, как правило, включает в себя три осн. компонента: поискового робота, службу индексирования и службу поиска. Задачей поискового робота является пополнение и постоянное обновление СИФ. Робот постоянно «обходит» веб-серверы Всемирной паутины, получает их текст и формирует входной массив документов для службы индексирования. Эта служба занимается построением поискового индекса (хранилище данных, в котором сосредоточены копии всех посещённых роботом страниц), обновлением индексов цитирования страниц и др. статистич. информации. Служба поиска использует поисковый индекс для выполнения пользовательских запросов. Кроме трёх осн. компонент, существенную роль в П. с. играет служба журналирования (файловая система, в которой осуществляется ведение журнала, хранящего список изменений и позволяющего сохранить целостность системы при сбоях) и анализа действий пользователей, предназначенная для сбора обратной связи и последующего улучшения качества поиска.

Для формирования поисковой выдачи П. с. может провести прямой просмотр текстов всех документов СИФ (используются эффективные способы представления и сжатия информации, статистич. расчёты и др.). Такая стратегия поиска имеет некоторые преимущества, но при росте коллекции документов становится чрезмерно трудоёмкой.

Альтернативой является предварит. обработка документов и формирование индексов. Возможна комбиниров. стратегия, когда вместе с индексом П. с. использует в процессе выполнения запроса тексты документов. Релевантность документа определяется П. с. на основе множества факторов: рассматриваются (в зависимости от модели поиска) информативность поисковых терминов, степень совпадения документа и запроса, индекс цитирования документа, абсолютное расположение поисковых терминов в тексте документа и их расположение относительно друг друга и многие др. показатели.

П. с., обслуживающие большие потоки поисковых запросов и большие коллекции документов, способствуют развитию технологий хранения и передачи данных. Активно развиваются объектографич. поиск, в котором в поисковой выдаче, помимо текстовых документов, присутствуют факты об объектах предметной области, персонализиров. поиск, придающий существенный вес информации о предпочтениях конечного пользователя и истории его поисковых запросов.

## **Литература**

Лит.: Сегалович И. В. Как работают поисковые системы // Мир Internet. 2002. № 10;  
Маннинг К. Д., Рагхаван П., Шютце Х. Введение в информационный поиск. М., 2011.