

Основные понятия

Инструменты поиска информации в Интернет разделяют на **поисковые и метапоисковые** системы.

Поисковые системы – это специально созданные узлы WWW, которые свободно (бесплатно) доступны, и основная функция которых состоит в поиске информации среди различных информационных ресурсов Internet (серверы WWW, FTP, Gopher, телеконференции, UseNet, списки адресов E-mail).

Поисковые системы классифицируются по методам поиска информации. Существует три основных типа таких систем:

1. поисковые каталоги;
2. автоматические индексы;
3. гибридные системы поиска.

Поисковые каталоги.

В поисковый каталог (поисковую систему первого рода) – информация заносится по инициативе человека. Специальные администраторы выполняют работу по поиску и классификации информации. Добавляемая страница должна быть жестко привязана к принятым в каталоге категориям. Результатом является постоянно обновляющийся иерархический (древовидный) каталог, на верхнем уровне которого собраны самые общие категории, такие как «бизнес», «наука», «искусство» и т.п., а элементы самого нижнего уровня представляют собой ссылки на отдельные web-страницы и сервера вместе с кратким описанием их содержимого.

Предметные каталоги предоставляют также возможность автоматического поиска по ключевым словам. Только поиск происходит не в содержимом самих web-серверов, а в их кратких описаниях, хранящихся в каталоге. Поисковые каталоги обеспечивают высокое качество поиска и по качеству сортировки документов намного превосходят поисковые индексы, проигрывая им по количеству просмотренных документов.

Наиболее известные поисковые каталоги:

- | | |
|-------------|---|
| 1. Yahoo! | http://www.yahoo.com |
| 2. Magellan | http://www.magellan.com |
| 3. InfoSeek | http://guide.infoseek.com |

Автоматические индексы.

Автоматические индексы (поисковые индексы, поисковые системы второго рода, spiders, crawlers) постоянно исследуют Интернет с целью пополнения своих баз данных. Обычно это не требует никаких усилий со стороны человека. Программа, в которую загрузили несколько тысяч общеизвестных URL-адресов, будучи запущена на компьютере с доступом к WWW, начинает автоматически скачивать из сети документы по этим URL, причем из каждого нового документа она извлекает все содержащиеся в нем ссылки и добавляет их в свою базу адресов. Данные инструменты полностью скрывают от пользователя организацию индекса и его содержимое.

Автоматический индекс состоит из трех частей:

1. программы-робота(паука),
2. базы данных, собираемой этим роботом и
3. интерфейса для поиска в базе, с которым и работает пользователь.

Преимуществами этих систем являются большие базы данных и высокая скорость обработки запросов.

К недостаткам можно отнести очень большое количество ссылок на выходе, причем большая часть этих документов имеет весьма отдаленное отношение к нужной теме.

Наиболее известные поисковые индексы:

1. Alta Vista <http://www.altavista.com>
2. HotBot <http://www.hotbot.com>
3. Open Text <http://www.opentext.com>

Гибридные системы поиска.

Сочетают в себе поисковые каталоги и поисковые индексы.

Наиболее известные гибридные системы поиска:

1. Lycos <http://www.lycos.com>
2. Excite <http://www.excite.com>
3. WebCrawler <http://www.webcrawler.com>

Метапоисковые системы.

Метапоисковые системы (поисковые службы) рассылают запросы пользователя одновременно нескольким поисковым серверам и обращаются к некоторым другим источникам информации. Затем они объединяют полученные результаты и представляют их пользователю в виде HTML-страницы с активизируемыми URL-ссылками.

Например:

1. AccuFind Search Service - <http://accufind.com>
2. SuperSearch Sere - <http://www.robtex.com/search/query.htm>
3. Meta Search - <http://metasearch.com>

Web-порталы – специализированные страницы, обеспечивающие удобный интерфейс доступа к поисковым системам и Web-узлам, представляющим всеобщий интерес. В MS Internet Explorer существует возможность обращения к мини-порталу, поддерживаемому компанией Microsoft. Для этого предназначена панель **Поиск**. С нее можно получить доступ сразу к нескольким поисковым серверам. В зависимости от версии и языка Internet Explorer этот список меняется.