Практическая занятие

Тема: Поисковые системы.

Цель: Закрепить понятия информационно-поисковые системы, программы-роботы, виды информационно-поисковых систем, язык запроса.

Задание:

1. Внимательно изучить теоретические сведения.
2. Составить таблицу пяти популярных поисковых систем.
3. Дать письменную характеристику каждой поисковой системе.
4. Ответить на контрольные вопросы.

Теоретические сведения

В настоящее время существует множество справочных служб Интернет, помогающих пользователям найти нужную информацию. В таких службах используется обычный принцип поиска в неструктурированных документах по ключевым словам.

Поисковая система- это комплекс программ и мощных компьютеров, способные принимать, анализировать и обслуживать запросы пользователей по поиску информации в Интернет.

Как работает поисковая система?

Многие пользователи почему-то считают, что в ответ на их вопрос поисковая система начинает быстро-быстро шерстить весь Интернет, после чего выдает список страниц, где отыскалось что-то похожее на строку поиска, введенную пользователем. Разумеется, в действительности все происходит совершенно не так, потому, что, если бы поисковая система при каждом запросе шерстила весь Интернет (или даже его маленькую часть), ответа приходилось бы дожидаться годами.

На самом деле почти каждая поисковая система (скажем так - классическая поисковая система) состоит из трех основных компонентов:

1) веб-паук (web spider)
2) индексатор
3) алгоритм поиска и оценки результатов

Веб-паук, несмотря на всю экзотичность подобного названия, - это всего-навсего специальная программа, которая запускается на компьютере (компьютерах), подключенном к Интернету, и ее основная задача - шерстить весь этот Интернет (точнее, интернет-странички) во всех возможных направлениях. Ведь странички, как мы говорили, состоят из гиперссылок, так вот паук носится как угорелый по гиперссылкам "паутины" сайтов и

скачивает полученные таким образом странички для второй компоненты - индексной базы.

Индексатор - это обработчик скачанных веб-пауком страниц. Он извлекает оттуда все слова и складывает их в поисковую базу (индексную базу). При этом индексатор записывает, где именно было найдено то или иное слово, и эта информация потом используется в поиске.

Алгоритм поиска - это главное ноу-хау любой поисковой системы. От алгоритма зависит эффективность полученного результата - то есть насколько быстро и точно пользователь найдет то, что его интересует.

Внимание! Таким образом, когда пользователь вводит свой запрос, поисковая система ищет ответ в своей индексной базе и выводит результаты в соответствии со своим алгоритмом поиска.

Для хорошей работы поисковой системы важны все три компоненты. Причем каждая из них на самом деле, весьма сложна, и ее работа подчиняется огромному количеству всевозможных хитрых правил, которые к тому же постоянно корректируются.

Поисковик должен иметь и шустрого паука, и мощную индексную базу, и эффективный алгоритм поиска - только тогда вы будете получать действительно полезные результаты.

Служба World Wide Web (WWW) - это единое информационное пространство, состоящее из сотен миллионов взаимосвязанных электронных документов.

Информационно-поисковые системы – системы, позволяющие хранить, обрабатывать различные каталоги, списки, справочники, т.е. представляют собой электронный вариант соответствующих бумажных аналогов.

В Интернет обращаются за определенной информацией. Чтобы открыть нужную Web-страницу, надо иметь либо ее адрес, либо другую страницу со ссылкой на нее. Если нет ни того ни другого, обращаются к поисковым системам. Поисковая система представляет собой специализированный Web-узел. Поисковые системы классифицируются по методам поиска.

Специальные программы-роботы круглосуточно сканируют пространство Интернет на предмет обнаружения новых документов. Новые документы просматриваются, выдаются ключевые слова и краткое содержание. Выявленная информация размещается в БД поисковой системы, упорядочивается (индексируется) и классифицируется по темам. При поступлении запроса на поиск от пользователя поисковая система находит

ответ в своей БД и выдает пользователю ранжированный по релевантности список адресов серверов с аннотациями, на которых “по мнению поисковой системы” находится искомая информация. Из вышесказанного следует, что индексные поисковые системы – это системы, в которых все операции производятся автоматизировано (например, www.yandex.ru , www.rambler.ru, www.Google.ru …).

Поисковый индекс обеспечивает поиск по заданным ключевым словам. В результате поиска формируется набор гиперссылок на Web-странице, содержащие указанные термины. Поисковые индексы предоставляют грандиозную широту поиска.

Различают следующие виды информационно-поисковых систем: юридические, отраслевые, электронные справочники, электронные обучающие системы.

К средствам поисковых систем относится язык запросов, он различен для разных систем, но имеет определенные характерные особенности. Рассмотрим язык запросов одной из популярных систем Яndex. Характерной особенностью этой поисковой машины является учет морфологии языка, т. е. поиск осуществляется по всем формам слова или слов в запросе. Например: если задано слово “лить”, то в результате поиска будут предложены и документы со словами “льет”. Стоит отметить, что не все поисковики обладают такой “способностью”.

Контрольные вопросы:

1. Что такое поисковая система?
2. Что лежит в основе поисковой системы?
3. Чем отличаются друг от друга поисковые системы?
4. Какую поисковую систему Вы предпочитаете и почему?
5. Что такое алгоритм поиска?