

## Лабораторная работа №25

**Тема:** «Использование информационно-поисковых систем Internet»

**Цели работы:** познакомиться с существующими в интернете каталогами и поисковыми машинами, видами информационных ресурсов Интернета; получить представление о способах хранения и классификации найденной информации; создать поисковик в Google CSE по теме проекта.

### Теоретическая часть

**Телекоммуникационная сеть** реализует синтез компьютерных сетей и средств телефонной, телевизионной, спутниковой связи. Общедоступные телекоммуникационные сети связаны друг с другом, чтобы их пользователи могли обмениваться информацией и на сегодняшний день ярким примером сообщества таких сетей является Интернет.

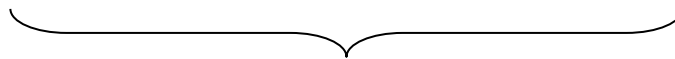
Интернет – это крупнейшая всемирная компьютерная сеть, которая является огромной открытой информационной системой, способной удовлетворить запросы широкого круга пользователей.

Одним из наиболее распространённых сервисов сети Интернет является WWW (World Wide Web) или всемирная паутина – распределенная информационная среда, содержащая разнообразную информацию (в том числе мультимедиа-компоненты Web-страниц), обладающая определенными возможностями распределенной базы данных, хотя и не предполагающая наличие единой структуры хранящейся информации.

Основными компонентами всемирной паутины являются гипертекстовые документы (веб-страницы), каждая из которых имеет свой адрес, определяющий её местонахождение в сети Интернет.

Все адреса ресурсов сети Интернет имеют одинаковую структуру, которая в общем виде может быть представлена следующим образом:

**<Протокол передачи данных>://<сервисная служба Интернета>.<имя сервера>**



Доменное имя

Например,

<http://www.google.ru>, где **левая часть адреса** указывает на то, что доступ к информационному ресурсу осуществляется при помощи протокола передачи данных HTTP (http – протокол передачи данных); **правая часть адреса** указывает на многоуровневое доменное имя компьютера, в котором домены отделяются друг от

друга точками: www – всемирная паутина (сервисная служба Интернета), google – имя сервера, ru – географическое расположение сервера.

Для указания полного адреса к документу в сети после доменного имени через символ / - прямой слеш указывается имя веб-ресурса, которое записывается так же, как путь от коревого каталога диска к нужному файлу.

**<Протокол передачи данных>://<доменное имя>/<полный путь к файлу>**

Например,

<http://www.google.com/coop/cse/>

Такой адрес веб-ресурса, называется **URL** (Uniform Resource Locator) – унифицированная ссылка на ресурс.

Сегодня в сети Интернет накоплены значительные русскоязычные ресурсы для образовательных целей. Они сосредоточены на образовательных порталах, серверах университетов, школ, других образовательных организаций, библиотек. Обучающие курсы и программы, интерактивные учебные пособия, полнотекстовые базы данных, учебная литература, справочники, энциклопедии, статьи из журналов и сборников, материалы научных конференций, готовые рефераты, авторефераты, диссертации – вот далеко не полный перечень ресурсных возможностей.

Важно помнить, что ресурсы Интернет динамично развиваются, поэтому и предметные указатели на их местонахождение в сети постоянно обновляются.

Рассмотрим, как осуществляется поиск информации в сети Интернет.

Поиск информации в сети - это последовательность действий, от определения предмета поиска, до получения ответа на имеющиеся вопросы с использованием всех поисковых сервисов, которые предоставляет сегодня Интернет.

Поиск информации в сети интернет может осуществляться двумя способами:

- по известному URL-адресу ресурса в сети (это самый быстрый и надёжный вид поиска);

- с использованием поисковых систем (осуществляется по запросу пользователя с использованием поисковых инструментов WWW).

1. Для поиска веб-ресурса по известному URL-адресу необходимо запустить программу-браузер, например Opera и в адресной строке ввести <http://gbpouber.tt.ucoz.net>;

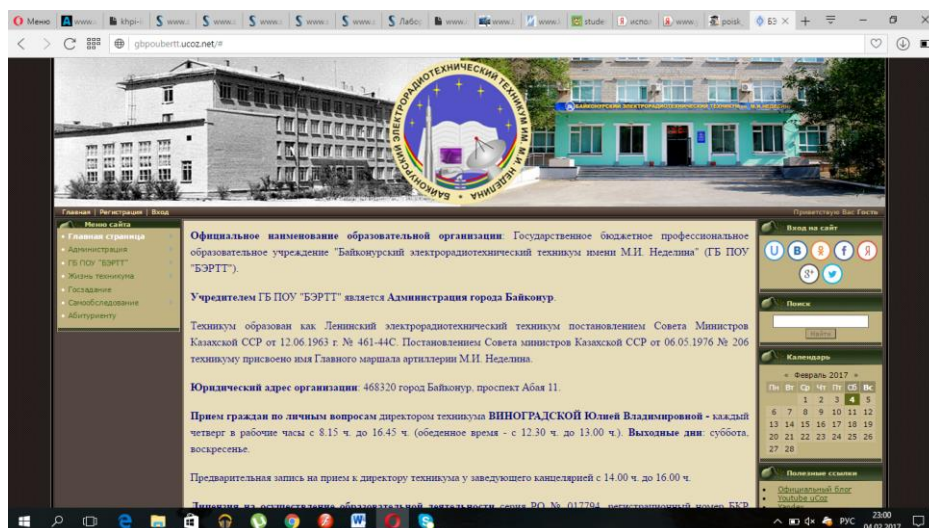


Рисунок 1. Внешний вид окна браузера при загрузке веб-страницы

2. Вопрос выбора поисковых инструментов WWW. В Интернет представлена информация на любые темы, которые только можно себе представить. Но найти нужную информацию не так-то легко из-за того, что сеть по своей природе не имеет четкой структуры. Поэтому для ориентировки в Интернете и быстрого получения свежей справочной информации разработаны системы поиска информации. Все системы поиска информации Интернет располагаются на специально выделенных компьютерах с мощными каналами связи. Ежеминутно они бесплатно обслуживают огромное количество клиентов.

Поисковые системы можно разбить на два типа:

- предметные каталоги, формируемые людьми-редакторами;
- автоматические индексы, формируемые специальными компьютерными программами, без участия людей.

### *Предметные каталоги*

Предметные каталоги представляют собой систематизированную коллекцию (подборку) ссылок на ресурсы Интернета. Ссылки организованы в виде тематического рубрикатора, представляющего собой иерархическую структуру, перемещаясь по которой, можно найти нужную информацию.

Механизм поиска в предметных каталогах и поисковых системах практически одинаков. Различия возникают только на этапе 2 (составление списка ключевых слов). Для поисковой машины - это составление списка ключевых слов, а для каталога на этом этапе производится определение тематики разделов, в которых может находиться необходимая Вам информация.

Как показывает практика, при решении поисковой задачи именно каталог, а не поисковая машина оказываются более приемлемыми для начала поиска.

Поисковые интернет-каталоги можно найти на любом общедоступном поисковом сервисе.

### *Автоматические индексы*

Специальные программы-роботы (Spider, Robot или Bot, известные также как «пауки») в автоматическом режиме периодически обследуют Интернет на основе определенных алгоритмов, проводя индексацию найденных документов.

Индекс – это хранилище данных, в котором сосредоточены копии всех посещённых роботами страниц. Созданные индексные базы данных используются поисковыми машинами для предоставления пользователю доступа к размещенной на узлах Сети информации. Индексы периодически обновляются и дополняются, поэтому результаты работы одной поисковой машины с одним и тем же запросом могут различаться, если поиск производился в разное время.

Пользователь в рамках интерфейса, выбранной поисковой системы, формулирует запрос. Далее запрос обрабатывается программой обработки запроса, которая просматривает индекс на предмет наличия нужной информации и возвращает ссылки на найденные документы. В окне браузера отображаются результаты обработки запроса.

В Интернете существует большое число поисковых систем. Как правило, при поиске можно использовать не одну систему, а несколько различных. В этом случае вероятность того, что искомая информация будет найдена, выше, поскольку разные системы используют разные алгоритмы поиска. Ниже приведен список некоторых из них:

<http://www.yandex.ru>

<http://www.list.ru>

<http://www.google.ru>

<http://www.rambler.ru>

<http://www.aport.ru>

<http://www.metabot.ru>

Успех поиска в таких системах в значительной степени зависит от формулировки запроса. Рассмотрим основные моменты, которые помогут вам правильно сформулировать запрос к поисковой машине:

- не задавайте только одно слово. Используйте целые фразы, или, по крайней мере, несколько слов;
- указывайте слова, которые не должны встречаться в искомых документах. Обычно для этого используют либо знак «-», либо ключевое слово NOT;
- начинайте имена собственные: названия компаний, городов, фамилии людей, и др. с заглавных букв, а все остальные слова пишите только маленькими буквами;
- если вы хотите найти фразу целиком - заключите ее в кавычки;

– если вы вводите запрос к поисковой машине, состоящий из нескольких слов, то в результате получаете список документов, в которых встречается хотя бы одно слово.

Поисковая машина обычно производит сортировку найденных документов по принципу релевантности.

Рассмотрим, что такое **релевантность**. При индексации документов поисковые машины высчитывают так называемый «вес» слова на странице - соотношение количества повторов на странице заданного Вами слова к общему количеству слов на странице документа. Если заданный запрос, состоит из нескольких слов, то более релевантными будут документы, в которых совокупный вес слов будет максимальный. Однако, при подсчете веса не учитывается, рядом или отдельно стоят данные слова, и поэтому нет гарантий, что в первых документах содержится максимальное количество повторений словосочетания. Вполне возможно, что такого словосочетания там вообще не будет.

Поэтому, если Вы хотите найти заданное словосочетание - задавайте запрос в окне поисковой машины в кавычках. В этом случае будет высчитываться вес словосочетания целиком. Соответственно, гарантируется наличие именно данного словосочетания в найденных документах.

В заключение рассмотрим основные этапы поиска информации в сети Интернет.

№	Этап	Содержание работ этапа
1	Определение предмета поиска	На этом этапе определяем, что конкретно нас интересует.
2	Составление списка ключевых слов	На этом этапе выявляем, как может называться то, что нас интересует.
3	Выбор информационного пространства	На этом этапе определяем, где может находиться то, что нас интересует.
4	Определение инструмента для поиска	На этом этапе принимаем решение о том, как проще и быстрее найти то, что нас интересует.
5	Предварительный поиск	Пробуем найти.
6	Анализ полученной информации	Смотрим на полученные результаты. Если это необходимо (в том случае, когда полученные результаты нас не устраивают), проводим корректировку всех предыдущих действий.
7	Дополнительный поиск	Ищем дальше, пока не получаем ответ на свой вопрос.

### **Рекомендации по поиску информации:**

– Потратьте несколько лишних минут, чтобы максимально «сузить» описание предмета поиска - это поможет сэкономить Вам много времени и денег;

- Искать что-то конкретное лучше всего с помощью поисковых машин, так как если вы знаете «хорошие» ключевые слова, четко определяющие то, что вы хотите найти, то и поиск не представляет никакого труда;
- Используйте для поиска нескольких поисковых машин;
- Если на просматриваемой Вами странице существует несколько заинтересовавших Вас ссылок - открывайте несколько окон, пока Вы читаете информацию на одной странице, остальные успешно (или не очень) успевают загрузиться;
- Если вы хотите найти популярные, часто посещаемые ресурсы - ищите с помощью рейтинга;
- Если нужна очень редкая информация - попробуйте найти ресурсы, посвященные более общей теме. Возможно, там будут размещены ссылки на необходимые Вам ресурсы или будут опубликованы требующиеся материалы;
- Старайтесь найти ответ, а не задать вопрос;
- Создавайте свою коллекцию интересующих Вас ссылок;
- Если Вы нашли что-то интересное, сразу запишите адрес ресурса в «Избранное» либо в текстовый файл;
- Записывая адреса ресурсов в «Избранное», старайтесь дать им более четкое название, максимально соответствующее содержащейся в них информации, постарайтесь не использовать слишком длинные названия;
- Разработайте свой собственный классификатор и для каждой темы создавайте отдельную папку в «Избранном». Поверьте, поиск в своих собственных неразобранных архивах занимает не меньше времени, чем поиск в Internet.

## Ход работы

### Задание 1.

Поместите найденные страницы в список избранных.

Для этого

1. загрузите страницу в браузер;
2. в меню **Избранное** выберите пункт **Добавить в избранное...**
3. задайте имя ссылки в появившемся окне или оставьте то, которое предлагается по умолчанию.
4. проверьте сохранённую ссылку. Откройте новое окно браузера. В меню **Избранное** найдите созданную ссылку и щелкните по ней, чтобы перейти на указанную страницу.

### Задание 2.

Найдите дополнительную информацию, полезную в работе над проектом. Для поиска воспользуйтесь предметными каталогами или поисковыми системами.

1. Запустите браузер и загрузите в нем сайт одной из поисковых систем, например, <http://www.yandex.ru>.
2. В строке запроса введите ключевые слова для поиска и нажмите кнопку **Найти**.
3. Просмотрите список найденных ссылок и выберите те, которые могут быть полезны для Вашего проекта (щёлкните по выбранной ссылке и просмотрите ее содержание).
4. Сохраните страницы в списке избранных.

На основе поисковых двигателей нового поколения можно строить свои собственные индивидуальные или коллективные поисковые машины. Поиск можно адаптировать к определенной тематике и к определенному сообществу. Пример организации социального поиска - персональная или групповая поисковая машина от Google находится по адресу <http://www.google.com/coop/cse/>

Создание собственного поисковика в **Google CSE** начинается с определения списка сайтов, которые пользователь считает объективными и актуальными источниками информации в различных сферах деятельности. К работе по изменению созданной поисковой системы можно подключить других пользователей, которые могут стать соавторами и настраивать ее свойства коллективно. Домашняя страница созданной поисковой машины, помимо строки поиска, будет включать профиль создателя и список заданных сайтов.

### **Задание 3.**

Создайте систему пользовательского поиска в **Google CSE**. Определите список сайтов, которые будут включены в систему пользовательского поиска. Перейдите на страницу системы персонального поиска **Google CSE** - <http://www.google.com/coop/cse/> .

Нажмите кнопку *Создать систему персонального поиска*. Дайте название системе. Опишите систему (для чего она создается, что можно будет находить при помощи этой поисковой машины, среди каких сайтов будет вестись поиск и т.п.). Поищите ключевые слова для поиска системы в сети.

Добавлять сайты к системе можно на странице управление. Гораздо более удобный способ – воспользоваться расширением **Google-marker** - <http://www.google.com/coop/cse/marker>

После того как вы добавите расширение к панели своего браузера, добавлять нужные сайты к поисковой системе станет значительно проще. Достаточно щёлкнуть по иконке маркера и система пользовательского поиска предложит вам добавить аннотацию сайта или страницы.

## Панель управления: Поиск по образовательным Вики

[Основные сведения](#) | [Сайты](#) | [Уточнения](#) | [Внешний вид](#) | [Код](#) | [Совместная работа](#) | [Заработайте!](#) | [Версия для компаний](#)

### Основные сведения

Название и описание будут показываться на [домашней странице](#) Вашей поисковой машины в Google.

Название системы поиска:

Описание системы поиска:

Ключевые слова, которые описывают тему или содержимое Вашей поисковой машины. Эти ключевые слова используются Вашей поисковой машины. [Узнать больше](#).

Ключевые слова системы поиска:   
например: климат "глобальное потепление" "парниковые газы"

Язык системы поиска:

Кодировка системы поиска:

Рисунок 4. Описание персональной системы поиска

Добавить аннотацию в:  [изменить систему поиска »](#)

Включить  Исключить

Включить:

Найти на этом сайте ([www.ourstories.org](http://www.ourstories.org))

Только эта страница ([www.ourstories.org/ourpartners.html](http://www.ourstories.org/ourpartners.html))

Применить ярлыки: (необязательно)

примеры Гугл  младшие школьники  зарубежный опыт  наука

Рисунок 5. Добавление сайта к системе пользовательского поиска

К работе по изменению созданной поисковой системы можно подключить других пользователей. Вы можете разрешить им стать соавторами поисковой системы и настраивать ее свойства коллективно.

## Контрольные вопросы

1. Поисковые системы и их типы.
2. Этапы поиска информации в Интернет.
3. Структура адресов ресурсов сети Интернет.

## Содержание отчета

4. Наименование и цель лабораторной работы
5. Скриншоты выполнения лабораторной работы.
6. Выводы по лабораторной работе.
7. Ответы на контрольные вопросы.