

Лабораторная работа № 1-18. Поиск информации в сети Интернет

Цель: получение практических навыков по работе с основными современными сервисами сети Интернет. Приобретение практических навыков составления поисковых запросов.

Задачи:

1. Изучить принципы организации поиска в сети Internet.
2. Изучить правила составления поисковых запросов.
3. Ознакомиться с правилами адресации информационных ресурсов и узлов сети Интернет.
4. Научиться свободно ориентироваться в информационном пространстве сети Интернет.

Теория

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.

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

Одним из наиболее распространенных сервисов сети Интернет является WWW (World Wide Web), или Всемирная паутина, – распределенная информационная среда, содержащая разнообразную информацию (в том числе мультимедиа-компоненты веб-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 ее 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>



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](http://www.google.ru) – протокол передачи данных); правая – на многоуровневое доменное имя компьютера, в котором домены отделяются друг от друга точками:

[www](http://www.google.ru) – Всемирная паутина (сервисная служба Интернета),
[google](http://www.google.ru) – имя сервера,
[ru](http://www.google.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 Интернета 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.

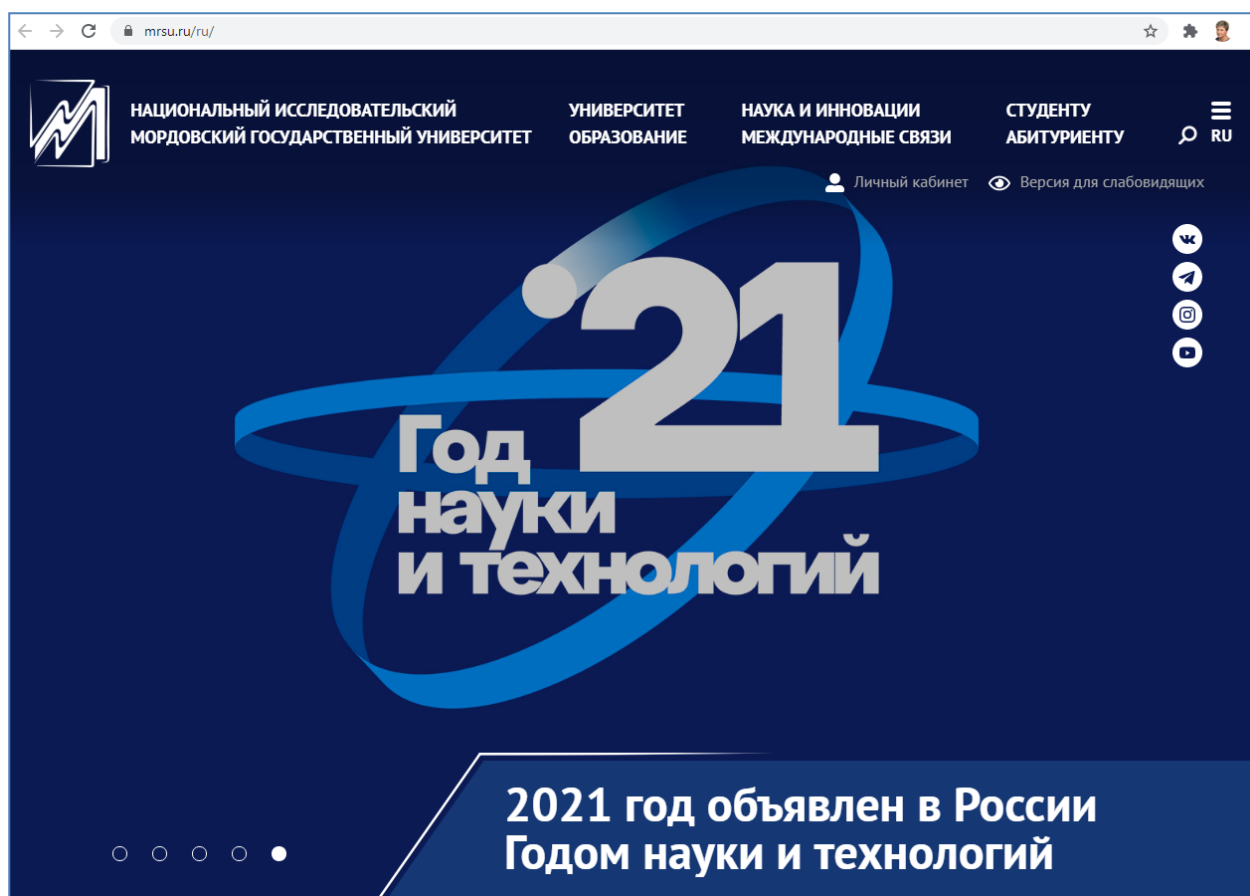
Поиск информации в сети Интернет может осуществляться двумя способами:

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

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

Для поиска веб-ресурса по известному URL-адресу необходимо запустить программу-браузер, например Яндекс, и в адресной строке ввести этот адрес.

- запустите браузер Internet Explorer (Пуск – Internet Explorer);
- введите в адресную строку следующий адрес: <http://www.mrsu.ru>;
- нажмите Enter;
- дождитесь, пока страница загрузится.



Остановимся более подробно на вопросе выбора поисковых инструментов WWW. В Интернете 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 Интернета 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 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 информация.

Как показывает практика, при решении поисковой задачи именно каталог, а не поисковая машина, оказывается более приемлемым 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 посещенных 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. Наиболее популярные поисковые системы

Название поисковой системы	Адрес
Яндекс (русскоязычная)	http://www.yandex.ru
Рамблер (русскоязычная)	http://www.rambler.ru
Google (русскоязычная)	http://www.google.ru
Апорт (русскоязычная)	http://www.aport.ru
Yahoo! (англоязычная)	http://www.yahoo.com
AltaVista (англоязычная)	http://www.altavista.com
	http://www.ra-gu.net

Поисковая система реализована в виде веб-страницы с обычным адресом, которая содержит так называемую строку для поиска и кнопку «Поиск» (Search), а также может содержать тематический каталог ресурсов, ссылки на популярные страницы и т.п.

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.

К средствам поисковых систем относится **язык запросов**. Используя различные приемы, можно добиться желаемого результата поиска:

! – запрет перебора всех словоформ;

+ – обязательное присутствие слов в найденных документах;

- – исключение слова из результатов поиска;

[] – сохранение порядка слов;

* – наличие звездочки в запросе свидетельствует о пропущенном слове, которое необходимо найти поисковику;

| – поиск любого из данных слов;

«» – поиск устойчивых словосочетаний;

mime – фильтрует выдачу по формату найденных документов;

date – поиск информации по указанной дате. Причем искать можно материалы, опубликованные как в конкретный день/месяц/год, так и за определенный интервал времени:

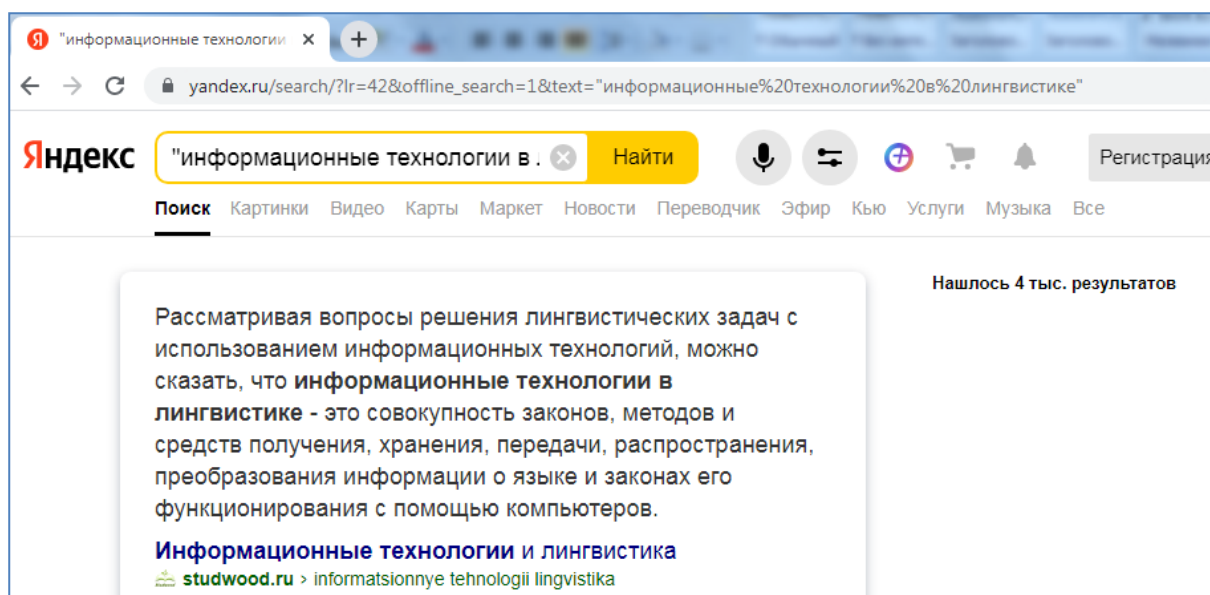
– date:ГГГГММДД – в указанный день;

– date:ГГГГММ* – месяц;

– date:ГГГГ* – год;

– date:ГГГГММДД..ГГГГММДД – в промежутке;

– date:>ГГГГММДД – после (или до) указанной даты (можно использовать: >, >=, <, <=).



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 слова к общему количеству слов на странице документа. Если задается запрос, состоящий из нескольких слов, то более релевантными будут документы, в которых совокупный вес слов будет максимальный. Однако при подсчете веса не учитывается, рядом или раздельно 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 необходимо 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	Определяем, что конкретно нас интересует
2	Составление списка ключевых слов	Выявляем, как может называться то, что нас интересует
3	Выбор информационного пространства	Определяем, где может находиться то, что нас интересует
4	Определение инструмента для поиска	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

Общие советы

- Потратьте несколько лишних минут, чтобы максимально «сузить» описание предмета поиска – это поможет сэкономить много времени;
- Используйте для поиска нескольких поисковых машин;
- Если на просматриваемой странице существует несколько заинтересовавших ссылок – открывайте несколько окон, пока читаете информацию на одной странице, остальные успешно (или не очень) успевают загрузиться;
- Если нужно 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 ресурсы или будут опубликованы требуемые материалы;
- Старайтесь найти ответ, а не задать вопрос;
- Создавайте свою коллекцию интересующих ссылок;

- Если 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».

Ход выполнения работы

Отчет по работе оформляется в виде документа Word и должен содержать: титульный лист, содержание работы, результаты выполнения заданий.

Задание 1. С помощью расширенного поиска на Яндексе сравните популярность следующих сайтов по количеству страниц, ссылающихся на них: Президента и Правительства Российской Федерации; Московского государственного университета и Мордовского государственного университета; Эрмитажа и Лувра. Сохраните найденные Web-страницы в отдельной папке. На отдельной странице отчета зафиксируйте количество ссылок на каждую из них (можно в виде таблицы).

Задание 2. Найдите информацию о том, когда и где родился ваш любимый писатель (поэт).

Составьте список его произведений. Найдите его фотографии в разные годы жизни. Сохраните всю информацию на отдельном листе отчета.

Задание 3. С помощью одной из поисковых систем найдите информацию о выдающихся людях XX века и занесите ее в таблицу, например:

Личности XX века		
Фамилия, имя	Годы жизни	Род занятий
Джеф Раскин	9 марта 1943 – 26 февраля 2005	Специалист по компьютерным интерфейсам
Лев Ландау	9 (22) января 1908 – 1 апреля 1968	Советский физик-теоретик
Юрий Гагарин	9 марта 1934 – 27 марта 1968	Летчик-космонавт СССР № 1, Герой Советского Союза, кавалер высших знаков отличия ряда государств, почетный гражданин многих российских и зарубежных городов. Полковник ВВС

Задание 4. Заполните таблицу, используя поисковую систему Яндекс.

Слова, входящие в запрос	Структура запроса	Количество найденных страниц	Адрес первой ссылки
Информационная система	Информационная +Система		
	Информационная –Система		
	«Информационная система»		
Персональный компьютер	Персональный компьютер		
	Персональный !компьютер		
	[Персональный компьютер]		
Информационные технологии	Информационные технологии		
	«Информационные» *		
	Информационные технологии mime:docx		
Компьютерная лингвистика	Компьютерная лингвистика		
	Компьютерная лингвистика		
	Компьютерная лингвистика date:>20190101		

Задание 5. Создайте аннотированный список интернет-ресурсов по теме, соответствующей вашему варианту:

1. Информационные технологии в лингвистике.
2. Прикладные программные средства для лингвистов. Обзор.
3. Текстовый процессор MS Word, его назначение, структура, основные возможности, приемы работы с текстами.
4. Сбор и обработка лингвостатистических данных средствами MS Word.
5. Специальные возможности программы MS Word для лингвистов.
6. Табличный процессор MS Excel, его назначение и основные возможности.
7. Сбор и обработка лингвостатистических данных средствами MS Excel.
8. Основы создания презентаций в MS Power Point.
9. Правильное использование заимствованных терминов и обозначений компьютерной лингвистики.
10. Особенности электронных переводческих словарей *Lingvo* и *Multitran* и их отличия от онлайн-переводчиков (*Google*, *Yandex* и т. п.).
11. Сравнение программ переводческой памяти.
12. Сравнение программ автоматического перевода.
13. Краудсорсинг, или модель «Википедии» в переводе.
14. Сравнение мультимедийных программ по обучению иностранным языкам (*English DeLuxe*, «*РЕПЕТИТОР English*» и т. п.).
15. Технология подкастинга в обучении языкам.
16. ВебКвесты в обучении языкам.
17. Возможности электронного письма в обучении языкам.
18. Сетевые формы коммуникации и их влияние на язык.
19. Ресурсы Всемирной паутины для обучения языкам.
20. Сравнительный анализ составления поисковых запросов в популярных русскоязычных поисковых системах (*Google*, *Yandex*, *Rambler*, *Mail.ru*, *Altavista*, *Yahoo*, *MSN*, *AOL*).

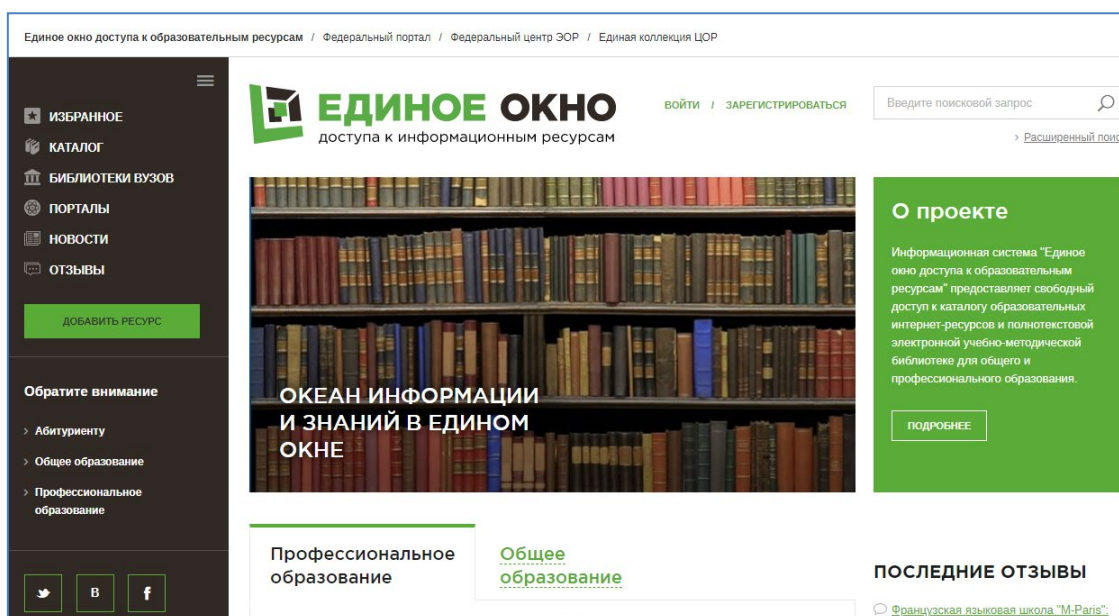
Аннотированный список – это список ресурсов, содержащий дополнительную информацию, которая позволяет дать предварительную оценку ценности ресурса. Подобный список обязательно сопровождается аннотацией ресурса.

В аннотированном списке должны быть представлены 10 интернет-ресурсов. В первую очередь электронные библиотеки, музеи информатики и вычислительной техники, персональные сайты различных авторов. Также допустимы ссылки на тематические социальные сети, блоги, твиттеры и другие ресурсы, содержимое которых соответствует выбранной теме.

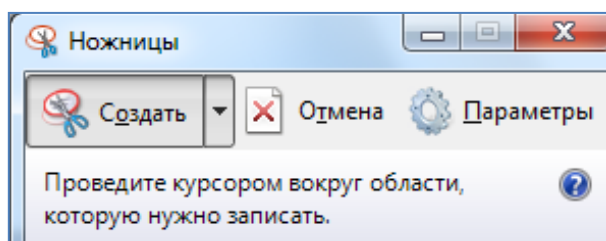
Информацию об авторе обычно можно найти в разделах **О сайте, О проекте, Об авторе, Контактная информация, Контакты** или внизу страницы © **Copyright**.

В аннотации укажите, чем данный ресурс может быть полезен для студентов и будущих специалистов. Также можно сказать о каких-либо важных особенностях ресурса, которые выделяют его среди остальных. Аннотация должна состоять из 3–5 предложений, *не являясь при этом скопированным текстом из раздела О проекте!!!*

№	Адрес ресурса	Автор ресурса	Аннотация
1	http://window.edu.ru	ФГАУ ГНИИ ИТТ "Информика"	Электронная библиотека цифровых образовательных интернет-ресурсов образовательного назначения. Данный сайт полезен для работы над заданием. В библиотеке и каталоге работает полнотекстовый поиск. Все материалы доступны в удобных форматах PDF и DJVU



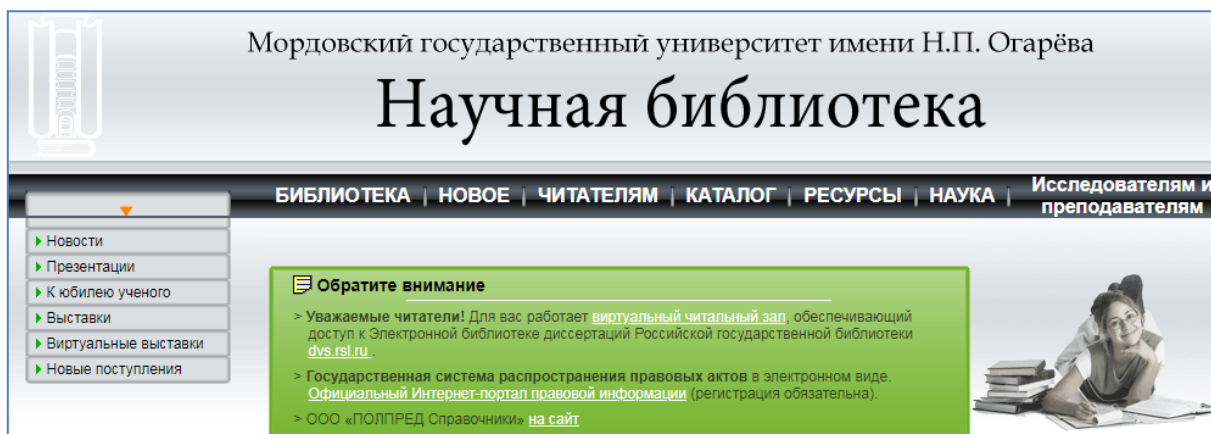
Информацию следует оформить в виде таблицы, сопровождающейся снимком экрана. В снимок экрана должна попасть только главная страница сайта, возможно с интерфейсом браузера. Сделать это можно с помощью компонента «Ножницы». Для запуска компонента нужно нажать кнопку **Пуск**, в поле поиска набрать **Ножницы** и кликнуть на соответствующий пункт.



Задание 6. Поиск информации в локальных поисковых системах

Поскольку знакомство с темой, указанной в задании 3, проще всего начать с учебных пособий, подберите в университетской библиотеке 10 учебных пособий, изданных за последние 5 лет.

Используйте электронный каталог Научной библиотеки им. М. М. Бахтина, на сайт которой можно перейти из личного кабинета ЭИОС.



В отчете опишите проделанные вами действия, для иллюстрации используйте Screenshot (моментальные снимки экрана). Обратите внимание, что Screenshot можно создать с помощью кнопки PrtScr на клавиатуре компьютера – в этом случае копируется все содержимое экрана. Комбинация клавиш Alt+PrtScr копирует только активное окно.

Задание 7. Поиск информации с помощью метапоисковых систем

Посмотрите, что представлено в Интернете по вашей тематике, используйте одну из метапоисковых систем, рассмотренных ранее. В отчете опишите свои действия и сделайте вывод по результатам поиска:

- какая информация встречается по заданной теме, много ли ее;
- насколько полезна информация, можно ли что-то использовать, например, в курсовой работе по заданной теме (5–6 предложений).

Задание 8. Расширенный поиск

Откройте поисковую систему Yandex и воспользуйтесь инструментами расширенного поиска. Помните, что информация должна быть достоверной (известен автор, издатель, авторитетный сайт) и актуальной (примерно за последние 5 лет).

В отчете опишите свои действия с использованием иллюстраций и сделайте выводы: смогли ли вы улучшить результаты поиска за счет

использования настроек расширенного поиска, какие настройки оказались наиболее полезными.

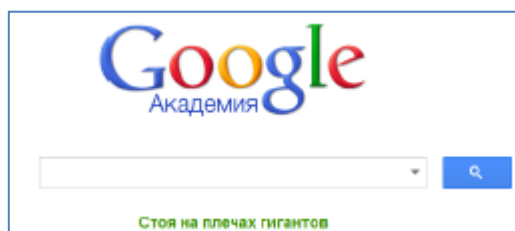
Представьте список из 5–7 источников, которые можно использовать для работы с заданной темой в дальнейшем.

Задание 9. Поиск книг в Google

Ознакомьтесь с каталогом книг Google <http://books.google.ru>.

Найдите несколько книг по вашей тематике. В отчете опишите свои действия, используйте иллюстрации. Сделайте вывод о полезности данного ресурса.

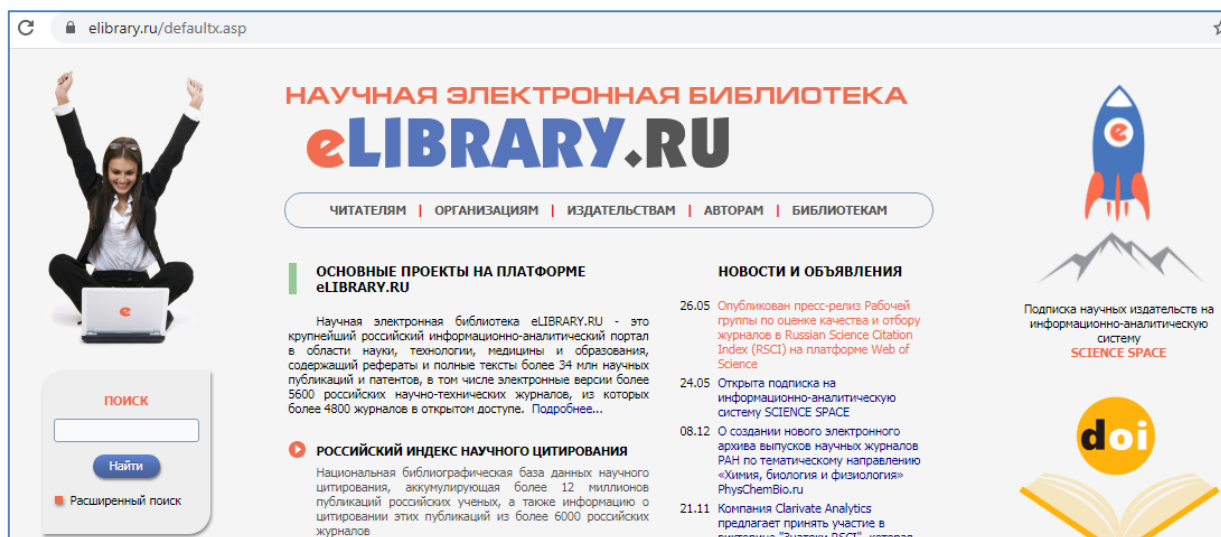
Откройте сервис Google Академия <http://scholar.google.ru> и подберите информацию по вашей теме.



В отчете опишите свои действия, напишите, в чем особенность данного ресурса. Представьте список из 5–7 источников, которые вы могли бы использовать при работе над вашей темой.

Задание 10. Ресурсы научной электронной библиотеки

Научная электронная библиотека elibrary.ru позволяет подбирать и просматривать статьи в ведущих российских журналах.



Перед началом работы необходимо авторизоваться в библиотеке. Если вы еще не зарегистрированы, то зарегистрируйтесь.

Подберите статьи по вашей тематике, которые для вас доступны.

В отчете опишите свои действия с иллюстрациями, представьте список из 10 статей, которые вы могли бы использовать в работе над своей темой, укажите способ доступа к полному тексту статьи (доступна pdf версия статьи на сайте библиотеки или есть бумажная версия в читальном зале библиотеки университета). Статьи должны быть за последние 5 лет, не старше.

Сформируйте общий список литературы и оформите его в соответствии со Стандартом организации СТО МордГУ006-2020 «Общие требования к построению, изложению и оформлению документов учебной деятельности».

Контрольные вопросы

1. Назовите средства поиска информации в сети Интернет.
2. Охарактеризуйте простые приемы поиска информации в сети Интернет.
3. Перечислите поисковые системы. Как их классифицировать?
4. Каковы принципы работы поисковых систем сети Интернет?
5. Сформулируйте основные правила составления поисковых запросов.

6. Что обозначают символы «+», «-», «|» в индексном поиске и как правильно их записать в вопросе?
7. Какие из поисковых систем имеют тематический каталог ресурсов?
8. Какие из поисковых систем имеют возможности поиска по различным категориям информационных ресурсов?
9. Назовите адреса наиболее популярных русскоязычных поисковых систем.
10. Как сохранить информацию, найденную в Интернете, на своем рабочем диске?
11. Как сохранить рисунок на своем диске? Какие типы графических файлов используются в Интернете?
12. Какие возможности для поиска предлагает расширенный поиск?
13. Как определить, какой вид поиска лучше использовать?