

Тема. Поиск информации в Интернет (2 часа)

Цель работы: изучение принципов организации поиска в сети Internet и приобретение практических навыков составления поисковых запросов.

Порядок выполнения работы:

1. Создайте копию этого файла и сохраните под именем: Лаб4 Фамилия студента № группы.
2. Вставьте в файл верхний колонтитул и добавьте в него информацию следующего содержания: «Лаб4 Фамилия студента № группы».
3. Ознакомьтесь с теоретическим материалом.
4. Выполните «задание», заполнив соответствующие таблицы и ответив на вопросы.
5. Отчет предоставить в Moodle в формате Word.

Теоретический материал

Системы поиска информации создаются с учетом особенностей информационных ресурсов и типов данных. Это позволяет разрабатывать адекватные алгоритмы поиска и повышать их эффективность.

Классификация сетевых информационных ресурсов зависит от выбранного признака классификации.

1. По способу представления информации:

- **Web-страницы** (технология WWW);
- **Gopher** (другой гипертекстовый стандарт, практически не развивающимся сегодня, т.к. поддерживает только текстовую форму представления информации);
- **базы данных;**
- **файловые серверы** (FTP – серверы);
- **телеконференции** (делятся на модерлируемые (управляемые - размещение статей производится специальным человеком - модератором, осуществляющим цензуру, и не модерлируемые, размещение статей в которых производится автоматически по запросу любого пользователя Сети).

2. По национально-территориальному признаку:

- **по языковому признаку:**

- основной язык в сети Интернет - английский,
- все основные языки мира также представлены в Сети,
- некоторые сайты поддерживают несколько языков

- **по географическому признаку**

- принадлежность какой-либо организации, осуществляющей свою деятельность на определенной территории,
- направленность на аудиторию в пределах другого региона

3. По виду и характеру информации (content):

- **тематическая информация** - наиболее ценный тип информации, относящейся к конкретной предметной области: техническая, технологическая, маркетинговая и др.;

- **научные публикации** - статьи, рефераты, обзоры и пр. публикации научного характера;

- **рекламная информация**

содержат информацию о фирме - профиль компании (company profile), о предлагаемых ею товарах или услугах, о занимающих ключевые посты людях; через сайт может осуществляться часть работы по поддержке клиентов, FAQ и т.п.;

- **справочная информация** - справочные материалы, ссылки на Web-сайты компаний, нормативную базу и т.д.;

- **новости** - информация легко доступна, представляет собой вид "сырой", необработанной информации, которая ценна не столько сама по себе, сколько в контексте прочих событий или в динамике развития, и часто нуждается в последующей обработке;

- **вторичная информация** - систематизированная и предварительно обработанную информацию, обладает значительной ценностью: специализированные тематические сайты, обзоры, подборки рефератов, каталоги и др.

4. По принципу организации и использования средства поиска:

- **каталоги** - справочники, директории - содержат списки адресов Интернет, сгруппированные по определенным признакам, объединяются по

тематике: наука, искусство, новости и т.д. Особенность - создание структуры, базы данных и их обновление осуществляется "вручную", коллективом редакторов и программистов, и процесс поиска требует непосредственного участия пользователя, самостоятельно переходящего по ссылкам.

• **поисковые машины** - постоянно исследуют узлы Интернет, доступные данной системе поиска, со всеми их связями. Осуществляют - постоянное обновление информации - машина поиска регулярно возвращается через определенный срок к уже изученным узлам, чтобы обнаружить и зарегистрировать изменения. Вся прочитанная информация индексируется, т.е. создается специализированная база данных, в которой закодированы все исследованные системой страницы Интернет. При поступлении запроса от пользователя машина поиска рассматривает всю индексируемую информацию и выдает список документов, соответствующих задаче поиска. Найденные документы ранжируются в зависимости от местоположения ключевых слов (в заголовке, в начале текста, в первых параграфах) и частоты их появления в тексте. Несмотря на схожий принцип работы, машины поиска различаются по языкам запроса, зонам поиска, глубине поиска внутри документа, методам ранжирования и приоритетов, поэтому применение разных поисковых машин дает различные результаты.

Для повышения **релевантности** поиска информации в Интернет применяют различные приемы для формирования поисковых запросов.

Найденный по запросу документ, содержащий нужную (искомую) информацию, называется **релевантным** (*англ. relevant - "относящийся к делу"*), в другом случае **нерелевантным** (*шумовым*)

Примеры параметров и операторов языка поисковых запросов приведены в табл.4.

Таблица 4

Параметры и операторы языка поисковых запросов

+	Слово обязательно должно присутствовать на странице
-	Слово обязательно должно отсутствовать на странице
;	Поиск фразы, состоящей из слов, разделенных запятой

""	Контекстный поиск
*	Может располагаться в любой части слова и заменять цепочку любых символов
AND (&)	Оба фрагмента должны находиться на странице
OR()	Хотя бы один фрагмент должен находится на странице
NOT(!)	Указанный фрагмент не должен находится на странице
NEAR(~)	Заданные фрагменты должны находится на расстоянии не более 10 слов
()	Объединение слов в выражение

Задание

1. С помощью поисковой системы Yandex (www.yandex.ru) проведите поиск информации по теме «Приложения Интернет в бизнесе и экономике».

Для повышения эффективности поиска заранее составьте тезаурус (набор ключевых слов) по теме и используйте его в поисковых запросах.

Ключевыми словами могут служить, например, слова «Интернет», «экономика», «бизнес», «коммерция», «реклама», «маркетинг», «прибыль», «деньги» и многое другое.

При составлении поисковых запросов проработайте разные варианты поиска, например:

- ключевые слова в различных комбинациях;
- изменение регистра;
- контекстный поиск (например, «Как заработать деньги в Интернет?»);
- поиск фразы (для этого фразу надо заключить в “ “);
- специальный язык запросов, т.е. знаки «+» и «-«, *, &? &&, |, (), AND, OR, NOT;

2. Составить 5-7 поисковых запросов, по каждому из которых провести поиск с помощью поисковой системы Yandex. Посмотреть несколько документов (сайтов) из предложенного Yandex списка сайтов, сделать вывод об их релевантности и о качестве поиска в целом. Понятие «**релевантность**» в данном контексте трактуется как субъективная оценка пользователем

соответствия полученного от поисковой системы ответа на запрос и содержанием самого запроса. В качестве ответа поисковой системы в данном случае рассматривается список сайтов, где содержится искомая информация. Сделать оценку релевантности ответа поисковой системы можно просмотрев несколько сайтов из предложенного списка. Если содержимое сайтов совпадает с вашими ожиданиями, т.е. содержит нужную информацию, то считается, что релевантность ответа поисковой системы высокая.

Результат работы представьте в виде таблицы Word. (**Внимание:** запросы в нижеследующей таблице 1 приведены **исключительно в качестве примера** и не должны копироваться!).

Таблица 1

Примеры запросов и результаты поиска (www.yandex.ru)

Поисковый запрос	Кол-во найденных документов	Комментарий о релевантности (высокая, средняя, низкая)
Интернет экономика бизнес	108923	низкая
Интернет, доход, прибыль, деньги	211098	низкая
Способен ли Интернет приносить прибыль?	7982	средняя
+Интернет &&/1 коммерция	210	высокая
(Интернет Интернет & бизнес) && доход	462	высокая

По аналогичным *поисковым* запросам проведите поиск информации с помощью поисковых систем Google (www.google.com), Rambler (www.rambler.ru). Результаты поиска представить в виде аналогичных таблиц. Сравнить результаты. Сделать выводы. Результаты сохранить в виде документа MS Word.

Примеры запросов и результаты поиска (www.google.com)

Интернет экономика бизнес	33 100 000	Высокая
Интернет, доход, прибыль, деньги	5 580 000	Средняя
Способен ли Интернет приносить прибыль?	390 000	Средняя
+Интернет коммерция &&/1	8	Низкая
(Интернет Интернет & бизнес) && доход	30 200	Средняя

Примеры запросов и результаты поиска (www.rambler.ru)

Интернет экономика бизнес	7000000	Высокая
Интернет, доход, прибыль, деньги	8000000	Высокая
Способен ли Интернет приносить прибыль?	10000000	Низкая
+Интернет коммерция &&/1	3000000	Средняя
(Интернет Интернет & бизнес) && доход	6000000	Низкая

Вывод: Лучший поисковик Гугл, потому что у него более чёткий поиск по запросу

С помощью любой поисковой системы проведите поиск информации по теме «Экономические приложения Интернет» по соответствующим рубрикам. Результат работы представьте в виде таблицы Word (не менее 7 строк) (см. Таблица 2). При выполнении данного пункта проявите свои творческие способности!

Таблица 2

Результаты поиска и уровень релевантности

Название найденного документа и его URL-адрес	Комментарий о релевантности (высокая, средняя, низкая)
https://habr.com/ru/post/295760/	Средняя
https://scienceforum.ru/	Средняя
https://studme.org/	Высокая
https://pf.hse.ru/582200445.html	Средняя
https://raec.ru/	Средняя
https://yvision.kz/	Низкая

Сравните результаты. Сделайте выводы.

Вывод: В гугле большая вариативность подбора ресурсов чем у других поисковиков

3. Сравнительный анализ работы поисковых систем, Yandex, Rambler, Google.

Оцените широту охвата поисковых систем. Придумайте несколько слов из различных областей и с различной частотой употребления (очень частотные, менее частотные, редкие). Для каждой поисковой системы задайте поисковый запрос из одного слова (чтобы исключить влияние языков

запроса) и зафиксируйте статистику результатов поиска в таблице 3 (не менее 7 строк).

Таблица 3

Оценка широты охвата поисковых систем

Слово	Количество найденных документов		
	Yandex	Rambler	Google
СВО	58 000 000	60 000 000	63 000 000
Футбол	32 000 000	26 000 000	68 400 000
Образование	45 000 000	19 000 000	334 000 000
Интеграция	10 000 00	6 000 000	19 200 000
Бакалавриат	2 500 000	1 000 000	12 900 000
Паросмия	12700	5000	86 800
Гиперболизация	14000	4000	66400

Сравните результаты. Сделайте выводы.

Вывод: Гугл лучший поисковик

Оцените качество сервиса поисковых систем: гибкость языка запросов, возможности ранжирования найденных документов, поиск похожих документов и др.

Рекомендации по поиску.

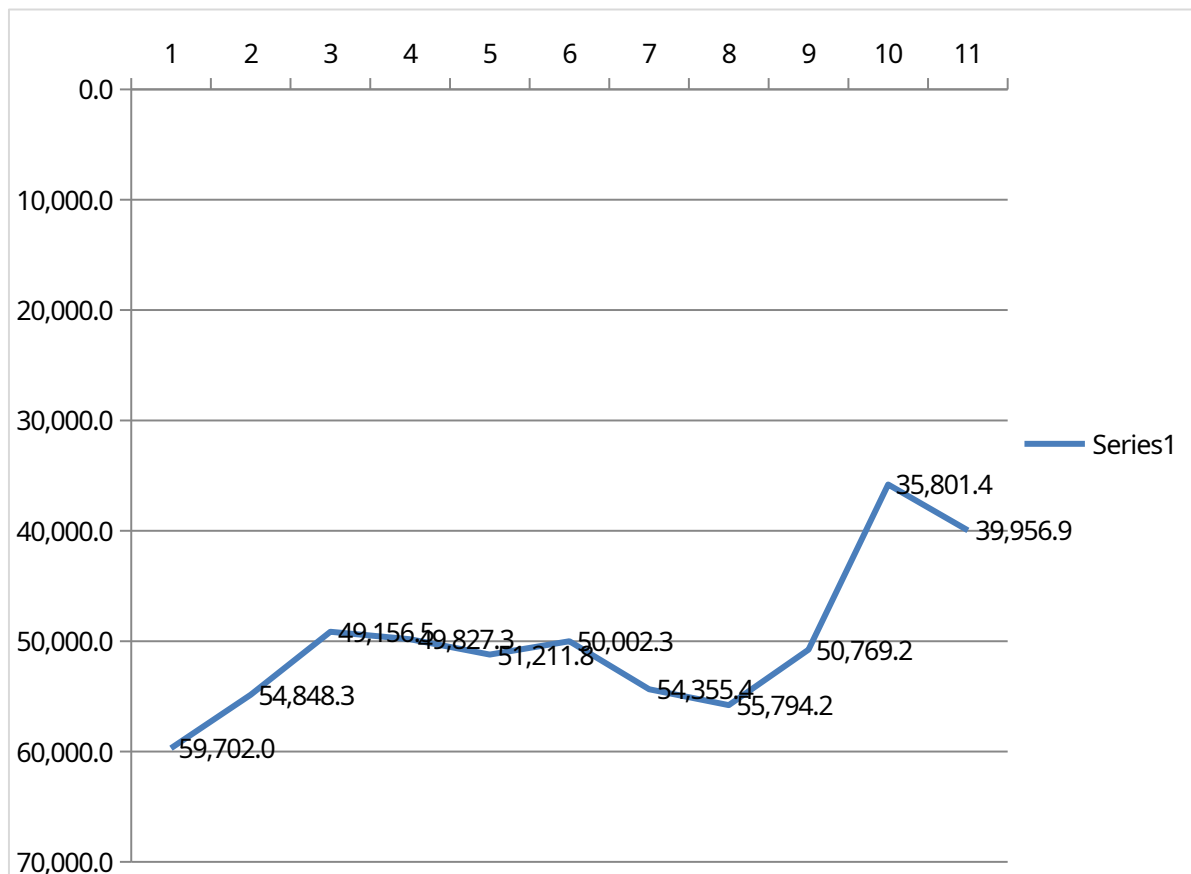
- Проверяйте орфографию.
- Используйте синонимы, например, *фотографии|фото|фотоснимки| снимки.*
- Составляйте запрос более, чем по одному слову.

- Используйте знаки "+" и "-". Между словами и знаками "+" "-" не должно быть пробела. Знак "-" (не тире и не дефис) надо писать через пробел от предыдущего и слитно с последующим словом.
- Используйте контекстный поиск для поиска цитат.
- Используйте операцию «Найти похожие документы», если запрос задан в абстрактной форме и описывает интересующую область;
- содержание документа конкретизирует запрос в интересующем аспекте;
- остальные найденные документы менее соответствуют запросу.

С помощью одной из поисковых систем поисковых систем: www.google.com, www.yandex.ru, www.rambler.ru, найдите официальную информацию о состоянии внешнего долга Российской Федерации за 2010 - 2021 гг. Подготовьте справку по данной теме.

59 702,0 экв. млн. долл. США

Отобразите с помощью мастера диаграмм Excel динамику внешнего долга Российской Федерации по годам 2010-2021 и скопируйте диаграмму в свой отчет.



Рекомендации по поиску.

- Варианты запросов по ключевым словам:

внешний долг России;

внешний долг России динамика;

внешний долг России 2010-2021;

- Контекстный поиск:

«Структура государственного внешнего долга по состоянию на 01 января 2022 года»

Найдите с помощью поисковых систем www.google.com и www.yandex.ru адреса сайтов электронных изданий «Экономика и жизнь», «Эксперт Сибирь», «Налоговый вестник», «Экономическая газета». Сравните полученные результаты по релевантности. Результаты представить в таблице. Таблицу разработать самостоятельно.

Ресурс	Гугл	Яндекс
Экономика и жизнь	Высокая	Высокая
Эксперт Сибирь	Средняя	Средняя
Налоговый вестник	Высокая	Высокая
Экономическая газета	Высокая	Высокая

По итогам выполнения лабораторной работы сформулировать **вывод** о преимуществах той или иной поисковой системы (www.yandex.ru, www.rambler.ru, www.google.com), обосновать его, используя соответствующие показатели (релевантность и др.).

Вывод: По моему мнению, на 2023 год, среди поисковых систем лидирует Гугл, т.к. эта поисковая система, лучше фильтрует результаты поисков, в сравнение с Рамблером и Яндексом.