

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Локтионова Оксана Геннадьевна

Должность: проректор по учебной работе

Дата подписания: 23.12.2021 11:30:27

Уникальный программный ключ:

0b817ca911e6668abb13a5d426d79e5f1c11eabbf73e943df4a4851fda56d089

МИНОБРАЗОВАНИЯ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра вычислительной техники



Технологии поиска информации

Методические указания
по выполнению лабораторной работы по дисциплине
«Информатика» для студентов направлений подготовки:
09.03.01 Информатика и вычислительная техника,
11.03.02 Инфокоммуникационные технологии и
системы связи, 10.03.01 Информационная безопасность,
11.03.03 Конструирование и технология электронных средств,
специальности 10.05.02 Информационная безопасность
телекоммуникационных систем

Курск 2021 г.

УДК 681.3

Составители: С.В. Дегтярев, Е.Н. Иванова, Т.Н. Конаныхина

Рецензент

Доцент кафедры программной инженерии,
кандидат технических наук

Ю.А. Халин

Технологии поиска информации: методические указания по выполнению лабораторной работы / Юго-Зап. гос. ун-т; сост.: С.В. Дегтярев, Е.Н. Иванова, Т.Н. Конаныхина. – Курск, 2021. – 15 с.: ил. 3., табл. 1. – Библиограф.: с. 15.

Рассматриваются основные виды поиска информации в сети Интернет, правила формирования запроса в популярных поисковых системах. Теоретический материал сопровождается примерами.

Методические указания соответствуют требованиям программ, утвержденным учебно-методическим объединением по направлениям Информатика и вычислительная техника, Инфокоммуникационные технологии и системы связи, Информационная безопасность, Конструирование и технология электронных средств, специальности Информационная безопасность телекоммуникационных систем.

Предназначены для студентов очной и заочной форм обучения.

Текст печатается в авторской редакции

Подписано в печать . Формат 60x84 1/16.
Усл.печ.л. Уч.-изд.л. . Тираж 20 экз. Заказ . Бесплатно.
Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94.

Цель работы

Получить знания об адресации в сети Интернет, методах поиска информации с использованием информационных технологий; научиться осуществлять выбор методики поиска информации; овладеть навыками использования различных инструментов профессионального общения, предоставляемых глобальной сетью Интернет.

Краткая теоретическая информация

Поиск по адресам URL

Самый быстрый и надежный вид поиска информации в Интернете – поиск по адресам URL (Universal Resources Locator – универсальный указатель ресурса).

Многие из них приводятся в печатных изданиях, специальных справочниках, звучат в эфире популярных радиостанций и с экранов телевизора.

Поклонникам группы «Аквариум» хорошо известен официальный сайт этой группы <https://www.aquarium.ru/ru>

Любители телеканала НТВ без труда найдут его сайт по адресу <https://www.ntv.ru/>.

Для быстрого доступа к вышеназванным ресурсам достаточно запустить программу-браузер, например Chrome, и набрать знакомый адрес URL в строке адреса.

Поисковые системы

В Интернете сосредоточено огромное количество документов. Чтобы облегчить поиск нужной информации, создаются специальные поисковые машины.

Поисковые машины – это автоматические системы, опрашивающие серверы, подключенные к глобальной сети, и сохраняющие в своей базе информацию об имеющихся на серверах данных. Поисковые машины (spiders, crawlers) постоянно исследуют Сеть с целью пополнения своих баз данных документов. Обычно это не требует никаких усилий со стороны человека. Примером может быть поисковая система Altavista (система открыта в декабре 1995. При-

надлежит компании DEC. С 1996 года сотрудничает с Yahoo).

По специальному образом сформулированному запросу поисковые машины предоставляют информацию о том, где можно получить необходимые данные.

Поисковые машины состоят из трех частей: робота, индекса и программы обработки запроса

Робот (Spider, Robot или Bot) – это программа, которая посещает веб-страницы и считывает (полностью или частично) их содержимое. Роботы поисковых систем различаются индивидуальной схемой анализа содержимого веб-страницы.

Индекс – это хранилище данных, в котором сосредоточены копии всех посещенных роботами страниц. Индексы в каждой поисковой системе различаются по объему и способу организации хранимой информации. Базы данных ведущих поисковых машин хранят сведения о десятках миллионов документов, а объемы их индекса составляют сотни гигабайт. 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.

Программа обработки запроса – это программа, которая в соответствии с запросом пользователя «просматривает» индекс на предмет наличия нужной информации и возвращает ссылки на найденные документы. Множество ссылок на выходе системы распределяется программой в порядке убывания релевантности, то есть от наибольшей степени соответствия ссылки запросу к наименьшей.

В настоящее время самыми популярными для российских пользователей Интернета являются две крупные поисковые системы индексного типа: Google, Yandex. Эти системы учитывают грамматические особенности русского языка, поэтому результаты их поиска в русскоязычных ресурсах отличаются более высоким качеством, чем у западных систем.

Согласно статистике поисковой системы Google, запросы, начинающиеся с «Где», «Кто», «Когда» и «Есть», составляют 8% всех поисковых запросов в Google.

Рейтинг поисковых систем среди русскоязычных пользователей представлен в таблице 1.

Таблица 1

Рейтинг поисковых систем среди русскоязычных пользователей

Поисковая система	Web-адрес
Google	https://www.google.ru
Яндекс	https://yandex.ru
Mail.ru	https://mail.ru
Bing	https://www.bing.com
Rambler	https://www.rambler.ru
Yahoo!	https://www.yahoo.com

Поисковые машины различаются по охвату информационных ресурсов:

Поисковые машины общей тематики имеют базу данных по всем направлениям знаний и отличаются обширным индексом и большим объемом накапливаемой информации;

Поисковые машины специального назначения просматривают только сайты по определенной тематике, например музыкальные или музейные.

Основные характеристики поисковых машин:

- объем документов в индексе;
- частота обновления информации;
- информационное пространство, которое охватывает робот поисковой машины, и разнообразие типов документов, о которых собирается информация;
- скорость обработки запроса;
- критерий определения релевантности (соответствия найденного документа поисковому запросу);
- возможность детализации и уточнения запроса.

Поиск по рубриктору поисковой системы

Поисковые каталоги представляют собой систематизированную коллекцию (подборку) ссылок на другие ресурсы Интернета. Ссылки организованы в виде тематического рубрикатора (см. рисунок 1), представляющего собой иерархическую структуру, переме-

щаяся по которой, можно найти нужную информацию.

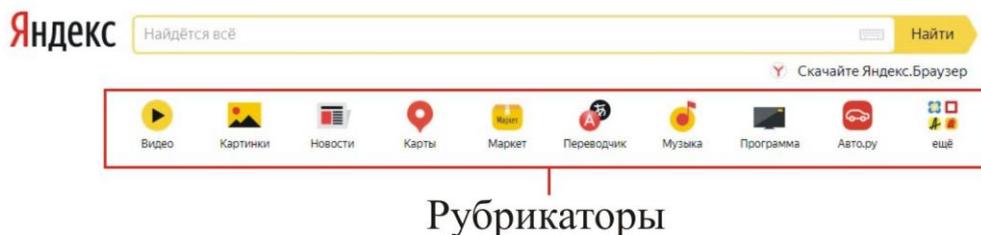


Рисунок 1 – Рубрикаторы Яндекс

Рассмотрим пример. Предположим, вы собираетесь приобрести мобильный телефон и хотите сравнить характеристики аппаратов разных фирм. Поиск мог бы вестись по следующим рубрикам каталога:

Яндекс > Каталог > Маркет > Электроника > Смартфоны и аксессуары > Мобильные телефоны.

Получив ограниченное количество ссылок, можно достаточно оперативно их просмотреть и выбрать телефон, исследовав характеристики по фирмам и модификациям аппаратов.

Поиск по ключевым словам

Большинство поисковых машин имеют возможность поиска по ключевым словам. Это один из самых распространенных видов поиска. Для поиска по ключевым словам необходимо ввести в строке поиска слово или несколько слов, которые следует искать, и щелкнуть на кнопке Поиск. Поисковая система найдет в своей базе и покажет документы, содержащие эти слова.

Поисковая система Google (рисунок 2).

Google позволяет осуществлять поиск в различных источниках информации: во Всемирной паутине, в каталоге, в дискуссионных группах, а также искать картинки.



Поиск в Google

Мне повезёт!

Рисунок 2 – Интерфейс Google

Возможен также расширенный поиск (рисунок 3), который позволяет разыскивать документ на разных языках и в различных форматах (PDF, DOC, XLS и т.д.), показывать страницы, измененные в течение определенного времени, и учитывать ряд других параметров.

Расширенный поиск

Найти страницы

со словами:

со словосочетанием:

с любым из этих слов:

без слов:

с диапазоном чисел: -

Как это работает в обычном поиске

Введите ключевые слова: Иван, ездорович, кружевницы

Заключите словосочетание в кавычки: "Иван, ездорович"

Вставьте оператор OR между словами: человек OR пародист

Поставьте знак минус перед словами: -пародист, -"Иванга в пародист"

Вставьте две точки между числами и укажите единицу измерения: 2000...10000 рублей, 10112...12345

Дополнительные настройки

Искать на: Поиск страниц на выбранном языке.

Страна: Поиск страниц, созданных в определенной стране.

Дата обновления: Поиск страниц, которые были созданы или обновлены в течение указанного времени.

Сайт или домен: Поиск на определенном сайте (например, wikipedia.org) или в домене (например, .edu, .org или .gov).

Расположение слов: Поиск по тексту, заголовку или адресу страниц, а также по ссылкам на них.

Безопасный поиск: Используйте **Безопасный поиск**, чтобы избежать от непристойных и непристойных сайтов и картинок в результатах поиска.

Формат файлов: Поиск страниц и файлов определенного формата.

Права на использование: Поиск страниц, которые можно свободно использовать, распространять и изменять.

Рисунок 3 – Расширенный поиск Google

Работая с Google, полезно знать несколько операторов, которые упрощают поиск. Например, для того чтобы получить толкование слова или сочетания слов, достаточно перед незнакомым термином (через пробел) напечатать «define».

Так, задав запрос «define Интернет», вы получите специальную ссылку, предоставляющую всевозможные толкования термина в Сети.

Слово «site», сопровождаемое двоеточием, позволяет ограничить поиск конкретным сайтом. Например, запрос «сеть site:www.hownetworks.boom.ru» позволяет найти документ со словом «сеть» на конкретном сайте.

Для каждого документа выдается ссылка на ресурс и ссылка [Показать найденные слова] — переход на подсвеченный документ, где контекст запроса выделен стрелочками, которые обеспечивают переход по выделенным словам Расширенный поиск.

Для достижения положительного результата пользователь должен прибегнуть к услугам нескольких поисковых систем. Можно это сделать самостоятельно, переходя из системы в систему, а можно поручить эту работу одной из метапоисковых систем (meta — первая составная часть сложных слов, обозначающая системы для описания и исследования других систем).

Одно из назначений метапоискового сервиса при поиске — тестирование Сети на предмет информации.

Язык формирования запросов

Для того чтобы сделать поиск более продуктивным во всех поисковых системах существует специальный язык формирования запросов со своим синтаксисом.

Правила формирования запроса и поисковой системе Яндекс:

Ключевые слова в запросе следует писать строчными буквами. Это обеспечит поиск всех ключевых слов, а не только тех, которые начинаются с прописной буквы.

При поиске учитываются все формы слова по правилам русского языка, независимо от формы слова в запросе.

Поиск слова в заданной форме — оператор «!». Ставится перед словом, которое нужно найти. По умолчанию весь поиск ведется без учета морфологии языка, т.е. слово ищется в любом падеже,

склонении и числе. Для точного поиска используйте оператор, при необходимости применяйте в одном запросе несколько операторов. Пример: !города.

Поиск по цитате – оператор «”». Помогает найти точную фразу или отрывок текста. Включите в кавычки требуемый запрос. Пример: «Я к Вам пишу, чего же боле».

Поиск по цитате с пропущенным словом – оператор «*». Используется, когда в отрывке пропущена часть текста. Применяется только в сочетании с оператором поиска по цитате. Пример: «В августе 2020 рост индекса заработной платы составил * в Москве».

Поиск с нужным словом – оператор «+». Будут найдены документы, в которых обязательно присутствует слово с оператором «+». Основное слово поиска без оператора может присутствовать не на всех страницах. При необходимости используйте в запросе несколько операторов. Пример: История +Триумфальная +арка.

Поиск без определенного слова – оператор «-». Помогает найти страницы, на которых есть запрос, но отсутствует слова с оператором. Можно использовать несколько операторов в одном запросе. Пример: Театр Станиславского -Немировича -Данченко

Поиск на выбор – оператор «|». Ищите страницы, на которых присутствует любое слово из запроса, как минимум одно. Используйте несколько операторов в запросе. Пример: кино | фильмы | сериалы.

Правила формирования запроса в поисковой системе Google.

Сам Google не афиширует информацию о своих «помощниках». На официальном сайте поддержки поисковика есть информация лишь о некоторых операторах, но далеко не обо всех. Условно все операторы можно поделить на две группы: простые – представлены, как правило, в виде одного-двух спецсимволов, знаков препинания; сложные (документные) – целые слова (фразы) или предлоги.

К простым относятся:

«+» (плюс). Данный оператор позволяет учесть то или иное слово (предлог) в обязательном порядке. Для этого нужно добавить перед ним «+». В результате поисковая система покажет только те ресурсы, где отмеченное слово (предлог) содержится. В одном запросе может быть сразу несколько таких плюсов. Например: Дос-

тоевский идиот содержание +кратко.

«-» (минус). Если «+» позволял находить сайты, которые обязательно содержат, то «-» ищет ресурсы, которые обязательно не содержат указанное слово (исключает его).

Он также добавляется впереди слова и может употребляться несколько раз в одном запросе. Например: рецепт суши – филадельфия – хосомаки.

«..» (две точки). Этот оператор позволяет искать информацию в определенном числовом диапазоне. Это могут быть даты, цены, измерения и т.д. Все числа из диапазона будут подсвечиваться в сниппетах. Пример: купить игровой ноутбук \$200..\$300.

«|» (прямая черта). Чтобы Google искал совпадения по любому из ключевых слов, содержащихся в запросе, их необходимо связать символом «|» (или). Вместо прямой черты можно использовать предлог «OR» (обязательно заглавные) – результат идентичен. Пример: женское платье розовое | нежно-розовое | светло-красное | малиновое.

«» (кавычки). Применяются в случаях, когда необходимо точное совпадение словоформы, фразы. Система исключает сайты с иными формами написания, иным порядком слов и т.д. Данный оператор хорошо сочетается с другими, позволяя делать запрос максимально конкретизированным (например, с «-»). Примеры, «емкость батареи Samsung S8».

«~» (тильда). Чтобы помимо указанного ключевого слова поисковик искал и его синонимы, необходимо впереди добавить символ «тильда». Например, ~стагфляция причины.

«*» (звездочка). Символ «*» позволяет искать пропущенные слова из устоявшихся и популярных фраз, выражений, строк музыкальных и художественных произведений, цитат. Помогает, когда пользователь забыл или не может четко сформулировать свой запрос. Например, первый закон термодинамики для*процесса.

Оператор «at» (или, как его в просторечье называют, «собака») служит для поиска информации по тегам в социальных сетях. Пример: @lordjurrd.

«#» (решетка). Оператор для поиска информации по тегам. В данном случае поиск осуществляется по хештегам и не только в социальных сетях. Все сайты с указанным хештегом отобразятся в ре-

зультатах выдачи. Пример, как прошли #выборы2020.

«\$» (знак доллара). Чаще всего употребляется в коммерческих запросах для того, чтобы найти товар или услугу по конкретной цене или в конкретном ценовом диапазоне, если применить оператор «..».

Добавив символ «\$» перед числом, пользователь заставит поисковую систему искать совпадения именно в ценах (не только в долларах, но и в любой другой валюте). В результатах выдачи совпадения будут подсвечены. Пример, купить телевизор LG \$300.

«AND». Данный оператор аналогичен обычному пробелу – позволяет учитывать все ключевые слова, расположенные слева и справа от него (прописывать только заглавными буквами). Например, детские AND футбольные AND бутсы.

«site:». Если в поисковой строке написать «site:», а после указать домен ресурса, то поиск будет осуществлен только внутри указанного сайта. Особо часто применяется для поиска информации на крупных площадках и порталах, где поиск вручную не настолько эффективен. Например, site:ru.wikipedia.org что такое математический анализ.

«related:». Оператор позволяет искать информацию на площадках со схожей тематикой указанного ресурса. Употребляется, когда пользователь хочет найти аналогичный контент на других сайтах. Так же, как и с предыдущим «помощником», после двоеточия необходимо указать домен исходной площадки. Чтобы поисковик нашел не просто сайты-аналоги, а конкретную информацию на их страницах, после домена необходимо задать свой запрос. Пример, related:amazon.com ноутбуки Lenovo.

«cache:». Поисковая система Google при индексации страниц сохраняет их в кэше у себя на серверах. Если по каким-то причинам сайт перестал работать (например, по техническим), можно найти их кэшированные версии при помощи оператора «cache:», указав URL-адрес искомой веб-страницы. Также можно добавить ключевые слова для поиска определенной информации на указанной странице. Пример, cache:https://auto.ru bmw m6.

«filetype:». Чтобы найти не просто html веб-страницу, а документ в конкретном формате (pdf, doc, rtf и т. д.), необходимо добавить к запросу «filetype:» и указать нужный формат. Т.е. мы указы-

ваем поисковику, документы какого формата нужно искать. Например, форма 16 filetype:pdf.

«info:». Оператор «info:» представляет большую ценность для веб-мастеров, так как позволяет получить различную информацию об указанном веб-адресе. Пример использования: info:vk.com.

«link:». Позволяет получить список сайтов-доноров, ссылающихся на указанный домен. Также используется веб-мастерами для поиска полезной информации (в том числе и о конкурентских ресурсах). Например: link:https://destacar.de.

«allintitle: (intitle)». Используется для поиска страниц, содержащих заданную фразу в тайтле (в заголовке). Учитываются все слова, введенные после двоеточия.

«intitle:» – аналогично, но учитывается только одно слово, стоящее сразу после оператора (остальные ключи будут содержаться не обязательно в заголовке). Пример, allintitle: демонтаж деревянных окон.

«allinurl (inurl)». Если использовать его, поисковик Google будет искать совпадения в URL-адресе:

- «allinurl:» учитывает все слова после двоеточия;
- «inurl:» учитывает только первое.

Пример, allinurl: продвижение группы ВК.

«allintext (intext)». Данный оператор работает по тому же принципу, что и два предыдущих. Изменяется лишь место для поиска совпадений.

«allintext:» и «intext:» ищут заданные ключи непосредственно в тексте на страницах (в контенте), учитывая все или только первое слово соответственно. Примеры использования: allintext: остеология это наука.

«allinanchor: (inanchor)». Последние операторы из категории «allin (in)» в языке поисковых запросов Google. В данном случае совпадения ищутся в анкоре.

Анкор – это текст гиперссылки (то, что видит пользователь). Так как в анкор можно поместить любой текст, в него зачастую добавляют ключевые слова веб-страницы, на которую ведет ссылка. И именно по этим ключам будет осуществляться поиск, если в запрос добавить «allinanchor» или «inanchor». Например, allinanchor: купить авиабилеты онлайн.

«define:». Один из самых полезных «помощников» в поиске информации для тех, кто часто имеет дело с терминологией. «define:» дает команду поисковой системе искать именно определения (страницы, на которых есть определение). В большинстве случаев Google сам даст пояснение в отдельном окне в самом начале выдачи. Например, define: бюджетная ссуда.

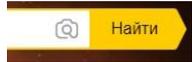
«movie:». Применяется, если необходимо найти информацию именно по фильмам. Названия некоторых фильмов состоят из одного-двух слов, которые употребляются и в обиходе. Получая такой запрос, Google не понимает, что именно пользователь хочет найти: определение, тематическую статью, фильм или что-то еще. Оператор «movie:» уточняет намерение пользователя. Пример: movie: любовь.

Возможно, на момент выполнения лабораторной работы некоторые из операторов уже не будут работать (в силу появления новых технологий). Ведь Google никогда не стоит на месте и постоянно обновляет свои технологии. Он стремится по максимуму облегчить процесс поиска информации, и такие «помощники» лишнее тому подтверждение.

Поиск по картинке

Яндекс умеет искать изображения по заданному образцу. Это может быть целое изображение или его фрагмент. При этом образец может размещаться в интернете или храниться на вашем компьютере или другом устройстве.

Поиск по картинке основан на алгоритмах компьютерного зрения. В результате поиска могут быть найдены как точные копии исходного изображения, так и картинки, незначительно отличающиеся от оригинала. Например, это могут быть разные фотографии одного и того же памятника архитектуры.

Как начать поиск по картинке? Нажмите кнопку  Найти в поисковой строке сервиса и в открывшемся диалоговом окне загрузите исходное изображение одним из следующих способов:

- перетащив исходную картинку в диалоговое окно;
- нажав ссылку Выберите файл и указав путь к картинке на жестком диске;

– введя URL картинки в поисковую строку.

Затем нажмите кнопку Найти.

Результаты поиска по картинке объединяются в группы. Вы можете открыть копию картинки в другом размере или узнать, на каких сайтах встречаются такие же картинки.

Задание

1. Поиск по справочникам адресов.

Пользуясь одним из справочников адресов Интернета, найти адреса следующих объектов:

- музей Эрмитаж;
- компьютерный путеводитель по Москве;
- музей А.С.Пушкина;
- страница одной из музыкальных групп.

2. Сравнение основных рубрик поисковых систем.

Поочередно открыть 2-3 поисковые системы. Просмотреть перечень основных рубрик. Сравнить рубрики поисковых систем.

Какая поисковая система, на ваш взгляд, обладает более продуктивной системой каталогов?

3. Поиск по каталогам.

Пользуясь каталогом поисковой системы, найти следующую информацию:

- текст песни М.Танича (по своему выбору);
- характеристики последней модели мобильного телефона известной фирмы (по своему выбору);
- рецепт приготовления любимого блюда;
- долгосрочный прогноз погоды на 10 дней;
- фотографию любимого артиста, певца, спортсмена, ...;
- информацию о вакансиях на должность учителя в г.Брянске;
- гороскоп своего знака зодиака на текущий день, месяц;
- советы по уходу за домашним животным, комнатным растением (по своему выбору).

4. Знакомство с языком запросов:

- сформировать запрос по точному названию или цитате;
- сформировать сложный запрос на поиск информации по закаливанию (исключить из поиска грудных детей, занятия с тренером).

5. Поиск по картинке:

Возьмите в интернете любое изображение. С помощью поиска по картинке найдите точную копию изображения.

Возьмите то же самое изображение, скачайте на компьютер.

Укажите путь к картинке на жестком диске и найдите её точную копию.

Список использованных источников

1. Колокольникова, А.И. Информатика [Текст] : учебное пособие / А.И. Колокольникова. – Москва ; Берлин : Директ-Медиа, 2020. – 289 с. : ил., табл. – Режим доступа: URL: <https://biblioclub.ru/>

2. Язык поисковых запросов Google [Электронный ресурс]. – Режим доступа: [www.url: https://www.rookee.ru/learn/kak-polzovatsya-operatorami-poiska-google-yazyk-zaprosov/](http://www.url:https://www.rookee.ru/learn/kak-polzovatsya-operatorami-poiska-google-yazyk-zaprosov/)

3. Язык поисковых запросов Яндекса [Электронный ресурс]. – Режим доступа: [www.url: https://www.ashmanov.com/education/articles/yazyk-zaprosov-yandeks/](http://www.url:https://www.ashmanov.com/education/articles/yazyk-zaprosov-yandeks/)