

Современные методы поиска научной информации

Курс для магистров 1-го года обучения

читает: к.б.н., доцент Гарипова Светлана Равилевна

Лекция 2

Поиск информации в электронных ресурсах интернет

Ресурсы научной информации в интернет

- Электронные каталоги (библиотек, организаций, личные хранения);
- Поисковые системы (машины): Google, Yahoo, Яндекс, Рамблер, Мейл;
- Реферативные базы данных (e-library, Scopus, Google Scholar, Open Access, PubMed, Elsevir, Sage и др.);

Проблемы поиска информации в интернет

- Говорят, что в Интернете есть все. На самом деле материалы для размещения в сети готовят живые люди, и потому там можно найти лишь то, что они сочли нужным опубликовать.
- Сегодня в Интернете несколько десятков миллиардов Web-страниц. При этом процент каталогизированных ресурсов неуклонно падает. Вывод: пространство Web быстрее наполняется, чем систематизируется
- Самый быстрый и надежный вид поиска – запустить браузер и набрать адрес web-страницы (если Вам он известен)

Поисковая система

- Это большая база ключевых слов, связанных с Web-страницами, на которых они встретились.
- Ключевые слова – это запрос поисковой системе;
- Результаты поиска выдаются в виде списка адресов Web-страниц, на которых встретились эти слова.

Как это работает?

- Поисковая система состоит из трёх частей:
 1. Робот (Spider, Robot или Bot) – это программа, которая посещает веб-страницы и считывает (полностью или частично) их содержимое.
 2. Индекс – хранилище данных, в котором сосредоточены копии всех посещенных роботами страниц
 3. Программы обработки запроса – это программа, которая в соответствии с запросом пользователя «просматривает» индекс на предмет наличия нужной информации и возвращает ссылки на найденные документы

Глобальная поисковая система Google

- Запущена в 1998 году, ныне единоличный лидер среди глобальных поисковых систем;
- Объем его индексного файла составляет на сегодня более 4,2 миллиардов web-страниц;
- В сутки программы-роботы системы индексируют порядка трех миллионов новых и обновленных страниц, актуализация базы производится каждые 28 дней

Возможности Google

- Способность индексировать документы в разных форматах (PDF, RTF, PS, DOC, XLS, PPT, WP5) и моментально конвертировать их в обычный HTML-файл, что освобождает пользователя от необходимости иметь специальное программное обеспечение;
- Имеет специально разработанный модуль ранжирования результатов, позволяющий отбирать наиболее авторитетные по числу и значимости ссылок страницы
- Отличается высокой степенью комфорта для пользователя, имеет интерфейс на русском языке.

Дополнительные сервисы Google

- Ссылки на документы, помимо традиционных сведений о заглавии документа, контексте искомых слов и данных о размере, содержат функцию Cached "Сохранено", позволяющую полностью восстановить весь документ, если по каким-либо причинам сайт, на котором он расположен, недоступен.
- Функция "Похожие страницы" позволяет получить перечень страниц, содержание которых схоже с указанным источником.
- Дополнительным платным сервисом **Google** является поиск труднодоступной информации непосредственно человеком. Стоимость этого вида обслуживания - \$2.50 за ответ

Конкретизация запроса через «Расширенный поиск»

- Можно добавить фильтры, ограничивающие язык документа, его формат (к примеру, "только документы в PDF"), время опубликования ("последние три месяца"), место термина в самом документе ("в заголовке страницы") или расположение страницы в определенном домене или даже сайте.
- при работе с файлами большого объема он индексирует не весь источник, а лишь его первые 101 Кб. (для PDF-файлов лимит ограничен 120 Кб.), поэтому индексация документа пока не всегда гарантирует возможность его нахождения по любому фрагменту текста.

Синтаксис поисковых запросов

AND

- Поисковый механизм игнорирует стоп-слова (предлоги, союзы, артикли), однако, если какое-либо из таких слов существенно, перед ним необходимо поставить "+", давая понять системе, что в данном случае термин даже из одной буквы является значимым (например, Александр +I).
- Обратите внимание: знаки “+” и “-” в поисковом запросе размещаются не как в математической формуле. Перед каждым из этих знаков в запросе обязательно ставиться пробел, а после знака сразу должно идти следующее искомое слово.
- Можно также поставить знак AND или символ & (профессиональный and поиск профессиональный & поиск)

Синтаксис поисковых запросов

OR и NOT

- **OR** (“ИЛИ”) – указывает, что в искомом документе, должно быть хотя бы одно из слов разделённых этим оператором. Данный оператор заменяют знаком “|”. Пример: *профессиональный | сложный +поиск в Интернете*
- **NOT** (“БЕЗ”) – предполагается что будет найдены документе в которых будут встречаться слова указанные перед оператором и исключаться документы в которых будут слова стоящие после оператора NOT. Также заменить оператор могут знаки “-” (минус) и “~” (тильда). Пример: *профессиональный +поиск -расширенный*

Синтаксис поисковых запросов

NEAR и кавычки

- **NEAR** (“близко”) - слова объединённые этим оператором, должны быть расположены в максимальной близости друг от друга. Заменяется оператор знаком “&&” или использованием кавычек. Тем самым вы укажем поисковику, что ищем именно словосочетание, а не просто несколько слов.
профессиональный && поиск “профессиональный поиск”
- Если оператор повторяется один раз (например, & или ~), поиск производится в пределах предложения. Двойной оператор (&&,~~) задает поиск в пределах документа.

Синтаксис поисковых запросов

Символы подстановки «*» и «?»

- Чтобы найти документы в которых встречается не только слово “поиск”, но и слова начинающиеся с “поиск”, такие как “поиском”, “поисковики”, “поисковые запросы”, нужно подставить символ “*” и выполнить следующий запрос: `поиск*`
- Если же нам нужно заменить какой либо символ в искомом слове, можно воспользоваться символом подстановки “?”, например так: `по??к` При этом будут найдены все документы в которых встречаются слова из 5 букв начинающиеся с “по” и заканчивающиеся на букву “к”.
- Ещё одним интересным знаком в поисковых запросах является символ “@”. Символ “@” применяется в запросе если нужно найти слово со всеми его словоформами: `@поиск`

Особенности языка запросов в Яндекс

- Вместо слов “AND”, “OR” и “NOT”, используются символы “&”, “|” и “~” соответственно. Действие этих знаков ограничивается только пределами одного абзаца. Для того чтобы, применить действие оператора для всего текста документа, нужно “удвоить” оператор. К примеру, вместо “~” поставить “~~” и т.п.
- Яндекс имеет и спецсимволы, которые могут определять на каком расстоянии могу располагаться в тексте искомые слова. /n – где n является количество слов разделяющих искомые слова в тексте. Например по запросу: *расширенный /4 поиск* будут найдены документы, в которых слова “расширенный” и “поиск” встречаются рядом, но на расстоянии максимум четыре слова друг от друга. Также символами “+” и “-” (которые здесь подразумеваются в другом значении) можно указать в какую сторону мы ведёт отсчёт слов разделяющий искомые слова.

Еще о специфике запросов в Яндекс

- Поискковая система будет искать все формы слов, но различает части речи
- Чтобы сохранить в поиске нужную форму слова, ставим его в кавычки или впереди восклицательный знак (пример: !сентября), два восклицательных знака перед словом означает поиск нормальной формы слова (именительный падеж единственное число) слова, то есть возможны падежи, но не омонимы его вариантов
- Если в цитате пропущено слово, ставим знак *
- Если нам нужны любые варианты из заданных слов -

Поиск изображений

- Режим "Поиск изображений" (Images, "Картинки") позволяет найти схемы, фотографии, рисунки, которых в базе данных Google учтено более 880 тысяч.
- Разыскание иллюстраций, ведется аналогично поиску текстовых документов с возможностью ограничения определенным размером, форматом или цветностью графических файлов - все через опцию "Расширенный поиск изображений".

Задания по поиску информации в интернет с помощью поисковых систем и реферативных баз данных

- Задание 1.** Сформулируйте поисковый запрос по своей теме с использованием разных вариантов синтаксиса и проведите поиск информации в Google и Яндекс. Укажите результат, полностью удовлетворяющий Вашему запросу. Используйте разные варианты поиска:
- Библиографический: определить источник по цитате и дать его полное библиографическое описание;
 - Фактографический: найти ответ на вопрос;
 - Документальный: найти полный текст статьи по названию или автору;
 - Аналитический: найти обзор литературы по заданной теме

- 3) Как можно просмотреть выпуски журналов, появившихся на сайте за последнее время?

Задание 2. Изучите ресурсы e-library и кратко дайте письменный ответ на вопросы

- 4) Какие сервисы являются не доступными при пользовании ресурсами Научной электронной библиотеки elibrary.ru без регистрации в ней?
- 5) Какие возможности дает создание своего профиля в Научной электронной библиотеке elibrary.ru?
- 6) Какие разделы можно настроить в панели “навигатор”?
- 7) Какие результаты позволяет получить использование функции “Простой поиск” и каковы недостатки этого подхода?
- 8) Что означает знак звездочки * при введении слов в простом поиске?
- 9) Каково назначение клавиши «добавить» при проведении расширенного поиска? Ограничиваем или расширяем мы при этом поиск? Какие можно при этом выбрать позиции поискового запроса?
- 10) Какие особенности использования в e-library синтаксиса и слов AND, OR, AND NOT, NEAR?
- 11) Как узнать, входит ли интересующий Вас журнал в перечень, рекомендуемый ВАК? Как выяснить его импакт-фактор?
- 12) Как узнать, доступен ли полный текст интересующей Вас статьи в Научной электронной библиотеке elibrary.ru?

Задание 3. Проведите поиск литературы в e-LIBRARY и запишите результаты поиска

- По ключевым словам
- По автору
- По тематическому рубриктору
- По каталогу журналов
- По каталогу книг
- С помощью опции «найти похожие статьи»

Практическую работу оформите в виде таблицы с графами:
вариант поиска, содержание поискового запроса, результаты

Задание 4

- Найти, распечатать и принести тексты четырех статей по Вашей теме в разных жанрах: 1) научная экспериментальная статья; 2) научная обзорная статьи (обзор литературе); 3) научно-практическая статья; 4) научно-популярная статья.
- Назвать по 2-3 журнала, в которых публикуются статьи по вашей теме в разных жанрах.

Задание 4. Изучите возможности Scopus, используя руководство, и сравните их с eLibrary

- По особенностям поиска источников;
- По возможностям экспорта данных;
- По анализу цитируемости;
- По анализу журналов;
- По инструментам для авторов;
- По правилам поиска (особенности синтаксиса поискового запроса)