



Министерство образования и науки Республики Казахстан

Павлодарский государственный университет им. С. Торайгырова

Кафедра Вычислительная техника и программирование

МЕТОДИЧЕСКИЕ УКАЗАНИЯ И РЕКОМЕНДАЦИИ

к лабораторным работам

по дисциплине Интернет технологии

для студентов специальности 050704 Вычислительная техника и программное
обеспечение

Павлодар



УТВЕРЖДАЮ

Проректор по УР

_____ Пфейфер

Н.Э.

(подпись)

(Ф.И.О.)

«__» _____ 201__ г.

Составитель: ст. преподаватель _____ Глазырина Н.С.

Кафедра Вычислительная техника и программирование

Методические указания и рекомендации

к лабораторным работам

по дисциплине Интернет технологии

для студентов специальности 050704 Вычислительная техника и программное
обеспечение

Рекомендовано на заседании кафедры

«__» _____ 201__ г., протокол №__

Заведующий кафедрой _____ Потапенко О.Г. «__» _____ 201__ г.
(подпись) (Ф.И.О.)

Одобрено УМС Физики, математики и информационных технологий
(наименование факультета)

«__» _____ 201__ г., протокол №__

Председатель УМС _____ Муканова Ж.Г. «__» _____ 201__ г.
(подпись) (Ф.И.О.)

ОДОБРЕНО ОПиМОУП:

Начальник ОПиМОУП _____ Варакута А.А. «__» _____ 201__ г.
(подпись) (Ф.И.О.)

Одобрена учебно-методическим советом университета

«__» _____ 201__ г. Протокол №__

Лабораторная работа № 1

Тема: Поиск в Интернете.

Цель работы: Научиться осуществлять поиск нужной информации в Интернете.

В результате выполнения работы студент должен:

Знать:

- какие поисковые системы используются;
- как найти нужную информацию в Интернете;
- как пользоваться языком запросов для поиска информации.

Уметь:

- работать в программе Internet Explorer;
- создавать запросы, используя язык запросов;
- осуществлять поиск нужной информации.

1. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

1.1 Поисковая система Улитка

Для работы в Интернете используют программу Internet Explorer. В поле Адрес набирается нужный адрес сайта. Например, введем адрес каталога Улитка- <http://www.ulitka.ru>.

Через несколько секунд или минут (в зависимости от качества связи) в окне браузера возникнет картинка, показанная на рисунке 1.

Автомобиль	Бизнес, финансы
Домашние страницы	Иностранные языки
Интернет	Компьютеры и программы
Культура и искусство	Люди
Медицина	Наука, техника, производство
Общество и государство	Отдых, развлечения
Пиво	Погода
Разное	Спорт
Справочник	Средство массовой информации

Рисунок 1- Каталог Улитки

Улитка имеет прямое отношение к поиску информации. Документы, доступные в Интернете, разбиты в этом каталоге по темам. Среди тем показанных на рисунке, Автомобили, Бизнес, Финансы, Экономика, Интернет и т.д.

Попробуем теперь найти в каталоге Улитка нужную информацию. Предположим, нас интересует «телеработа», то есть работа, выполняемая для заказчика, живущего в другой стране, например Америке. Чтобы начать поиск, нужно выбрать правильный раздел. Поэтому спускаемся в нижнюю часть окна (с помощью, например, клавиши ↓), находим ссылку Работа и Обучение, подводим к ней указатель мыши и нажимаем левую кнопку. В окне браузера появиться картинка. Раздел Работа и Обучение разбит в каталоге Улитка на семь подразделов: Вузы, Коллекции рефераты и т.д.

Работа и обучение (556 ссылок)

Вузы

Поиск работ и вакансий

Семинары

Школы

Коллекция рефератор

Разное

Учебные курсы

Наиболее подходит раздел Поиск работы и вакансий, поэтому выбираем его.

В окне появляется список, поиска работ. Всего документов 223, но документ под номером 2- Телеработа (Telework)

Поиск работы и вакансий

1) Работа в Москве

2) Телеработа. Работа в других странах через интернет.

1.2 Поисковая система Апорт: сайты и страницы

Рассмотрим работу поисковой системы Апорт (www.aport.ru)¹ и попробуем найти то же, что и в каталоге Улитка, слово – «телеработа». Итак, соединимся с поисковой системой Апорт, наберем в белом поле (адрес) которое находится в верхней части страницы слово телеработа и нажмем кнопку Найти.

Через некоторое время результаты поиска покажутся в окне браузера и первым будет виден именно сайт, в нашем случае это List.ru. «Рубрицированный каталог аннотированных ссылок», эта фраза относится к сайту www.list.ru и означает набор ссылок, то есть адресов сайтов, помещенных в разные разделы-рубрики. Аннотированные ссылки – это описания и адреса сайтов.

Этот раздел начинается с заголовка страницы Карьера/работа: Работа в Интернете: Телеработа, дальше идет ее адрес <http://www.list.ru/catalog/11506.html>, а за ним — те места из полного текста страницы, где встречается слово, которое мы ищем (в нашем случае — «телеработа»). По ним легко понять, о чем данный документ, интересен ли он. Самая подходящая страница называется 11506.html и находится на сайте www.list.ru в папке catalog. Сайты состоят из отдельных документов-страниц. Большие сайты занимают целый компьютер, а страницы — это отдельные файлы, которые для порядка помещены в разные папки. Но может быть и так, что на одном компьютере есть несколько сайтов, расположенных в разных папках одного жесткого диска.

Легко заметить, что Апорт понимает под сайтом именно набор тематически связанных страниц, независимо от того, каков их адрес состоит ли он только из слов, разделенных точками, или более сложен.

Этот сайт найден и системой Апорт. Естественно, Апорт не может показать сразу все 145 найденных сайтов, а показывает только первые 10. Чтобы посмотреть второй десяток, нужно перейти в самый конец списка сайтов, выданных системой в первом десятке, и выбрать мышью изображение листочка с номером 2.

Улитка, как мы помним из раздела «Окно в Интернет», содержит только адреса и краткие описания сайтов. В то же время Апорт, похоже, хранит содержимое всех известных системе сайтов, иначе было бы невозможно показать те места на странице, где встречается указанное нами слово «телеработа».

Апорт - поисковая система совершенно другого типа – так называемый тематический индекс. Апорт «навешает» все известные системе сайты, переписывает тексты всех доступных страниц и хранит эти тексты вместе с их адресами в своей базе данных. Обращаясь к системе Апорт, мы ищем слово «телеработа» во всех, текстах страниц, которые есть в ее базе данных. Обращаясь к системе Улитка, мы ищем слово «телеработа» в описаниях сайтов, которые в ней хранятся. Ясно, что описания содержат гораздо меньше информации, чем сами сайты, вот почему в системе Апорт удалось найти 389 документов, а в системе Улитка — всего 2.

Так зачем же тогда вообще нужны такие сайты как Улитка? Если Апорт предоставляет гораздо больше информации, так давайте будем пользоваться Апортом. и забудем про Улитку! Чтобы понять, почему делать этого не стоит, попробуем разобраться в том, как формируются базы данных систем Апорт и Улитка.

Так вот. Апорт — это *автоматический индекс*. То есть Апорт навешает все известные системе сайты и переписывает все встреченные тексты страничек в свою базу данных. Кроме того, посетив сайт. Апорт запоминает все ссылки на другие страницы и сайты и дальше, когда выпадет свободное время, навешает эти сайты и также переписывает их себе в базу. Ясно, что через какое-то время работы в базе данных окажутся все русскоязычные сайты (в Интернете есть сайты на разных языках, Апорт — это местная поисковая система, которая интересуется документами на русском языке).

Апорт, как и любой автоматический индекс, состоит из двух «половинок». Первая — это сама база данных и программа поиска нужного слова. Набирая в соответствующем поле слово телеработа и нажимая кнопку Найти мы взаимодействуем именно с этой частью системы Апорт. Вторую часть называют по-разному: пауком (spadier), роботом (robot), червем (worm); она ищет новые сайты и переписывает их содержимое в базу данных. Когда какой-нибудь сайт или страница оказываются в базе данных, говорят, что они проиндексированы.

Автоматический индекс, как гигантский пылесос, засасывает все, что встретит на своем пути. В то же время, каталог ресурсов, такой, как Улитка, гораздо более разборчив. Сюда попадают только те сайты, которые посмотрел и одобрил редактор — специальный сотрудник, занимающийся их поиском и оценкой.

Чтобы понять, почему предварительный просмотр сайта может быть полезен, нужно знать, что Интернетом никто не управляет, здесь нет никакой цензуры и за содержание сайта в подавляющем большинстве случаев отвечает только его хозяин. Значит, никакой гарантии, что данный сайт соответствует своему названию, нет. И тем более нет гарантий, что документы, помещенные на сайт, содержат полезные и правдивые сведения.

Когда Апорт записывает все сайты подряд в свою базу данных, система, наряду с ценными и точными сведениями, засасывает горы словесного мусора. Вместе с красивыми и тщательно спроектированными сайтами Апорт индексирует неряшливые домашние страницы, где в беспорядке на безумном фоне разбросаны фотографии хозяина, его жены и собаки. Такие страницы заведомо не попадут в Улитку, потому что их не пропустит редактор.

У каталогов ресурсов, таких, как Улитка, есть, как ни странно, еще одно преимущество - малый размер.

Из описания систем Улитка и Апорт можно понять, что сайты попадают в каталог или индексируются, только когда их найдет редактор Улитки или автоматический робот Апорта. На самом деле, большинство сайтов представляют поисковым системам их хозяева. Для этого и в Улитке и в Апорте есть специальные ссылки Добавить сайт. Хозяин сайта вводит его адрес и краткую характеристику. Если сайт представлен системе Апорт, он будет обязательно помещен в базу данных (проиндексирован). Если же сайт представлен Улитке — его судьба зависит от редактора, который может занести его адрес в наиболее подходящий (с его точки зрения) раздел каталога, а может и не занести, если сочтет сайт неинтересным или не соответствующим теме.

1.3 Устройство страниц. Рейтинги

Занимаясь поиском слова «Телеработа» в Апорт можно было заметить, что из 389 документов, некоторые оказываются в первом десятке, очень важно, чтобы первыми были показаны самые лучшие документы.

Значит перед любой поисковой системой стоит задача сортировки найденных страниц в порядке убывания важности.

У каждой поисковой системы своя правила сортировки. И все было бы хорошо. Если бы не желание хозяев страниц оказаться в первом десятке. Для этого они пытаются понять алгоритм сортировки конкретной поисковой системы с целью обмануть ее - создать страницу, которая обязательно попадет в первый десяток.

Чтобы обмануть поисковые системы, поначалу достаточно было десятки раз набрать невидимыми, сливающимися с фоном буквами нужное слово, и страница оказывалась в числе первых. Постепенно между поисковыми системами и хозяевами сайтов развернулась настоящая война. Поисковые системы стали часто менять алгоритмы сортировки, хозяева сайтов старались обмануть новые алгоритмы. Появились фирмы, специально создающие страницы, которые будут показаны в числе первых при поиске определенного слова.

Эта война идет до сих пор с переменным успехом. И, как бы понимая ее бессмысленность, хозяева поисковых систем стали изобретать новые способы сортировки, среди которых самый естественный — показывать первыми те страницы, которые чаще посещаются. Посмотрим, как организован один из самых популярных российских рейтингов - Rambler (www.rambler.ru).

Сайты, включенные в рейтинг, разбиты по темам – совсем как в каталоге ресурсов. Имена разделов (Авто/Мото, Литература, Религия) показаны в средней нижней частях рисунка.

Выбрав ссылку Развлечения, увидим список самых посещаемых сайтов этого раздела. Согласно рейтингу Rambler, больше всего посетителей у сайта, посвященного анекдотам. Это сайт «Анекдоты из России» — <http://www.anekdot.ru>.

Как вообще Rambler оценивает посещаемость сайта. Все, оказывается, просто. Сайт, участвующий в рейтинге, размещает у себя ссылку на картинку, которая хранится на сайте Rambler. Выдавая картинку тому или иному сайту, Rambler увеличивает на единицу счетчик посещений — так и формируется рейтинг.

Место сайта в рейтинге со временем меняется. Популярность некоторых сайтов растет, и они поднимаются вверх.

Рейтинг Rambler, с которым мы только что познакомились, во многом похож на тематический каталог Улитка. В нем сайты разбиты по темам. Разница лишь в том, что попасть в Rambler можно без предварительного отбора — достаточно зарегистрироваться на сайте (выбрав ссылку Добавить).

1.4 Язык запросов

Мы искали слово «телеработа» в индексе Апорт, то есть во всех страницах, хранящихся в базе данных индекса. Чтобы понять, насколько эта возможность замечательна и фантастична, вспомним, как приходится искать книги в обычной библиотеке.

Читателю традиционной библиотеки доступны два каталога - алфавитный и систематический. В алфавитном каталоге можно искать книги по фамилиям авторов или по названиям. То есть найти в нем можно лишь то, что уже известно из каких-то других источников, например из списка цитируемой литературы в книге или статье. Тематический каталог похож на «интернетовский» каталог ресурсов (Улитку, например). В нем книги расставлены по темам, согласно принятой у библиотечарей классификации. И в тематическом, и в алфавитном каталоге нам доступна только библиотечная карточка, где указываются авторы книги, ее название и дается краткая аннотация.

Можно представить себе, что Интернет — огромных размеров библиотека, у которой нет единого каталога. Библиотеку никто не охраняет, в ней десятки миллионов книг, валяющихся где попало. Хозяева книг переставляют их с полки на полку, вырывают из них страницы или вставляют новые. Иногда книга полностью переписывается, иногда хозяин забирает ее с собой и больше не возвращает. Часто он забывает о книге, и она валяется где-то в углу, никому не нужная.

Иногда в библиотеке появляются люди, которые пытаются навести порядок. Они ходят вдоль бесконечных полок, просматривают книги и заносят сведения о них в тематический каталог (Улитку, например). Некоторые группы библиотечарей считают, что кратких сведений о книгах в тематическом каталоге недостаточно; они берут с полок книги, одну за другой, и переписывают каждую в специальный индекс.

Примерно так составляются каталоги в Интернете, но библиотекари здесь быстры и не утомимы, работают круглые сутки, не болеют и в отпуск не ходят — это, специальные программы - роботы, навещают каждый сайт и переписывают его содержимое в специальную базу данных — индекс (например Апорт).

В индексе можно искать не только названия, но и любое слово. Причем не нужно перебирать бумажные карточки, достаточно набрать слово в поле поиска, а всю остальную огромную работу выполнит специальная программа.

Вот для этого почти в любой поисковой системе есть язык запросов, то есть правила, которые позволяют искать несколько слов и целые фразы. Пусть, например, вас интересует не просто телеработа, а телеработа в Германии. В этом случае разумно искать документы, где встречаются оба слова — «телеработа» и «Германия». На языке запросов Апорта это выглядит так, телеработа & Германия.

Как видно, поставив между словами знак &, система Апорт ищет документы, где обязательно встречаются оба слова. При этом число найденных документов уменьшится. Апорт находит всего 32 документа.

2. Задание

1. Используя язык запросов осуществить поиск фраз в заголовках документов в системах Апорт, Яндекс и Рэмблер.
2. Осуществить расширенный поиск, сгруппировав слова, соединенные логическими операторами.

Содержание отчёта

Отчёт должен содержать:

1. задание к работе;
2. программу.

Контрольные вопросы

1. Чем отличаются системы Улитка и Апорт?
2. Чем отличается рейтинг Рэмблер от Улитки?
3. Каково устройство страниц?
4. Как осуществляется рейтинги страниц?
5. Что такое язык запросов?

Литература

Основная литература:

1. Шафран Э. Создание Web-страниц. Санкт-Петербург: Питер,2000.
2. Истабрук Н. Освой самостоятельно Internet за 24 часа. М: Бином,1998.
3. Крупник А. Поиск в Интернете. Санкт-Петербург:Питер 2001.

Дополнительная литература:

1. Электронные вычислительные машины в 8-ми книгах . Учебное пособие для Вузов/под ред.Савельева А.Я.

Лабораторная работа № 2

Тема: Работа в системе Altavista.

Цель работы: Познакомиться с основным назначением поисковой системы Altavista.

В результате выполнения работы студент должен:

Знать:

- как работать в системе Альтависта;
- элементы языка запросов;
- как использовать язык запросов в системе Альтависта.

Уметь:

- осуществлять поиск информации в системе Альтависта;
- осуществлять детальный поиск информации;
- осуществлять запрос в режиме расширенного поиска.

1. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

1.1 Поисковая система Altavista

До сих пор мы искали слова в русскоязычных документах, и сейчас самое время вспомнить, что Интернет - мировая сеть, где есть документы практически на всех явных языках, но подавляющая их часть все-таки написана на английском. Попробуем поэтому поискать слово «telework» — английский эквивалент слова «телеработа» в индексе Altavista (www.altavista.com), который включает в свою базу данных тексты на разных языках.

Нажав кнопку Search (поиск), получим список адресов тех страниц, в которых есть слово «telework» В Альтависте нашлось 45 580 страниц, содержащих слово «telework».

Индекс Альтависты, с которым мы только что познакомились, принципиально ничем не отличается от индекса Апорта. Разница только в размерах.

Альтависта - глобальная поисковая система. Она индексирует все страницы подряд, независимо от того, на каком языке они написаны. В ее индекс попадают французские, португальские, русские, греческие и китайские страницы. Индекс Альтависта - один из самых больших в мире, и включает в себя примерно 200 миллионов страниц.

Индексы, подобные Альтависте, претендуют на то, чтобы вобрать в себя весь Интернет.

И чем дальше, тем хуже это получается, ведь Интернет стремительно растет и никакой даже самый мощный индекс не в состоянии за ним угнаться. Мало того, что скорость появления новой информации все увеличивается, но хуже то, что информация постоянно меняется. Web – сайты переезжают на другие компьютеры, а сами страницы меняют адреса. Вот почему любой индекс должен не индексировать новые страницы, но постоянно следить за старыми, периодически обходя весь известный ему Интернет.

Ясно, что Апорт, индексирующий только русские документы, объем которых едва ли составляет сотую часть всего Интернета, может лучше сделать свою маленькую часть работы: проиндексировать новые страницы, успевая следить за старыми.

Другое преимущество системы Апорт в том, что она учитывает особенности русского языка, о которых Альтависта даже не догадывается. Апорт, как мы уже видели, ищет не только заданное слово, но и все его грамматические формы.

Теперь поговорим о поиске в Альтависте слова «telework» — ведь это хороший повод еще раз обратить внимание на основную проблему поиска в Интернете — слишком большое число находок. Когда число найденных страниц в пределах одной-двух сотен, их еще можно просмотреть. Но никакая голова не выдержит 45 тысяч страниц (а могут быть и миллионы). Один из способов борьбы с этим мощным язык запросов, который позволяет, с одной стороны, уменьшить число документов, в которых идет поиск (например, можно искать слово «telework» в заголовках документов), а с другой стороны, сделать очень точный запрос в котором можно указать, какие слова или фразы должны быть в документе, а каких слов и фраз быть не должно. Так вот - Альтависта обладает мощным языком запросов.

Посмотрим, как уменьшается число документов, если искать слово «telework» в их заголовках. Чтобы Альтависта искала только в заголовках документов, нужно предварить слово, которое ищем, префиксом title:, введя в поле поиска словосочетание title: telework.

Нажав кнопку Search, получим чуть более 2 тысяч ссылок. Это почти в двадцать раз меньше, но все еще много. Дальнейшие действия зависят от наших целей и предварительных знаний. Если никаких знаний нет, стоит посмотреть первые выданные ссылки, попробовать освоить терминологию и познакомиться с предметом. Если же цель более определенная, нужно уточнить запрос.

Пусть, например, нам интересна телеработа в Германии. Тогда можно поискать документы, в заголовках которых есть слово «telework», а в тексте обязательно встречается слово «Germany». На языке запросов Альтависты это выглядит так

+title:telework +Germany

Знаки + означают «обязательное присутствие» и позволяют извлечь из десятков тысяч документов только те, которые содержат нужную нам информацию. Так, по введенному нами запросу Альтависта нашла всего 74 страницы.

Первая страница (скорее, это целый сайт) называется «Teleworking Germany», то есть «Телеработа в Германии». После названия страницы Альтависта дает ее краткую характеристику: «Teleworking Germany offers and searches for home based work opportunities...», то есть «Teleworking Germany предлагает и ищет возможности работать дома». Как видим, это как раз то, что мы искали.

Важным элементом языка запросов Альтависты является знак - (минус). Если поставить его перед словом, Альтависта будет искать все документы, где этого слова нет. С помощью знака минус легко определить, сколько вообще документов в индексе Альтависты. Для этого нужно поискать все документы, в которых есть данное (не важно какое) и все документы, где его нет, а потом сложить оба числа. Ведь сумма числа страниц, в которых слово есть, и числа страниц, в которых его нет, как раз и равна общему числу страниц в индексе.

Второй крайне важный элемент языка запросов - кавычки. Закрывая ряд слов в кавычки, мы приказываем Альтависте искать фразы, то есть последовательности слов, где важен их порядок.

1.2 Альтависта. Детальный поиск

Для поиска документов по имени в Альтависте есть специальный префикс url:

Например:

url.engines

По этому запросу Альтависта выдаст документы, в *адресах* которых есть слово «engines».

Можно пытаться угадывать название документа. Например, доменные имена сайтов, посвященных поиску, могут содержать фрагмент «searchengine», который с помощью префикса url: можно искать следующим образом:

url: searchengine

Другой замечательный способ «очистки» — поиск слов только в заголовках документов, часто отражающих их суть. Можно, например, искать документы, в заголовках которых есть фраза «search engine» или «search engines». Делается это с помощью префикса title:, как показано ниже.

title: “search engine*”

Звездочка в конце слова «engine» поставлена для того, чтобы Альтависта искала как слово «engine», так и слово «engines». Эта звездочка — очень важная мелочь, ее надо ставить, когда нет уверенности в том, как выглядит слово. Встретив звездочку, замыкающую слово «engine»,

Альтависта будет, конечно, искать не только слово «engines», но и все слова, начинающиеся с «engine».

Полезен и префикс anchor:, приказывающий Альтависте искать только среди ссылок, которые переносят нас к другим страницам и сайтам. Поиск среди ссылок может быть полезен, потому что ссылки короткие и тот, кто их создает, часто вынужден в немногих словах отразить суть документа, на который ссылка указывает.

Может помочь и префикс link:, который заставляет Альтависту искать все страницы, ссылающиеся на данную. Например, по следующему запросу будут показаны многочисленные адреса страниц, в которых есть ссылки на поисковую систему Альтависта:

Link:www.altavista.com

Префикс link: полезен прежде всего владельцам сайтов, потому что позволяет понять, кто на них ссылается. Но его можно использовать и для поиска похожих сайтов. Если найденный сайт посвящен какой-то специальной теме, то можно предположить, что сайты, которые на него ссылаются, тоже имеют отношение к этой теме.

Можно подумать, что знаков +, -, * и всевозможных префиксов, описанных в этом разделе, хватит, чтобы найти любой документ. Но это не так. Чтобы среди сотен миллионов страниц найти нужную, приходится порой сочинять очень сложные запросы, для которых язык плюсов и минусов оказывается слишком бедным.

Вот на этот случай в Альтависте предусмотрен так называемый расширенный поиск, перейти к которому можно, выбрав вкладку Advanced Search (расширенный поиск).

Если в режиме простого поиска перед каждым словом ставился плюс, и это означало, что ведется поиск документов, в которых есть все указанные слова, то в расширенном поиске с той же целью между словами ставятся знак &, который можно перевести как И. Следующий запрос в режиме расширенного поиска означает, что ведется поиск документов, в которых есть слово «bred» и слово «shows», то есть обязательно *оба* слова:

Bred & shows

Но если хочется, чтобы в найденных документах было *или* слово «bred», *или* слово «shows», или оба слова вместе, запрос становится таким:

bred | shows

А вот по следующему запросу Альтависта найдет документы, в которых есть слово «Clinton», но нет слова

«Lewinsky»:

Clinton ! Lewinsky

В языке запросов есть еще один знак — , который может быть заменен словом NEAR. Если его поставить между словами, то будут найдены страницы, в которых между словами «Bill» и «Clinton» не более 9 других слов. Причем неважно, какое слово встречается первым. Например:

Bill NEAR Clinton

В таблице 1 собраны запросы на языке простого поиска и точно такие же запросы режима расширенного поиска.

Таблица 1- Одни и те же запросы в режимах простого и расширенного поиска.

Простой поиск	Расширенный поиск
+bread +shows	Bread & shows; bread AND shows
Bread shows	Bread shows; bread OR shows
+Clinton -Lewinsky	Clinton ! Lewinsky: Clinton AND NOT Lewinsky
“Who is Mr. Putin?”	“Who is Mr. Putin?”

Как видим, знак & можно заменить оператором AND, а знак | действует так же, как оператор OR, восклицательный знак ! эквивалентен оператору AND NOT.

Языки расширенного поиска, как это видно из таблицы 1, позволяет делать то же, что язык простого поиска. Но, конечно же, он может и многое другое. Например, следующий запрос вполне обычен в расширенном поиске, но не реализуем в поиске обычном:

(bread AND shows) OR (freedom AND peace)

Его словесное описание выглядит так: ведется поиск документов, в которых есть пара слов bread, shows *или* пара слов freedom, peace. В этом последнем примере применен новый, очень

важный элемент языка расширенного поиска — скобки, которые позволяют группировать слова, соединенные знаками &, | и !.

В поисковых русскоязычных системах есть некоторые отличия в языке запросов от системы Альтависта. Например, Рэблер(Rambler.ru) не поддерживает поиск фраз, система просто не замечает окружающие слова кавычки. В Рэблере также нет плюсов и минусов для выделения слов, которые должны обязательно присутствовать или отсутствовать в документе. Вместо этих знаков нужно использовать логические операторы. Например, для присутствия обоих слов используется следующий запрос письменность AND алфавит. Если же нужно найти любое из нескольких слов, используется оператор OR.

Заставить Рэблер искать все грамматические формы слова можно, если поставить перед словом знак #.

Например

#русская #письменность

В автоматическом индексе Апорт для присутствия обоих слов в поиске необязательно ставить знак &. Апорт по умолчанию ищет документы, где есть оба слова. Однако для присутствия одного из слов или обоих слов вместе, слова соединяются оператором ИЛИ, например

Станислав | Лем

В Апорте можно использовать кавычки, задающие поиск фразы.

“Шерлок Холмс”

Поиск фразы в заголовке документа задается так:

title=(“Шерлок Холмс”)

Автоматический индекс Яндекс считает все грамматические формы слова одним словом. Но если нужно найти в точности те слова, которые указаны в запросе, перед ними ставится восклицательный знак.

!двенадцать !стульев

Для поиска фразы слова заключаются в кавычки. Также в поиске можно использовать знаки + и – как в Альтависте. Поиск фразы в заголовке производится следующим образом:

\$title(“двенадцать стульев”). Следующий запрос ищет графический файл:

#image=”visa.gif”

Задание

1. Используя язык запросов осуществить поиск фраз в заголовках документов в системе Альтависта.
2. Осуществить расширенный поиск, сгруппировав слова, соединенные логическими операторами.

Содержание отчёта

Отчёт должен содержать:

1. задание к работе;
2. программу.

Контрольные вопросы

1. Как производится поиск в заголовках документов в системе Альтависта?
2. Что такое детальный поиск?
3. Что такое расширенный поиск?
4. Чем отличается детальный поиск в русских системах от системы Альтависта?
5. Какие префиксы используются для поиска информации?

Литература

Основная литература:

1. Шафран Э. Создание Web-страниц. Санкт-Петербург: Питер,2000.
2. Истабрук Н. Освой самостоятельно Internet за 24 часа. М: Бином,1998.
3. Крупник А. Поиск в Интернете. Санкт-Петербург:Питер 2001.

Дополнительная литература:

1. Электронные вычислительные машины в 8-ми книгах . Учебное пособие для Вузов/под ред.Савельева А.Я.