**Лабораторная  работа № 1**

                    **ТЕХНОЛОГИЯ ПОИСКА ИНФОРМАЦИИ**

                                                    1. **Цель работы**

         Практическое освоение технологии эффективного поиска информации.

                                        2. **Общие сведения**

                 2.1.  **Средства поиска информации**

  В Интернет можно найти почти любую требуемую информацию.  Информационные ресурсы Интернет характеризуются необъятным количеством  материалов, накопленным за десятилетия существования компьютерных систем. Они содержат текстовые файлы, программы, картинки, музыку, фильмы; постоянно обновляются и лавинообразно нарастают.  Ресурсы Интернет широко используются практически во всех сферах человеческой деятельности.  Всевозрастающую роль играют они и в обучении.

       Навыки технологии профессионального поиска информации  крайне необходимы специалисту в области информационных технологий, которая невероятно обширна и динамична. Профессиональный поиск позволяет не только свести до минимума вероятность пропуска искомой информации, но и многократно сократить временные и финансовые затраты на ее поиск .

Для поиска информации в Интернет используются: поисковые машины, метапоисковые средства, индексированные каталоги, онлайновые энциклопедии и справочники. Современные поисковые порталы  содержат  не только  набор  вышеназванных поисковых средств,  но обеспечивают и дополнительные услуги, такие как бесплатные адреса электронной почты, места для размещения домашних Web-страниц и  т. д. Для эффективного поиска в Интернет необходимо знать принципы функционирования  поисковых  средств и уметь правильно формировать поисковый запрос.

  Поисковые машины,  осуществляя постоянное сканирование доступных  узлов Интернет, скачивают найденные страницы  в базу данных и формируют специальную базу данных, в которой хранится индексированная информация о скаченных страницах. При поступлении  запроса поисковая машина, использую индексированную информацию, выдает список документов, ранжированных  по местоположению ключевых слов в поисковом запросе, их частоте в тексте и другим параметрам. Имея схожий принцип работы, поисковые машины, тем не менее, различаются по   используемым алгоритмам и принципам поиска, которые к тому же постоянно совершенствуются; поэтому результаты  поиска у разных машин  отличаются.

В настоящее время существует большое количество поисковых средств.  Наиболее популярными среди нескольких сотен различного вида   "поисковиков" являются   поисковые средства следующих порталов.

**Google** (http://www.google.com/) мировой лидер по объему проиндексированных документов и скорости обработки запросов.   Google самая популярная поисковая система, обеспечивает поиск независимо от языка источника,  проста в использовании,  имеет  хороший язык запросов и   простой интерфейс.

    **Яndex** (http://www.yandex.ru/)  крупнейший российский портал, предлагающий пользователям  многочисленные услуги, включая поисково- информационные (12 служб). Робот поисковой системы Яндекс постоянно сканирует Интернет, автоматически отслеживая изменения; результаты поиска упорядочиваются в соответствии с установленными критериями релевантности (степени соответствия искомого и найденного). Поисковая система обеспечивает возможность расширенного поиска, позволяющего детализировать многочисленные параметры поиска.

**Rambler** (http://www.rambler.ru/) российский поисковый портал. Как профессиональная информационно-поисковая система Rambler  существует с 1996 года. Обеспечивает различные виды поиска, включая расширенный поиск по комбинации различных параметров с учетом морфологии русского языка.

**Aport!**  (http://aport.ru/) российский поисковый портал Апорт имеет  расширенные возможности по формулированию запросов; обеспечивает поиск графических изображений и мультимедийных файлов.

**AltaVista** (http://www.altavista.com/).  Поисковая система AltaVista появилась в конце 1995 года и до триумфа Google была мировым лидером  в области поиска (индекс системы уже тогда содержал более 20 млн. страниц). Предназначена для поиска информации в Интернет независимо от региона мира; обеспечивает поиск графики, аудио- и видеофайлов,  а также   возможность перевода на основные европейские языки; одним из первых начала поддерживать поиск на русском языке.

**Yahoo!** (http://www.yahoo.com/).  Как поисковая система является  наиболее авторитетным справочником ресурсов Интернета. Выдает по поисковой теме максимальное количество зарубежных сайтов;  обеспечивает поиск на русском языке.

       Для поиска информации применяются также метапоисковые системы. метапоисковая система обеспечивает поиск по нескольким поисковым системам сразу (до нескольких десятков).   Однако использовать метапоиск имеет смысл в  основном в случаях поиска конкретного документа  или  по очень узкой тематике.

       2.2.  **Язык поисковых запросов**

Поисковый запрос в общем случае может состоять из одного или нескольких слов,  логических операторов и знаков препинания. Простые запросы не требуют знания  **языка поисковых запросов**, так что если ввести в поисковую строку несколько слов без знаков препинания и логических операторов, то будут найдены документы, содержащие все эти слова, причем на произвольном  расстоянии друг от друга. Знание же языка запросов конкретной поисковой машины и правильное его применение позволяет сделать поиск быстрым и эффективным.

**Операторы языка запросов**

       Оператор **AND** (логическое **И;** сокращенное обозначение **«&»)** образуют сложный запрос, по которому будут найдены только те документы, которые одновременно содержат оба аргумента оператора. Например, по запросу: «**информационная AND технология»** будут найдены только те документы, которые содержат и слово «**информационная»**, и слово  **«технология»**. Аналогичный результат будет получен и при таком запросе: «**информационная & технология».**

**Примечание.** Оператор **AND** используется по умолчанию, так что запрос: «**информационная технология»** даст тот же результат, что и запрос: «**информационная AND технология».**

Оператор **OR** (**ИЛИ;** сокращенное обозначение **« | »**) образуют запрос, по которому будут найдены все документы, удовлетворяющие хотя бы одному  из аргументов оператора. По запросу: «**информационная OR технология»**  будут найдены  документы, которые содержат хотя бы одно из слов    «**информационная»** или **«технология»,** либо оба эти слова.

Оператор **NOT** (**И-НЕ,** сокращенное обозначение **«&!»**) образует запрос, по которому будут найдены   документы, удовлетворяющие левой части запроса и не удовлетворяющие правой. По запросу: «**информационная NOT технология»**  будут найдены  документы, которые содержат слово  «**информационная»** и не содержат слово **«технология».**

**Примечание.** Если сложный запрос включает несколько операторов, то он будет выполнен согласно традиционным приоритетам этих операторов. Изменить порядок выполнения можно посредством использования скобок.

**Кавычки**

Для поиска цитат или слов в заданной форме можно использовать двойные кавычки. Слова по той части запроса, которая заключена в двойные кавычки, ищутся в документах именно так, как они в этих кавычках представлены. Например, запросу «**информационная технология  “применяться”»** удовлетворяет документ, который содержащий текст **«…информационная технология будет применяться…»**, но не удовлетворяет документ, содержащий **«…информационная технология   применяется…»**. А на поисковый запрос **«”информационная технология   применяется”»** будут выданы только документы, содержащие с точностью до буквыименно такую комбинацию слов идущих подряд, при этом количество выбранных документов резко сократится.

**Примечание.** Так интерпретируютсядвойные кавычки всеми поисковыми машинами названных выше поисковых порталов за исключением **Апорта**, у которого оператор двойные (или одинарные) кавычки позволяет находить словосочетание, указанное в них, или близкое к нему, так что Апорт в поисковом запросе с двойными кавычками слова «**применяться»** и **«применяется»** различать не будет.

**Скобки**

       Скобки могут использоваться для построения вложенных запросов, изменения области действия операторов, а также для изменения  принятых по умолчанию приоритетов логических операторов. При использовании скобок часть запроса, которая заключена в скобки, интерпретируется как запрос, так что на нее распространяются правила языка запросов.

Например, по запросу «**данные** **AND информация OR сигнал»** поисковая  машина найдет документы, содержащие либо **«данные»** и **«информация»**, либо **«сигнал»**.  А на запрос «**данные** **AND (информация OR сигнал)»** будет найдено «**данные»** и одно из слов **«информация»,** или  **«сигнал».**

**Ограничение расстояния**

При простом запросе будут найдены документы, в которых встречаются все слова запроса не превышающие **ограничение расстояния** по умолчанию (например, у Рамблера оно равно расстоянию в 40 слов). Так что на запрос **«информация данные»** документ будет выдан только в том случае, если слова  **«информация»** и **«данные»** отстоят в этом документе не далее чем на 40 слов, включая и эти два слова.

       Значение **ограничения         расстояния** можно изменять, например, у **Рамблера** конструкцией **(n, запрос)**, где  **n** -  положительное число, **запрос** -  корректный поисковый запрос. Например, по запросу «**(2, информация данные)»**  будут выданы только те документы, в которых  слова  **«информация»** и **«данные»** стоят рядом хотя бы один раз.

Другие поисковые машины обычно  используют иные операторы ограничения  расстояния. Поисковые машины могут позволять задавать этот параметр и в меню расширенного поиска (см. например, справку по расширенному поиску у Яндекса).

         Приведенные выше элементы языка поисковых запросов используются, как правило, всеми поисковыми машинами. При этом поисковые машины  могут использовать и  другие конструкции, включая метасимволы испециальныеоператоры, так что за деталями следует     обращаться к справочной системе конкретной поисковой машины.

                        **3. Порядок выполнения работы**

          3.1. Ознакомьтесь  с материалом, изложенным в пункте 2, этой  работы.

          3.2. Включите закрепленный за Вами компьютер и получите у преподавателя  индивидуальное   задание.

          3.3. Ознакомьтесь, используя гиперссылки, с возможностями поисковых машин порталов,  указанных в пункте 2.1.

            3.4. Составьте  в соответствии с полученным заданием варианты поискового запроса.

                     3.5.  Осуществите в соответствии с составленными вариантами поискового запроса поиск  необходимых документов.

                  3.6. Проанализируйте  полученные результаты.

           3.7. Оформите отчет и защитите лабораторную работу.

         3.8. Выключите компьютер и приведите в порядок рабочее  место.

**4. Содержание отчета**

            4.1.  Цель работы.

       4.2.  Краткую характеристику использованных операторов и конструкций языка запросов.

       4.3. Описание результатов поиска документов по вариантам поискового запроса.

            4.4.  Анализ полученных результатов.

            4.5.  Выводы по работе.

**5. Контрольные вопросы**

              1. Как проинтерпретирует  поисковая машина Яндекса запрос ***“информационная технология”*** ?

            2. Выдаст ли **Апорт** на поисковый запрос ***“информационная технология”*** документы, содержащие слова ***информационные технологии***, но не содержащие слова  ***информационная технология***?

              3.  Каким образом можно изменить область действия логических операторов в поисковом запросе?

        4.  Перечислите основные логические операторы языка запросов.

              5.  Чем отличаются метапоисковые системы от поисковых машин?