

Консультационная сессия по работе с информацией (поиск и использование полученной информации для эффективного решения профессиональных задач)

– эффективные способы поиска информационных ресурсов необходимых для решения профессиональных задач.

Поиск информации является процессом выявления в определенном массиве текстовых документов тех данных, которые касаются конкретной темы и удовлетворяют указанным условиям, и в которых имеются необходимые сведения и факты (к примеру, вся необходимая информация по теме самообразования).

Состоит процесс поиска информации из нескольких последовательных этапов, посредством которых обеспечивается сбор данных, их обработка и предоставление. Как правило, поиск осуществляется следующим образом:

- определяется информационная потребность и формулируется запрос;
- определяется комплекс источников, в которых может находиться нужная информация;
- информация извлекается из выявленных источников;
- происходит ознакомление с данными, и оцениваются результаты поиска.

Поиск обычно осуществляется посредством браузеров или поисковых систем.

Браузеры. Браузер (от англ. слова browse - просматривать) представляет собой специальную программу для просмотра информации в Internet. Существует большое число разных браузеров, среди которых наибольшей популярностью пользуются Google Chrome, Яндекс браузер, Mozilla Firefox, Opera и Internet Explorer.

Браузеры хранят историю поиска по сети интернет на протяжении определенного времени. В любом из них есть система электронных закладок, которыми можно воспользоваться при необходимости.

Поисковые системы. Информационные технологии, обладающие высокой скоростью передачи информации и значительным объемом памяти, стали влиять на все сферы деятельности человека, в том числе на работу. Для поиска в интернете обычно требуется ввести запрос в строке поискового сервиса, нажать кнопку «Найти» и выбрать наиболее подходящий из предложенных вариантов – интернет-страниц.

Для правильного отбора источников информации необходимо стилистическое понимание источников информации, которых существует немалое количество. Представим лишь самые распространенные:

- научные исследования, имеющие под собой реальные доказательства, полученные эмпирическим путем;
- научно-популярные размышления, включающие в себя как фактические эмпирические данные, так и субъективные точки зрения людей, являющиеся специалистами в той или иной области;
- философские трактаты и рассуждения, отличающиеся наибольшей оригинальностью, субъективностью и формой подачи;

– художественная литература, служащая, как правило, источником информации – пищи для размышления, но не достоверных эмпирических данных;

– публицистические произведения – категория произведений, которые посвящены актуальным явлениям и проблемам текущей социальной жизни. Нередко в таких произведениях можно отыскать немало достоверных данных и фактов;

– средства массовой информации – комплекс органов публичной передачи информации, таких как телевидение, радио, журналы и газеты, а также Интернет.

Необходимость и важность проблемы информационного поиска привела к образованию в Интернет целой отрасли, задача которой заключается именно в оказании помощи пользователю в его навигации в веб-пространстве. Составляют эту отрасль специальные поисковые службы или сервисы. Условно их можно разделить на справочники (directories) и поисковые системы (search engines).

Эти разновидности внешне очень похожи, поскольку каждый справочник, как правило, обладает собственной поисковой системой, а каждая поисковая система – собственным справочником. Однако принципы их работы базируются на абсолютно разных подходах и технологиях. При этом каждая разновидность поисковых сервисов применяется для решения определенного типа задач. Правильным выбором инструмента во многом определяется стратегия поисковой деятельности и, в конечном итоге, результат поиска.

Поисковые средства первой группы, называемые также **каталогами** или **рубрикаторами**, представляют собой электронные **справочники**, имеющие четкую иерархическую систематическую или логико-тематическую структуру, во многом напоминающую структуру систематического каталога библиотеки. Работа со справочниками позволяет ориентироваться в ресурсах Интернет в пределах отдельных отраслей знания, углубляясь от общего к частному, меняя иерархические ветви, возвращаясь на несколько шагов назад, и т.д. К наиболее значительным и объемным справочникам относятся **Yahoo!** (www.yahoo.com), **About** (www.about.com). Среди российских разработок в этой области значатся **Апорт** (www.aport.ru), **List.ru** (list.mail.ru), **Weblist** (www.weblist.ru), **Пингвин** (www.pingwin.ru), **Иван Сусанин** (www.susanin.net), и другие.

Главной отличительной особенностью справочников является то, что все они сделаны «вручную». Редакционные коллегии каждого из справочников, которые по характеру труда более всего напоминают отделы каталогизации и систематизации крупных библиотек, регулярно просматривают содержимое вновь появившихся серверов и следят за изменениями на уже существующих. Выявленные данные анализируются и заносятся в разделы справочника в соответствии с принятой классификацией. Описание сервера в целом или его раздела, если он представляется вполне самостоятельным блоком, снабжается краткой аннотацией, содержащей самые общие сведения о характере имеющейся информации. В некоторых случаях добавляются сведения о языке документов, посещаемости ресурса, его физическом месторасположении, и т.п. Естественно, что подобным образом можно описать лишь часть ресурсов Сети, да и то только весьма крупные объекты – такие как сервера организаций и учреждений. По причине огромной трудоемкости такой работы, справочники отражают лишь

незначительную часть ресурсов Сети, и в первую очередь - наиболее заметные сайты.

Справочники предназначены для решения трех типов задач:

- ориентация в незнакомой отрасли знания;
- поиск крупных объектов, каковыми являются, к примеру, серверы целых организаций или значительных проектов;
- получение готового перечня ресурсов, имеющих размытый поисковый образ, например, «все о теннисе», библиотек определенного типа, транспортных расписаний или сайтов различных политических партий.

Созданием и актуализацией справочников глобального масштаба в большинстве случаев занимаются информационные компании, прежде всего американские. Практически все справочники декларируют всемирный охват материала, однако, практика показывает, что основной упор в них делается все-таки на североамериканские сайты. Естественно, что обращение к глобальным справочникам ресурсов требует минимального владения английским языком.

Рассмотрим в качестве примера два каталога-справочника – зарубежный (Yahoo!) и отечественный (Aport).

Yahoo! (<http://www.yahoo.com>). Yahoo! является одним из самых первых, надежных и авторитетных справочников Всемирной паутины. У каталога два основных достоинства - внушительный объем (на сегодня отражено порядка 2 миллионов сайтов в более чем 25 тысячах категорий) и научность и логичность используемой схемы классификации. За счет этого в справочнике объективно отражаются все отрасли знания без каких-либо приоритетов. В Yahoo! учитываются не только веб-сайты, но и конференции UseNet. Эти достоинства на протяжении долгого времени позволяли Yahoo! оставаться незаменимым инструментом для первоначального ознакомления с информационным наполнением в Интернет по любой области деятельности.

Апорт (<http://aport.ru>). В настоящее время является единственным профессионально поддерживаемым отечественным справочником Интернет-ресурсов. Данный каталог выступает ключевой составной частью одноименного портала, включающего также поисковую систему, характеристики которой рассмотрены ниже.

Справочник имеет многоуровневую иерархическую структуру, отличающуюся достаточной логичностью и продуманностью. Ее полный вид можно получить, щелкнув по пункту верхнего меню «Все рубрики».

В основу работы поисковых систем (поисковых машин) заложены совершенно иные технологические принципы. Задача поисковых машин – вести поиск по содержанию каждого веб-документа, то есть проводить максимально детальный поиск информации в Интернет. В отличие от справочников, все поисковые машины имеют одинаковый принцип деятельности и функционируют в автоматизированном режиме.

Все поисковые системы состоят из двух базовых блоков. Первый блок представляет собой программу-робот, задача которого путешествовать с сервера на сервер, находить там новые или изменившиеся документы и скачивать их на главный компьютер системы. Робот, просматривая содержимое сервера, также находит новые ссылки, как на другие документы данного сервера, так и на внешние сайты. Программа самостоятельно направляется по указанным

ссылкам, после чего процесс повторяется. Выявленные документы обрабатываются (индексируются) вторым блоком поисковой системы. При этом, как правило, учитывается всё содержание страницы, включая текст, иллюстрации, аудио и видео файлы, и пр. Индексации подвергаются все слова в документе, что как раз и дает возможность использовать поисковые системы для детального поиска по самой узкой тематике. Образующие гигантские индексные файлы, хранящие информацию о том, какое слово, сколько раз, в каком документе и на каком сервере употребляется, и составляют базу данных, к которой происходит обращение пользователей, вводящих в поисковую строку сочетания ключевых слов.

Среди внушительного числа поисковых систем выделяются несколько наиболее признанных, позволяющих выявлять информацию с высокой степенью полноты и надежности. К наиболее авторитетным поисковым системам всемирного масштаба в настоящее время относятся Google (www.google.com), AlltheWeb (www.alltheweb.com) и AltaVista (www.altavista.com). Российские системы этого класса ныне представлены сайтами Яндекс (www.yandex.ru), Апорт (www.aport.ru) и Rambler (www.rambler.ru).

Поисковые системы распространены в гораздо большем количестве, чем электронные справочники, и число их, составляющее сегодня нескольких десятков, продолжает неуклонно увеличиваться. Профессиональная работа с ними требует специальных навыков, поскольку простой ввод искомого термина (ключевого слова) в поисковую строку скорее всего приведет к получению списка из сотен тысяч документов, содержащих данное понятие, что практически равносильно нулевому результату.

Google (www.google.com). Эта поисковая система запущена в 1998 году и является ныне единоличным лидером среди глобальных поисковых систем по всем значимым параметрам. Главное достоинство Google - объем его индексного файла, который составляет на сегодня более 3 миллиардов веб-страниц и статей из групп новостей по интересам. В сутки программы-роботы системы индексируют порядка трех миллионов новых и обновленных страниц, при этом актуализация базы производится каждые 28 дней.

Google отличается высокой степенью комфорта для пользователя. Несмотря на то, что это глобальная поисковая система, пользователи из неанглоязычных стран автоматически переадресовываются на интерфейс на их родном языке. Русскоязычный интерфейс, в частности, находится по адресу www.google.com.ru. Длительность процесса в большинстве случаев не превышает одной секунды, несмотря на огромный объем индексного файла системы.

Яндекс (www.yandex.ru или www.ya.ru). Запущен в сентябре 1997 года. В настоящее время - признанный лидер российского поискового сервиса. Демонстрирует высокие показатели как по объему проиндексированных документов, так и по релевантности поиска. Декларируется еженедельное обновление базы. Интерфейс максимально прост - состоит из единственной строки ввода ключевых слов. За счет встроенной системы морфологической обработки терминов Яндекс приспособлен для формирования запросов на естественном русском языке. Лингвистические инструменты позволяют автоматически учесть практически все возможные оттенки употребления

ключевых слов и составить поисковое предписание в высшей степени широко, охватив все возможные сочетания терминов. В процессе обработки запроса поисковый механизм самостоятельно производит расширения, исключает стоп-слова, анализирует расстояние терминов друг от друга, и пр. Типичный запрос в этом случае задается путем ввода отдельных терминов или целой фразы в поисковую строку.

Для формирования более точного запроса целесообразно обратиться к «Расширенному поиску» (пункт в нижней части титульной страницы). С помощью структурированного меню можно легко задать ограничения по различным сочетаниям ключевых слов, местоположению термина в документе, времени и языку публикации, месте на сайте.

Апорт (www.aport.ru). Поисковая система, запущенная в феврале 1996 года, ныне по праву занимает место в лидирующей группе отечественных навигационных сервисов. Объем проиндексированных документов на настоящий момент точно неизвестен, но тесты позволяют поставить ее на третье место по объему индексного файла. Особенностью Апорт является то, что он строго не ограничивает своих роботов пределами доменов России и СНГ - в перечне результатов можно встретить ссылки на многие зарубежные сервера.

Апорт снабжен массой различных возможностей, заставляющих отнести его к числу самых удобных для пользователя. Именно в нем одним из первых в российской практике был применен сложный язык запросов (<http://aport.ru/help.htm>), позволяющий максимально детализировать поисковое предписание и избавить пользователя от информационного шума, возникающего при случайном сочетании ключевых слов. Некоторые дополнительные опции ныне доступны при обращении к «Расширенному поиску».

Список результатов содержит сведения об адресе, дате опубликования и последней проверке документа. Документы с одного сайта сгруппированы вместе. Доступна ставшая уже привычной функция реконструкции текста без обращения и исходному документу. Положительным моментом является то, что найденные ключевые слова выводятся в окружающем контексте из любой части документа, а не только из его начала, что позволяет окончательно и точно определять соответствие страницы запросу уже на этапе просмотра ссылок.

Мета-поисковые машины. Отличия в стратегии и широте охвата материала различных систем часто приводят к тому, что разные поисковые средства дают разноречивые ответы на один и тот же запрос. Этим не замедлили воспользоваться разработчики поисковых орудий особого рода, основанных исключительно на использовании потенциала других поисковых систем. Все они получили название мета-поисковые машины. К наиболее известным их представителям, в частности, относятся MetaCrawler (www.metacrawler.com), KartOO (www.kartoo.com), Cyber 411 (www.cyber411.com), и некоторые другие.

Мета-поисковые системы не имеют собственных поисковых роботов и баз данных (индексных файлов). Их достоинство заключается в умении рассылать вводимые в них запросы по другим системам, а затем суммировать результаты. Таким образом, пользователь, вводя поисковое предписание, например, в Cyber 411, фактически одновременно обращается к десятку поисковых систем. Этим гарантируется «объективность» полученных результатов, однако, учитывая

различия в подходах к обработке терминов разными системами, результат может оказаться не всегда релевантным запросу.

Поиск информации. Все поисковые системы реализуют несколько алгоритмов поиска:

1. Простой поиск. В поле запроса вводится одно или несколько слов, которые могут характеризовать содержание документа. Если это слово одно, то в ответ выдается большое количество ссылок. Если несколько слов, то результат зависит от того, как эти слова введены, что, в свою очередь, зависит от каждой конкретной системы.

2. Расширенный поиск подразумевает запрос из группы слов. Слова связываются логическими операторами «И», «ИЛИ», «НЕ» и др.

3. Контекстный поиск – реализован не во всех поисковых системах. Если этот метод есть, то ключевая фраза должна быть заключена в кавычки.

4. Специальный поиск позволяет найти дополнительную информацию. Например, такие команды позволяют определить, как часто в Сети встречаются гиперссылки, указывающие на какой-то ресурс, с их помощью можно найти ключевые слова, входящие в заголовки веб-страниц и т.п.

Особенности группировки слов в поисковых системах:

1. Поисковые системы по-разному трактуют группы слов, введенные через пробел, как «И» или как «ИЛИ». В большинстве систем в качестве оператора «И» используется «+» перед словом без пробела. Большинство российских поисковых систем по умолчанию считают, что два слова должны присутствовать в документе. На Яндексе - в одном предложении. «ИЛИ» в системах задается так: Апорт – ИЛИ, Рамблер – OR, Яндекс - &.

2. Прописные буквы. «хлеб» = «ХЛЕБ», но «ХЛЕБ» ≠ «хлеб». Если введены строчные символы, то разыскиваются как строчные, так и прописные символы, но если использованы прописные, то ищется точное совпадение с прописными.

3. Зарезервированные слова – это слова, которые не учитываются при обработке запроса. К ним относятся неинформативные слова: предлоги, союзы, местоимения, артикли и др. слова малого размера. В некоторых системах зарезервированными могут быть слова, которые часто встречаются, поэтому информативными не являются. Например, в системе, ориентированной на поиск книг слово «книга» будет не информативным. При контекстном поиске необходимо точное (!) соответствие между заказом и результатом поиска. Если поисковая система «зачистила» Web-документы от зарезервированных слов, то с контекстным поиском она справиться не может. На Яндексе и Рамблере только делают вид контекстного поиска (через кавычки). В России честный контекстный поиск проводит только Апорт, но у него невелика база указателей.

4. Формы поиска. С 1997 г. поисковые системы предоставляют услуги по розыску данных, записанных в разных форматах: рисунки, видеофайлов, звуковые клипы и т.д. Для этого следует включить переключатель, соответствующий типу разыскиваемых данных. Яндекс позволяет найти тексты, товары и рисунки, Апорт – плюс MP3, Рамблер – только тексты.

5. Действия после поиска. Если представлен объемный документ, то можно воспользоваться кнопкой Правка – Найти на этой странице.

6. Приемы поиска в системе «Яндекс». Поиск по одному слову ведется на основе корня этого слова. Например, если введено слово «снег», то система выдаст документы, в состав которых входят однокоренные слова. Если поиск словоформ не требуется, то его отменить с помощью «!», например, !снегом.

Поиск по группе слов:

– Знак «+» пишется слитно со словом без пробела.

– Можно ввести оператор И (&), справа и слева от & должны быть пробелы.

– Если требуется присутствие слов не только в предложении, но и во всем документе, применяется оператор &&.

– «~» - строгое исключение из предложения

– «~~» - исключение из всего документа

– Пример: «Москва &Петербург&&+столица»

Поиск с указанием расстояния:

– «/±n», «/(-n+n)» Оператор NEAR позволяет находить документы, в которых искомые слова находятся близко друг от друга. Например, «Город/+1Москва» = «Город Москва».

– «/±n &&» или «/±n~~» , например, «Город/+1&&Москва» - словосочетание будет ли в одном предложении или в соседних.

Использование скобок. Служат для управления порядком действий. «Москва &Петербург& (столица &город-герой)».

Управление ранжированием производится с использованием весовых коэффициентов. Его можно присваивать любому ключевому слову или выражению, если оно заключено в скобки. «Москва:5 Петербург столица».

Ввод уточняющего слова осуществляется после знаков <-, например «Николай +Еременко<-младший артист+театра».

Современные стандарты образования ориентированы на формирование у специалистов профессиональных компетенций, к которым относится и умение использовать информационно-коммуникационные технологии в профессиональной деятельности. Поэтому поиск и является таким важным условием в развитии данных компетенций.

К сожалению, в последнее время все поисковые системы используют технологии SEA-продвижения, что означает коммерциализацию поиска. Следствием является затрудненность поиска серьезной научной информации посредством простого (через браузер) поиска в сети интернет. Выходом может стать поиск в электронных библиотеках, научных базах цитирования и прочее.