

Методика информационного поиска

Центральная задача информационного поиска (ИП) – удовлетворить информационную потребность. Информационная потребность формулируется как некоторый запрос, представляющий собой набор ключевых слов, характеризующий то, что ищет пользователь.

Запрос – это формализованный способ выражения информационных потребностей пользователем системы.

Информационный запрос в широком смысле – текст, выражающий информационную потребность. Информационный запрос в узком смысле – входное сообщение в автоматизированную систему, содержащее требование на выдачу информации.

Важнейшей проблемой, связанной с проведением эффективного с точки зрения пользователя ИП, является разработка поисковых стратегий.

Практика выработала универсальный алгоритм поисковой деятельности (совокупность операций, выполняемых в строго установленном порядке дискретными — прерывистыми — «шагами»):

1. Информационная подготовка поиска. На этом этапе уточняется запрос, дополняются исходные данные о предмете поиска. Вместе с информационным консультантом (чаще всего библиотечным работником) или самостоятельно определяется тип поисковой задачи, используется опыт решения аналогичных задач.

2. Моделирование источников поиска. На этом этапе выявляется идеальный источник информации, который бы полностью отвечал характеру поисковой задачи. Если в библиотеке есть информационный источник, соответствующий виду запроса, поиск можно считать успешно завершённым. Однако чаще всего сложные запросы требуют обращения к многим полезным источникам для сплошного поиска, поскольку один источник не даёт релевантной информации. В этом случае исследуются все информационные источники данной информационно-поисковой системы.

3. Выбор оптимального пути поиска. Среди имеющегося множества информационных источников ищем близкие к характеру поисковой задачи. На этом этапе используются каталоги, картотеки, справочные пособия, библиографические указатели, списки новых поступлений, базы данных, справочный аппарат первичных документов.

4. Реализация поиска — это этап получения конкретного ответа в соответствии с запросом, просмотр информационных источников, выявление и отбор нужной информации из этих источников.

5. Оценка результатов поиска — этап, на котором проверяются точность, качество проведенного информационного поиска. Пользователь должен быть уверен в том, что если информация не найдена, её нет в данном информационно-поисковом массиве. Определение результатов поиска — очень сложный этап, он требует высокого, подчас профессионального уровня информационной деятельности, знания методики информационного процесса, средств выполнения запроса. На данном этапе более всего нужна помощь посредника — специалиста информационной службы. На первом этапе — информационной подготовки поиска — уточняется запрос, то есть обращение пользователя в информационную службу, отражающее потребность в информации.

Формулировка информационного запроса должна отражать действительные потребности и интересы его автора. В запросе важны ясность образа предмета потребности, представление о конечном результате решаемой информационной задачи

По характеру необходимой информации запросы подразделяются на адресные, тематические, уточняющие, фактографические. Исходя из вида информационного запроса осуществляется, соответственно, адресный, тематический, фактографический, уточняющий виды поиска. В свою очередь, вид информационного поиска определяет

моделирование источников и оптимальный путь поиска, близкие характеру информационного запроса.

Рассмотрим виды информационного поиска в единстве методов осуществления и источников информации.

Адресный поиск — это поиск данных о наличии и/или местонахождении, точном адресе хранения документа. Предположим, вам захотелось прочитать книгу Булгакова «Мастер и Маргарита». Самый необременительный путь поиска — прийти в библиотеку и там искать данную книгу. Такой поиск сведений о наличии и местонахождении конкретного документа часто и называется библиотечным. В библиотеке можно вовсе не прикладывать собственных усилий, а сделать запрос библиотечному работнику, который найдет книгу и выдаст ее вам. Но если вы дорожите своей самостоятельностью и сознательно развиваете, повышаете информационную культуру, то сами проделаете эту достаточно простую процедуру. Надо только знать, что в библиотеке имеется специальный справочно-библиографический аппарат (СБА), важнейшей функцией которого является обеспечение информационного поиска.

В СБА входят: библиотечные каталоги, библиографические и фактографические картотеки, библиографические базы данных; справочно-библиографический фонд (СБФ), состоящий из официальных, нормативных, справочных и библиографических изданий; фонд неопубликованных библиографических пособий, созданных самой библиотекой. Все элементы СБА создаются в традиционной и электронной формах.

Какой из этих элементов СБА надо выбрать как источник поиска книги Булгакова? Есть вероятность, что в каждом элементе можно найти нужные сведения. Но главным источником здесь является библиотечный каталог (алфавитный). Поскольку вам известны фамилия автора и заглавие книги, вы находите в алфавитном каталоге карточку с фамилией «Булгаков». Но есть определенная трудность: вы не знаете инициалов писателя. Следовательно, вам надо провести не только адресный, но и уточняющий поиск. Вам поможет знание заглавия книги. По нему среди нескольких карточек с фамилией «Булгаков» и разными инициалами вы легко найдете карточку с описанием нужной книги — «Мастер и Маргарита» — и с указанием инициалов ее автора.

Шифр книги в левом верхнем углу каталожной карточки является ключом, с помощью которого библиотечный работник найдет роман в книжном фонде библиотеки.

Если вы не нашли в каталоге карточку с описанием романа М. А. Булгакова, это еще не значит, что данного романа в библиотеке нет. Очень возможно, что он вошел в сборник или собрание сочинений писателя.

Тематический поиск (поиск документов по узким и широким темам запроса) требует владения более сложными операциями с большим количеством документов. Источниками тематического поиска являются самые разные, многочисленные информационные ресурсы. Достаточно полно обеспечивают тематический поиск СБА библиотек. При этом успех зависит от ресурсов конкретной библиотеки. Поэтому со сложным тематическим запросом лучше обращаться в крупную библиотеку, имеющую богатые фонды и хорошо организованный СБА (республиканскую, краевую, областную и т. п.). Это особенно важно при узких темах запроса, когда искать литературу намного трудней, чем по широким темам. Так, легче подобрать материал по истории русской литературы XVIII века, чем только о развитии одного жанра поэзии в этот период. А если, например, будущий литературовед пишет научную работу о ссыльных поэтах-декабристах, ему предстоит длительный многоуровневый поиск.

Традиционными источниками тематического поиска являются систематический и предметный каталоги библиотек, картотеки, библиографические пособия.

В систематическом каталоге карточки с библиографическими записями (сообщениями о первичном документе в особой документной форме) систематизированы по отраслям знания в соответствии с определенной системой библиотечно-библиографической классификации (ББК).

Предметный каталог доступнее и проще, его удобно использовать в дополнение к систематическому при запросах на узкие темы. Записи в предметном каталоге располагаются в алфавитном порядке предметных рубрик.

Прямыми источниками тематического информационного поиска являются библиографические пособия.

Важнейшее достоинство библиографического источника — возможность быстрого нахождения нужного документа благодаря наличию «ключей» — вспомогательных указателей в пособии. Вспомогательные указатели бывают самые разные — имен, авторов, заглавий, тематические, предметные, географические и т. п. С их помощью легко и быстро можно узнать, какие книги и статьи посвящены определенной теме, предмету, кто автор книги, кто переводчик, произведения каких авторов вошли в издание и т. д.

Фактографический поиск — это поиск какого-либо конкретного факта, данных, относящихся к какому-либо предмету, процессу, событию; поиск терминов, законов, дат, адресов, правил правописания и т. д., и т. п. Конечным результатом фактографического поиска является не документ, не список документов, а ответ по существу. В каком году произошла битва при Ватерлоо? Когда родился Эйнштейн? Что такое «спам»?

Как и при тематическом, и при адресном поиске, источниками фактографического поиска могут стать различные составляющие информационных ресурсов. Но особое значение здесь имеют справочные документы, специально собирающие наиболее важные и часто запрашиваемые сведения с целью их быстрого разыскания. Важнейшие виды справочных документов: энциклопедии, энциклопедические словари, справочники, путеводители, словари, хроники и календари, летописи жизни и творчества. Электронные фактографические ресурсы (фактографические базы данных) аккумулируют (собирают) в себе разные типы справочных документов, стремятся к целостному представлению информации

Уточняющий поиск — это восполнение или уточнение неполной или искаженной информации в документах. В информационных запросах содержатся типичные ошибки, связанные с искажением фамилии и инициалов индивидуального автора документа; сокращения и неточности в заглавии книг, периодических изданий; представление редактора, переводчика, издателя как автора; представление названия раздела, главы как названия всей книги и т. п.

Уточняющий поиск — один из важных видов информационного поиска. В его процессе часто ставятся специфические поисковые задачи, решить которые может только специалист — информационный работник. Уточнение неопределенных запросов является и чрезвычайно сложной психологической, а не только информационной процедурой, которая тесно связана с проблемой формирования информационной культуры личности.

В практике информационной деятельности все виды поиска взаимосвязаны. Одна поисковая задача может переходить в другую, любой информационный источник может использоваться для любого вида поиска. И, как правило, очень редко сложный информационный поиск можно провести по какому-либо одному из источников, одному из элементов ресурсов. Развитие сетевых коммуникаций может изменить соотношение источников поиска. Однако меняется форма, но не сущность процесса разыскания. Следует искать наиболее простой и эффективный способ решения каждой поисковой задачи, исходя из доступных средств и экономии времени. Традиционный, ручной и автоматизированный поиск, использование сети Интернет — все это составляющие единого процесса информационного поиска.

Т. о. при осуществлении ИП важно определить для себя технологию и алгоритм выполнения запросов: вести поиск традиционным путем, воспользоваться локальными БД или Интернет, и определить, каким образом это лучше осуществить (продумать методику). Умение вести эффективный информационный поиск является показателем высокого уровня информационной культуры личности. В целом культура информационного поиска дает пользователю следующие преимущества:

зная структуру информационного источника и структуру поискового образа документа, можно самостоятельно ориентироваться в источниках и отражаемых источниками описаниях документов, лично, без помощи консультанта вести информационный поиск; зная поисковые возможности элементов информационных ресурсов, умея в них ориентироваться, можно самому владеть стратегией каждого вида поиска; зная видовую классификацию информационного поиска, можно выбрать оптимальные информационные источники для его конкретной разновидности; зная структуру информационно-поисковой системы и владея универсальным алгоритмом поисковой деятельности, можно вычислить точность, полноту, оперативность поиска и самому судить о соответствии полученной информации информационному запросу.

Информационный поиск в Интернете

Интернет открыл доступ ко многим информационным ресурсам, получить которые другим способом в настоящее время невозможно (или слишком дорого). Объединяя локальные, национальные и глобальные сети, он стал громадным информационным ресурсом, наиболее динамичным из всех существующих в настоящее время.

Весьма важной чертой Интернета является то, что пользователь информации сам производит поиск, ведет диалог с поисковой системой, меняя тактику поиска в зависимости от полученных результатов. В связи с этим можно выделить два положительных момента:

- формируются основы научного мышления, навыки исследовательской работы: пользователь учится ориентироваться в сетевом пространстве, отбирать из большого количества документов то, что действительно необходимо и, наконец, варьировать свой запрос в зависимости от анализа уже выявленных документов;
- выявляется информация, которая не только соответствует запросу, но и может оказаться полезной в будущем.

Тем не менее, пользователи Интернета часто не получают удовлетворения от результатов поиска в новой информационной среде. Обычно это происходит по следующим причинам:

- как правило, пользователи в качестве инструмента поиска используют либо поисковые машины (AltaVista, HotBot), либо каталоги ресурсов (Yahoo). Несмотря на ряд достоинств, вышеупомянутые поисковые системы не обеспечивают научный контроль качества (т.е. процесс включения информации в БД этих поисковых систем не находится под контролем специалистов, отвечающих за точность, качество информации по этой тематике), что может привести к получению в результате поиска недостоверной информации;
- отсутствие знаний и навыков по составлению запроса к поисковой системе приводит к получению большого количества информационного шума. Информационный шум – сведения, которые не несут полезную информацию.

Таким образом, можно утверждать, что в Интернете есть больше возможностей, нежели используется в настоящий момент. И, следовательно, проблема заключается не в недостаточности информации в Интернете, а в неумении пользователя ее найти [4]. Поэтому необходима правильная методика поиска в Интернете и знание возможностей поисковых машин, каталогов и баз знаний. Рассмотрим их по отдельности.

Методика поиска в Интернете. Поиск информации в новой информационной среде мы рассматриваем больше как процесс решения поисковой задачи, стоящей перед пользователем, нежели как просто нахождение релевантной запросу информации. Релевантная информация – информация, актуальная для конкретной задачи, цели и периода времени.

Методика поиска в Интернете разбивает процесс поиска на 3 этапа [6].

Основными этапами являются:

1. Формулировка и уточнение информационного запроса.
2. Планирование поискового процесса.
3. Реализация поиска. Окончательное решение поисковой задачи.

Этап 1. Формулировка и уточнение информационного запроса

Этап формулировки и уточнения запроса традиционно играет ведущую роль. От того, насколько верно сформулирован запрос, будет зависеть выбор информационных ресурсов и результат поиска. На данном этапе целесообразно сделать следующие шаги:

ШАГ 1. Формирование представления об искомой информации.

Для того чтобы получить это представление, обладателю информационной потребности необходимо как можно более точно выразить ее на естественном языке.

ШАГ 2. Определение типа информационной потребности.

Даже достаточно грубое представление о типе информационной потребности позволяет уточнить обуславливаемые ею требования, которые следует учитывать при поиске. Традиционно различают два типа информационных потребностей:

- границы информационной потребности четко определены. Например: «Какие учебные курсы делового английского были опубликованы в России в 2001 г.?», «Автомобили каких марок выпускались в России в 1966 г.?». В данном случае наилучшей информацией признается та, которая наиболее полно и точно описывает конкретную информационную потребность данного потребителя и позволяет получить точный ответ;

- границы определены нечетко, а также с течением времени могут изменяться. Например: «Методики преподавания делового английского», «Изучение информационных потребностей». В данной ситуации потребитель будет ощущать комфорт при возможности самому разобраться в информации.

Еще раз напомним, что процессы информационного поиска, с помощью которых удовлетворяются информационные потребности разных типов, в целом отличаются друг от друга, поэтому существуют разные виды информационного поиска. **Об информационном поиске, удовлетворяющем информационные потребности первого типа, говорят как о фактографическом. Об информационном поиске, удовлетворяющем потребности второго типа, – как о документальном.** Однако часто, чтобы найти некоторые факты в Интернете, необходимо сначала провести поиск документов (веб-страниц), содержащих эти факты. **В этом случае можно говорить о документально-фактографическом поиске.**

Использование поисковых машин, каталогов и баз знаний.

Каталоги сайтов (справочники)

Каталог сайтов или справочник (Web-directory, синонимы: Директория, Каталог, Рубрикатор) - систематизированный перечень Интернет-ресурсов, имеющий иерархическую структуру и, фактически, представляющий собой систематический каталог Интернет-ресурсов. **Это список адресов сайтов, сгруппированных по темам и снабжённых краткими описаниями.** Объектом описания являются в большинстве случаев целые сайты или их большие разделы. Описания ресурсов включают название сайта, его адрес (URL) и аннотацию. Работа по классификации и аннотированию ресурсов производится вручную редакторами разделов. Каталоги делятся на универсальные и специализированные (отраслевые), а также на глобальные и национальные (региональные). С помощью справочников выполняется поиск ресурсов (сайтов), определенной тематики путем перехода от более общих его разделов, к более частным. Пример глобального справочника универсального характера: Yahoo!.

Как устроены каталоги. Информация о сайтах сгруппирована в каталогах по разделам. Каждый раздел имеет несколько подразделов. Кроме основных разделов многие каталоги имеют дополнительные, в которых сайты классифицированы по регионам, странам, сектору экономики. Внутри разделов сайты отсортированы по алфавиту. Встречается сортировка по посещаемости, дате создания, рейтингу, экспертной оценке. К названиям сайта прилагаются краткие описания, написанные редакторами.

К наиболее значительным справочникам всемирного масштаба относятся Yahoo! (www.yahoo.com), Open Directory (dmoz.org), About (www.about.com). Среди российских разработок в этой области значатся Апорт (www.aport.ru). Среди белорусских --

<http://www.akavita.by/>, <http://www.all.by/>, <http://www.br.by/>, <http://catalog.tut.by/>, <http://www.poisk.com/> каталог + поиск по белорусскому интернету и др.

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, снабжаются 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, например, сервера организаций и учреждений. По причине огромной трудоемкости такой работы, справочники отражают лишь незначительную часть ресурсов Сети, в первую очередь наиболее заметные сайты.

Основными параметрами, характеризующими достоинства справочников являются: объем; оперативность отражения новых или изменившихся ресурсов; логичность и последовательность иерархической схемы классификации; перекрестность структуры.

Объемом справочника определяется степень его надежности или "информационная прочность". От этого зависит уровень доверия, который пользователи питают к конкретному каталогу. Так как материалы в Интернет появляются, изменяются, а порой и бесследно исчезают ежедневно, важно то, насколько оперативно появляются ссылки на новые сайты и исправляются или удаляются ссылки на устаревшие ресурсы. В некоторых системах существует специальный механизм, периодически проверяющий доступность сайта и исключаяющий его из перечня при долгом "отсутствии" в Сети. Логичностью (научностью) применяемой схемы классификации определяется степень простоты, с которой пользователи могут находить требуемые сведения. Система же перекрестных ссылок позволяет выявлять информацию, используя разные подходы (например, территориальный или отраслевой). В этом случае схема классификации должна автоматически выводить пользователя на искомый объект, какой бы путь поиска не был выбран.

Возможности составления запроса для этого вида поисковых средств особой роли не играют, так как сложные разыскания, требующие большой детализации запроса, с помощью каталогов не проводятся.

Справочники предназначены для решения трех типов задач:

1. ориентация в незнакомой отрасли знания;
2. разыскание крупных объектов, каковыми являются, к примеру, серверы целых организаций или значительных проектов;
3. получение готового перечня ресурсов, имеющих размытый поисковый образ, например, клиник пластической хирургии, библиотек определенного типа, транспортных расписаний или сайтов различных политических партий.

Следует помнить, что, обращаясь к справочникам, **мы изначально можем рассчитывать на получение лишь очень общих сведений по тематике, и никогда - детальных данных.**

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. Эти достоинства на протяжении всей второй половины 1990 годов позволяли Yahoo! оставаться

незаменимым инструментом для первоначального ознакомления с информационным наполнением в Интернет по любой области деятельности.

Интерфейс справочника Yahoo. Сегодня его нелегко разглядеть среди множества других сервисов, предоставляемых компанией.

Основной ряд Yahoo! включает 14 категорий, в числе которых: БИЗНЕС И ЭКОНОМИКА, КОМПЬЮТЕРЫ И ИНТЕРНЕТ, НОВОСТИ И СМИ, РАЗВЛЕЧЕНИЯ, ИСКУССТВО И ГУМАНИТАРНЫЕ НАУКИ, ОБРАЗОВАНИЕ, ПОЛИТИКА, ЗДРАВООХРАНЕНИЕ, ЕСТЕСТВЕННЫЕ НАУКИ, ОБЩЕСТВЕННЫЕ НАУКИ и т.д. Использование Yahoo!, как и любых других справочников, наиболее эффективно для ознакомления с наполнением Интернет по различным областям деятельности, при размытости критериев поиска - когда неизвестно по каким ключевым словам осуществлять поиск. Разыскание осуществляется путем последовательного просмотра категорий: шаг за шагом, разворачивая пункты меню, построенные в иерархическом порядке, можно знакомиться с полным перечнем ресурсов, постепенно сужая и конкретизируя тему. Так, от самого общего раздела СПРАВОЧНЫЕ РЕСУРСЫ (REFERENCES) можно дойти до web-сервера конкретной библиотеки, миновав по пути разделы БИБЛИОТЕКИ, БИБЛИОТЕКИ ОПРЕДЕЛЕННОЙ СТРАНЫ, БИБЛИОТЕКИ ОПРЕДЕЛЕННОГО ТИПА. На каждой ступени иерархии после названий разделов в скобках указывается число отраженных в них ресурсов. При большом количестве объектов в одном разделе пользователям предлагается выбрать первую букву названия сайта (например, названия университета или фамилии популярного исполнителя).

Учитывая гигантскую популярность и, следовательно, посещаемость Yahoo!, в последние несколько лет его владельцы уделяют все большее внимание побочным сервисам, среди которых электронные магазины, аукционы, службы новостей, агентства путешествий и т.п. Собственно справочник занимает все более и более скромное место в интерфейсе титульной страницы сервера. Кроме того, с марта 2004 года за включение сайта в справочник взимается не только разовая, но и ежегодная абонентская плата - 300 долларов. Это не только затрудняет включение в Yahoo! новых сайтов мелкого бизнеса, но и приводит к исключению из каталога сайтов, не уплативших ежегодный взнос. Естественным следствием этого является то, что содержание справочника постепенно все менее соответствует реальному наполнению Интернет. Эксплуатация известной торговой марки без поддержания высокого качественного уровня основного продукта уже обернулась снижением авторитета Yahoo! как ведущего справочника, и если в ближайшее время в этой области не будут предприняты кардинальные шаги, его популярность будет снижаться, несмотря на громкое имя.

Поисковые машины

Поисковая машина- (Search engine, синонимы: поисковая система, поисковый сервер, “искалка”) - инструмент для поиска информации в Интернет. Поисковые машины состоят из двух основных частей: программы-бота и индексного файла (базы данных). Робот сканирует сайты Интернет, выгружая web-страницы на базовый компьютер, где производится их индексирование. Учитываются не только тексты, но и иллюстрации, а также аудио- и видеофайлы. Пользователь, вводя состоящий из ключевых слов запрос, обращается к индексному файлу и получает в ответ перечень страниц, содержащих искомые термины. По территориальному охвату поисковые машины различаются на глобальные и локальные (национальные или отраслевые).

К основным параметрам поисковых систем относятся:

1. объем индексных файлов (число проиндексированных серверов и отдельных документов); Первая величина, являющаяся ключевой, устанавливает широту охвата материала и определяется числом проиндексированных документов. Сейчас эта цифра для лидеров мирового сетевого поиска составляет более триллиона документов.
2. степень оперативности обновления базы данных за счет включения сведений о новых материалах и удаления устаревших; Учитывая тот факт, что в среднем интернетовский

адрес сохраняет актуальность до полугода, после чего документ или меняет местоположение или убирается с сервера, большое значение имеет уровень оперативности обновления данных, характеризующий степень соответствия индексного файла поисковой системы реальному местоположению документов на сайтах. В настоящее время этот параметр колеблется от двух недель до полутора месяцев.

3. возможности для составления запроса; Возможности поискового механизма выражать запрос максимально точно в значительной степени предопределяют долю релевантных документов в перечне полученных результатов. Каждая машина имеет свою собственную лексику, которая по-разному позволяет детализировать поисковое предписание.
4. интеллектуальность системы ранжирования результатов поиска; Все поисковые машины обладают модулем ранжирования результатов поиска. Создание таких модулей - целая область программирования, в которой конкурируют сложнейшие алгоритмы, созданные разными компаниями. Перечень факторов, принимаемых во внимание при определении места документа в перечне ссылок, необычайно широк: от местоположения слова на странице до рейтинга (авторитета) страниц, имеющих ссылки на найденный документ
5. наличие дополнительных сервисных функций, облегчающих работу пользователя. Не последнюю роль играет и простота интерфейса, наличие дополнительных сервисных функций, как, например, возможность перевода текста документа на иностранный язык, способность выделять все документы с определенного сайта, сужение критериев в ходе поиска, нахождение документов "по образцу" и т.д.

По этим параметрам 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 относится Google (www.google.com). Российские системы этого класса ныне представлены Яндекс (www.yandex.ru), Rambler (www.rambler.ru).

Практически все всемирно известные справочники и поисковые системы в настоящее время превратились во внушительные информационные корпорации с многомиллионными доходами. Заработав авторитет наиболее посещаемых мест в Сети, они предоставляют свои страницы для размещения рекламной информации, доходы от которой и составляют основу их бюджета. Постепенно поисковые сервера превращаются в многофункциональные порталы, в которых поисковый сервис остается главной приманкой для пользователей, но далеко не единственной и даже не основной из предоставляемых услуг. Помимо разыскания информации, такие сервера обычно предоставляют пользователям бесплатную электронную почту, возможность бесплатно размещать собственные страницы, сведения о погоде, текущих новостях, биржевые котировки, карты местности и т.д.

Профессиональная работа с поисковиками требует специальных навыков, поскольку простой ввод искомого термина 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)

Поисковая система, запущенная в 1998 году и являющаяся ныне единоличным лидером среди глобальных поисковых систем по всем значимым параметрам. Главное достоинство Google - объем его индексного файла, который составляет на сегодня более триллиона web-страниц и статей из групп новостей по интересам. В сутки программы-роботы системы индексируют порядка трех миллионов новых и обновленных страниц, при том, что актуализация базы производится каждые 28 дней.

Второе несомненное преимущество Google - его способность индексировать документы не только в виде HTML-файлов, но также документы в форматах PDF, RTF, PS, DOC, XLS, PPT, WP5 и ряде других. При этом Google позволяет моментально конвертировать

страницы в указанных форматах в обычный HTML-файл, что освобождает пользователя от необходимости иметь специальное программное обеспечение для доступа к файлу.

Следующим важнейшим достоинством является специально разработанный модуль ранжирования результатов - PageRank. Он основан на алгоритме, согласно которому вначале устанавливается структура ссылок во всей Сети, а затем каждая отдельная страница ранжируется в соответствии с числом и значимостью ссылок на нее с других страниц. При этом авторитетность внешних ссылок более важна, чем их количество. Подобный алгоритм позволяет существенно повысить релевантность ссылок, вследствие чего Google отличает высокая степень соответствия найденной информации интересам пользователя. Этот результат достигается, в частности, еще и за счет специальной подсистемы защиты пользователя от сайтов, которые продвигаются с помощью различных недобросовестных методов.

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.

Интерфейс первой страницы Google - на сегодня в Сети у него нет достойных конкурентов.

Методика поиска с помощью Google предельно проста. В поисковую строку водится запрос на естественном языке - неважно на русском, английском или любом другом. Язык запросов не допускает усекания терминов знаком "*", поэтому все возможные варианты слов (library, libraries, librarians) пользователю следует вводить самостоятельно. Все термины запроса по умолчанию объединяются условием AND (И) - перед ними нет нужды ставить знак "+". Таким образом, в список результатов попадают лишь страницы, содержащие все введенные ключевые слова. Для поиска по точной фразе традиционно используются кавычки: так, запрос "Кто убил кошку у мадам Поласухер?" прямо приведет к ссылке на полный текст "Собачьего сердца" Михаила Булгакова. 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).

Google имеет в своем арсенале множество опций для максимальной конкретизации запроса. Все они доступны через меню Advanced Search "Расширенный поиск". Помимо уже описанных возможностей, добавляются 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.

Помимо поиска текстовых материалов, Google обладает лучшими на сегодня возможностями поиска иллюстраций с помощью режима "Поиск изображений" (Images, "Картинки").

Кстати, совсем недавно Google сообщил о специальной индексации сайтов крупнейших учебных заведений, что должно несколько улучшить поиск различной информации академического характера. Новый вид поиска под названием Google Scholar (<http://scholar.google.com>) облегчит жизнь учащимся и научным работникам, так как он специально предназначен для поиска специализированной научной литературы, включая рефераты диссертаций, статьи, книги, различные научные публикации и ссылки на них. При этом в базу данных научного поисковика от Google войдут как документы, находящиеся в открытом доступе, так и материалы, доступные исключительно по подписке (статьи, которые только цитируются в проиндексированных материалах, выводятся в результатах с пометкой citation), причем последние включаются в базу данных по согласованию с издателями, полные версии статей будут доступны только

подписчикам, а все остальные пользователи смогут ознакомиться только с рефератами таких материалов.

Академия гугл – позволяет вести поиск научных, надёжных и проверенных источников информации.

Яндекс (<http://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. Актуализация базы осуществляется еженедельно. Кроме того, Яндекс - поисковая система, индексирующая документы в форматах PDF, RTF и DOC.

Первая страница Яндекс. Находится если не все, то многое.

Интерфейс максимально прост - состоит из единственной строки ввода. За счет встроенной системы морфологической обработки терминов Яндекс приспособлен для формирования запросов на естественном русском языке. Мощнейшая лингвистика позволяет 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.

Безусловным достоинством Яндекс является модуль ранжирования результатов поиска. В его основе лежит алгоритм анализа местоположения и повторяемости термина в документе и анализ внешних ссылок на страницу. Система производит группировку найденных страниц по сайтам на этапе формирования списка результатов и выдает по одному документу с сервера, который, по ее мнению, наиболее соответствует запросу. При этом в ссылке указывается общее число документов с данного сервера, также содержащих искомый термин. Яндекс показывает найденные термины в окружающем словарном контексте, что позволяет сразу же установить степень соответствия найденного документа информационной потребности пользователя.

Rambler (<http://www.rambler.ru>)

Запущен в октябре 1996 года и изначально предназначался для выявления материалов на серверах в пределах бывшего СССР. К началу нового тысячелетия Rambler на время утратил лидирующие позиции, устарев практически по всем показателям. Однако проведенная в декабре 2002 коренная модернизация всей программно-аппаратной части позволила поисковой системе вновь обрести былой авторитет. По результатам тестов, Rambler занимает второе место после Яндекса по величине базы данных.

Система обладает обычной и расширенной ("Расширенный поиск") формами ввода запроса. При серьезных разысканиях лучше обратиться к последней, которая предоставляет дополнительные возможности в ограничении запроса различными параметрами, задании критерия сортировки результатов и степени развернутости выводимых результатов. По умолчанию результаты поиска группируются по сайтам, что весьма логично, поскольку на одном сайте термин используется, как правило, в едином контексте. Всегда четко указывается дата создания документа и дата его последнего индексирования поисковым роботом.

Апорт (<http://www.aport.ru>)

Поисковая система, запущенная в феврале 1996 года, ныне прекратила своё существование является каталогом товаров и услуг. Хотя в своё время он был 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. Апорт на достаточно высоком уровне осуществлял поиск MP3-файлов.

Список результатов содержал сведения об адресе, дате опубликования и последней проверке документа. Документы с одного сайта сгруппированы вместе. Помимо поисковой системы, Апорт имел одноименный справочник ресурсов, являвшийся наиболее объемным каталогом в России.

Бинг является русскоязычным вариантом текстовой версии поисковика, запущенного корпорацией Майкрософт в 2009 году. Он в настоящий момент не опережает Гугл. Однако объем индексного файла сопоставим с аналогами в Гугл, но по всем остальным параметрам, по определению релевантности поисковых запросов, Бинг пока ещё уступает своему сильному конкуренту. Система обладает стандартными возможностями поисковика, возможность поиска по картинкам и видеозаписям.

GoGo.ru— является российским поисковиком, разработкой компании Mail.Ru, запущенный в июне 2007 года. Компания приобрела домен gogo.ru ещё в 2000 году. Спустя 6 лет началась разработка, которую возглавил Михаил Костин, создатель поисковика Апорт. Сумма инвестиций составила 700 тысяч долларов.

Результаты выдачи отличаются от встроенной системы на основе поисковика Яндекс. По данным специализированных статистических компаний, качество её поисковой выдачи находится на хорошем уровне.

Отличительными особенностями являются поиск на русском языке по видеозаписям, а также на основе данных проекта ответы. Мэйл.ру. разработчиками предусмотрена возможность фильтрации результатов поиска по теме. Есть возможность ограничить выдачу по коммерческим, информационным, а также пользовательским источникам.

В 1996 году в калифорнийском городе под названием Беркли была разработана новая поисковая система — поисковик Ask.com.

Одна из особенностей поисковика **Ask.com**. состоит в том, что она вопросно-ответная. В ее основе лежит обычный разговорный язык, использующийся каждым из нас в повседневной жизни. Почему разработчики отдали предпочтение такому специфическому подходу. Прежде всего, потому, что количество пользователей глобальной паутины ежедневно возрастает. К тому же, многие новички не обладают нужными навыками использования глобальной паутины, а также временем, необходимым для изучения всех тонкостей составления верных поисковых запросов.

Благодаря, как уже было отмечено выше, «вопросно-ответной» концепции, пользователи этой поисковой системы избавлены от необходимости составлять запросы с нужными ключевыми словами. Кроме того, им не требуется использовать и логические операторы поиска и множество прочих поисковых функций, требующих специальных умений и знаний.

В поисковик Ask.com входят следующие возможности — AskCity (результаты поиска для выбранного города, региона и т.д.), функция перевода страниц на английский язык, автоподбор синонимов ключевых слов, получение информации о времени размещения веб-сайта. Последняя функция позволяет узнать пользователям время появления выбранного портала в глобальной паутине. Благодаря всему этому данная ПС получила своих приверженцев во всемирной паутине.