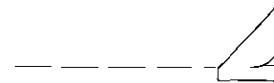


Профессиональный ПОИСК В ИНТЕРНЕТЕ



Поиск информации является одной из наиболее распространенных и наиболее сложных задач, с которыми приходится сталкиваться любому пользователю Сети. Однако если для обычного человека владение методами эффективного информационного поиска является желательным, но далеко не обязательным качеством, то для профессионалов информационной деятельности, и в первую очередь библиотекарей, умение быстро ориентироваться в ресурсах Интернета и находить требуемые источники относится к числу базовых квалификационных навыков. По мере того как все большая часть общего объема документального массива человечества мигрирует в Интернет, умение оперативно и качественно находить там информацию и консультировать в этом вопросе читателей становится закономерным продолжением функции библиотеки в цифровом обществе.

Сложности при проведении поиска определяются двумя главными факторами. Во-первых, число источников в Сети чрезвычайно велико. В настоящее время самые приблизительные подсчеты указывают ориентировочную цифру более чем в триллион документов, расположенных на серверах по всему миру. Во-вторых, этот массив информации не только колоссален по объему, но еще и крайне динамичен. За те полминуты, что вы потратили на чтение первых строк статьи, в виртуальной вселенной появились тысячи новых или измененных документов, сотни были перемещены на новые адреса, а десятки – навсегда прекратили свое существование. Интернет никогда не спит, как никогда не спит наша

планета, по которой непрерывно катится волна деловой активности человечества в точном соответствии со сменой часовых поясов.

В отличие от стабильного и контролируемого фонда документов в библиотеке, в Сети мы имеем дело с гигантским и непрерывно меняющимся информационным массивом, поиск данных в котором является сложным процессом. Ситуация зачастую очень напоминает известную задачу поиска иголки в стоге сена, и порой сведения, представляющие огромную ценность, остаются невостребованными единственно по причине трудности их разыскания.

ИНСТРУМЕНТЫ ИНФОРМАЦИОННОГО ПОИСКА

Не прекращающееся ни на секунду обновление информационного массива Интернета в сочетании с одновременным ростом объема данных крайне усложняет учет имеющихся документов. Никакие списки сайтов, которыми наполнены печатные руководства по работе в Сети, не могут дать точных сведений. В лучшем случае они в состоянии лишь помочь сделать первые шаги: сверхбыстрая смена ситуации в киберпространстве приводит к тому, что подобные перечни безнадежно устаревают уже в момент своего выхода в свет, а поддерживать их в актуальном состоянии становится принципиально невозможным по причине стремительного роста числа новых и новых ресурсов.



Вадим Константинович СТЕПАНОВ,
профессор Московского государственного университета
культуры и искусств

Сверхважность информационного поиска привела к образованию в самом Интернете целой отрасли, задача которой заключается именно в оказании помощи пользователю в его навигации в киберпространстве. Составляют эту отрасль специальные поисковые службы или сервисы. Условно их можно разделить на **справочники (directories)** и **поисковые системы (search engines)**. Внешне эти разновидности зачастую 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 разысканий.

Поисковые средства первой группы (их еще называют **каталогами** или **рубрикаторами**) представляют собой электронные справочники, имеющие четкую иерархическую систематическую или логико-тематическую структуру, во многом напоминающую структуру систематических каталогов универсальных библиотек. Работа с ними позволяет ориентироваться в ресурсах Интернета в пределах отдельных отраслей знания, углубляясь от общего к частному, меняя иерархические ветви, возвращаясь на несколько шагов назад и т.д. К наиболее значительным справочникам всемирного масштаба сегодня относятся **Yahoo! (dlr.yahoo.com)** и **Open**

Directory (dmoz.org). Среди российских разработок в этой сфере сохраняет некоторое значение лишь **Апорт (www.aport.ru)**. Главной особенностью справочников является то, что все они сделаны вручную. Их содержательное наполнение осуществляется редакторами, труд которых более всего напоминает работу сотрудников отделов каталогизации и систематизации крупных библиотек. К редакторам постоянно поступают сведения о вновь возникших ресурсах. На заре интернет-эпохи для этой цели сотрудникам сетевых каталогов приходилось самостоятельно выявлять новые ресурсы. Начиная с конца прошлого века технология была изменена: сведения о новых ресурсах стали поступать главным образом через веб-форму, которая расположена на сайте самого справочника. Эта форма заполняется, как правило, самими создателями ресурса, желающими разместить сведения о нем в каталоге интернет-сайтов. Полученные данные анализируются

Апорт

The screenshot shows the Aport website interface with a search bar and navigation menu. The main content is divided into categories like 'Бизнес и экономика', 'Наука и образование', 'Общество', 'По странам и регионам', and 'Развлечения и отдых'. It also includes a 'Сервисы' section, a 'Почта' section, and a weather forecast for Moscow.

- ❖ Библиотека электронных ресурсов Исторического факультета МГУ **
Электронные тексты. Базы данных. Цифровые изображения рукописей. Коллекция ссылок на исторические источники на русском языке в Интернете.
Изд: М.: ИВ. Регион: Москва
- ❖ Научная библиотека им. Н.И.Лобачевского Казанского государственного университета **
Общие сведения о библиотеке и ее услугах. Преобразоват платных услуг. Информация о новых поступлениях. Перечень издаваемой библиотекой. Список сотрудников.
Изд: М.: ИВ. Регион: Казань
- ❖ Библиотека Санкт-Петербургского государственного университета экономики и финансов **
Об организационной структуре библиотеки. Характеристика фондов. Доступ к электронному каталогу. Об автоматизации и организации обслуживания читателей. Информация о научной, редакционной деятельности и междисциплинарных связях.
Изд: М.: ИВ. Регион: Санкт-Петербург
- ❖ Научная библиотека ГрГУ им. Я. Купалы **
О библиотеке: история, современно, услуги. Автоматизированная система поиска. Каталог без данных на компакт-дисках. Труды Ученых университета. Нормативные документы. Материалы конференций. Список литературы. [рус, бел]
Изд: М.: ИВ. Регион: Гродно
- ❖ Научная библиотека Петрозаводского государственного университета **
Историческая справка. Новости библиотеки. Описание структуры. Перечень фондов и каталогов. Возможность поиска в электронном каталоге. Сведения об издательской деятельности и международном сотрудничестве библиотеки. Доступ к базам данных. Информация для читателей. [рус, англ]
Изд: М.: ИВ. Регион: Петрозаводск
- ❖ СОРАС - доступ к каталогам университетских библиотек Великобритании и Ирландии **
Описание использования операционной программы поиска по библиотекам каталогам. [на англ. яз.]
Изд: М.: ИВ. Регион: Manchester

Описания ресурсов в одном из российских справочников

сайты организаций и учреждений. Отражение в справочниках отдельных web-страниц хоть и встречается на практике, но является скорее показателем низкого уровня работы их редакторов.

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 Интернете 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 исключающий его из перечня при долгом «отсутствии» в Сети. Логичностью (научностью) применяемой схемы классификации определяется степень простоты, с которой можно 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 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 детали-

редакторами и заносятся в разделы справочника в соответствии с принятой в нем классификацией. При этом описание сайта в целом или его раздела, если он представляется вполне самостоятельным блоком, снабжается краткой аннотацией. В некоторых случаях добавляются сведения о языке документов, посещаемости ресурса, его географическом местоположении и т.п. Естественно, что подобным образом можно описать лишь часть ресурсов, да и то лишь весьма крупные объекты, например,

защища запроса, с помощью каталогов не проводятся.

Справочники предназначены для решения трех типов задач: ориентации в незнакомой отрасли знания; разыскания крупных объектов, каковыми являются, к примеру, сайты организаций или значительных проектов; получения 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 весьма общих сведений по тематике и никогда – детальных данных. Понять это помогает простой пример, когда от сервера крупной корпорации, например, «Газпром», содержащего многие тысячи страниц, в справочнике будет представлено лишь название и несколько строк аннотации. Другим наглядным примером является сравнение справочника ресурсов Интернета с систематическим каталогом библиотеки, в котором от книги (в данном случае целого сайта) остается лишь описание и аннотация.

Высокая трудоемкость создания и поддержки в сочетании со сравнительно невысокой эффективностью разысканий привели в начале текущего столетия к повсеместному свертыванию работ по созданию справочников интернет-ресурсов. В настоящее время эта разновидность средств навигации относится к постепенно исчезающим.

В основу работы поисковых систем (поисковых машин) заложены совершенно иные технологические принципы. Задача поисковых машин – обеспечивать детальное разыскание информации в электронной вселенной, что может быть достигнуто только за счет учета (индексирования) всего содержания максимально возможного числа web-страниц. Поисковые системы функционируют в автоматизированном режиме и имеют одинаковый принцип действия. Они состоят из трех базовых программных компонентов. Первый компонент пред-

ставляет собой программу-робот, задача которого заключается в том, чтобы путешествовать с сайта на сайт, находить там новые или изменившиеся документы и скачивать их на главный сервер системы. При этом робот, «просматривая» содержимое документа, находит новые ссылки на другие документы данного сайта или на внешние ресурсы. Программа самостоятельно направляется по указанным ссылкам, находя все новые и новые документы с помещенными в них ссылками, после чего процесс повторяется вновь, напоминая хорошо известный в библиографии метод снежного кома.

Выявленные документы обрабатываются (индексируются) вторым компонентом поисковой системы – программным модулем, отвечающим за формирование базы данных. При этом учитывается все содержание страницы, включая текст, иллюстрации, анимации, аудио- и видеофайлы и пр. Индексации подвергаются все слова в документе, что как раз и дает возможность детальных разысканий 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 при вводе ключевых слов в строку запроса.

Выдача результатов осуществляется с помощью третьего программного модуля, производящего извлечение из базы сведений об искомым источниках и, что еще более важно, интеллектуальное ранжирование результатов. При этом берется в расчет местоположение термина в документе (название, заголовок, основной текст), частота его повторения в документе, процентное соотношение искомого термина к остальному тексту страницы, а также число и авторитетность внешних ссылок на данную страницу с других сайтов.

К основным параметрам поисковых систем относятся: объем индексных файлов (число проиндексированных сайтов и отдельных документов); степень оперативности обновления базы

данных за счет включения сведений о новых материалах и удаления данных об устаревших; возможности для составления запроса; интеллектуальность системы ранжирования результатов поиска; наличие дополнительных сервисных функций, облегчающих работу пользователя.

Первая величина, являющаяся ключевой, устанавливает широту охвата материала и определяется числом проиндексированных документов. Сейчас эта цифра для лидеров мирового сетевого поиска колеблется в пределах триллиона документов. 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.

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.

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. 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 перечисленным 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 относятся **Google** (www.google.com), **Yahoo! Search** (search.yahoo.com) и **Bing** (www.bing.com). Российские системы этого класса ныне представлены **Яндекс** (www.yandex.ru), **Rambler** (www.rambler.ru) и **Gogo.ru** (www.gogo.ru).

YAHOO! DIRECTORY

Search: the Web | the Directory

Search

Arts & Humanities

Photography, History, Literature...

Business & Economy

B2B, Finance, Shopping, Jobs...

Computers & Internet

Hardware, Software, Web, Games...

Education

Colleges, K-12, Distance Learning...

Entertainment

Movies, TV Shows, Music, Humor...

Government

Elections, Military, Law, Taxes...

Health

Diseases, Drugs, Fitness, Nutrition...

News & Media

Newsletters, Radio, Weather, Blogs...

Recreation & Sports

Sports, Travel, Autos, Outdoors...

Reference

Phone Numbers, Dictionaries, Quotes...

Regional

Countries, Regions, U.S. States...

Science

Animals, Astronomy, Earth Science...

Social Science

Languages, Archaeology, Psychology...

Society & Culture

Sexuality, Religion, Food & Drink...

New Additions

613, 612, 611, 610, 59...

By Liz Gill

Fri, May 14, 2010, 12:01 am PDT

May 14 marks the birth of an iconic invention. On this day in 1965, the late [Donald Goerke](#) unveiled SpaghettiOs, the meal in a can that was destined to make children come running like cats at the click of a can opener. It would be rivaled in canned-food fame only by [Spam](#), and would consistently sell more than 100 million count per year for generations to come. I admit I was one of those kids who cheered when the dinner menu consisted of a bowl of C-shaped noodles swimming in mysterious orange-ish sauce. I dare not taste them with an adult palate, however, for fear of spoiling the pleasurable memory.

If you decide to revisit the flavors of youth, you might make a casserole (either [plain](#) or [fancy](#)) using the Os. Rather than use genuine SpaghettiOs in a recipe, though, some have attempted to *copy* the recipe (in [cheesy](#) and even [vegan](#) versions). The Campbell Soup Company recently changed their SpaghettiOs recipe, claiming to have reduced the [sodium content](#) by more than a third, a move that many processed food manufacturers have been making voluntarily as the FDA considers regulating [sodium levels](#).

Along with potential adverse health effects of sodium, another consumer worry is plaguing the canned food industry. Although U.S. manufacturers have mostly abandoned the use of [lead solder](#) in food packaging, recent studies have identified potential risks associated with the chemical compound [Bisphenol-A](#), a component in the lining of cans. While differing experts will tell you that you either [should](#) or [shouldn't](#) worry, a [University of Missouri endocrinologist](#) has at least taken [canned tomatoes](#) out of his own pantry. The good news is that a [quest is on for BPA-free cans](#). They're already in use for some foods, but the FDA has yet to approve them for foods with high acidity (including, yep, those darn tomatoes).

Positive changes in the industry are certainly welcome in the name of health. I just hope they never change Spaghetti-Os too much. That would be a big [uh-oh](#).

Suggested Sites...

[MySlurp.com](#) - as if the food wasn't fun enough - kids can play with their virtual Os in these online games.

[Mealtime.org](#) - recipes, research, and resources from the Canned Food Alliance.

[Can and Aerosol News](#) - latest news from the industry.

[10 Sophisticated Variations On Your Favorite Childhood Foods](#) - from The Learning Channel.

[College Eating and Fitness 101](#) - for students living off canned food, some nutrition tips.

Bottle of chardonnay and a can of SpaghettiOs

What the beautiful people eat on their nights off
(Photo by Inaam)

всего, американские. Практически все справочники декларируют всемирный охват материала, однако практика показывает, что основной упор делается все-таки на североамериканские сайты. Естественно, что обращение к глобальным справочникам ресурсов требует минимального владения английским языком.

Yahoo! (dir.yahoo.com). Является одним из самых первых и авторитетных справочников Всемирной паутины. У каталога два основных достоинства – внушительный объем (на сегодня отражено порядка 2 млн сайтов в более чем 25 тыс. категорий) и научность, логичность используемой схемы классификации. За счет этого отражаются все отрасли знания без каких либо видимых приоритетов. Эти достоинства на протяжении всей второй половины 1990 гг. позволяли **Yahoo!** оставаться незаменимым инструментом для первоначального ознакомления с информационным наполнением в Интернете по любой области деятельности.

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 этого справочника наиболее эффективно для ознакомления с наполнением Интернета 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 пойти до web-сайта конкретной библиотеки, миновав по пути разделы «Библиотеки», «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 пользовате-

Интерфейс справочника Yahoo!. Ныне это всего лишь одна из многочисленных служб гигантского портала

Практические все известные справочники и поисковые системы в настоящее время превратились в крупные информационные корпорации с многомиллионными (а некоторые и многомиллиардными) доходами. На основе поисковых сервисов созданы гигантские высокотехнологичные системы интернет-рекламы, построенные в значительной мере на основе учета и анализа пользовательских поисковых запросов. Поисковые серверы постепенно превращаются в многофункциональные порталы, в которых поисковый сервис остается главной приманкой, но далеко не единственной и даже не основной из предоставляемых услуг. Помимо разыскания информации, такие серверы обычно предоставляют пользователям бесплатную электронную почту, сведения о погоде, текущих новостях, биржевые котировки, карты местности, возможность бесплатно размещать собственные страницы и тому подобные сервисы.

ГЛОБАЛЬНЫЕ СПРАВОЧНИКИ РЕСУРСОВ И ПОИСКОВЫЕ СИСТЕМЫ

Электронные справочники глобального масштаба. Созданием и актуализацией справочников глобального масштаба в большинстве случаев занимаются информационные компании, прежде

лям предлагается выбрать первую букву названия сайта (например, названия университета или фамилии популярного исполнителя).

Справочник имеет перекрестную структуру, позволяющую находить данные, используя различную логику поиска. Например, Московский государственный университет может быть найден как по цепочке «образование/высшее образование/университеты/университеты России», так и через «регионы/страны/Россия/образование/университеты».

При необходимости выявить данные о конкретных объектах целесообразно прибегнуть к непосредственному поиску. В поисковую строку должны быть введены ключевые слова, максимально полно обозначающие искомый объект. Следует учитывать большой объем Yahoo! и, следовательно, опасность возникновения большого информационного шума при выдаче результатов. Поэтому, прежде чем прибегать к поиску по ключевым словам, рекомендуется предварительно пройти несколько соответствующих иерархических ступеней. Например, для поиска сайтов, посвященных хоккеисту Павлу Буре (Bure), рекомендуется предварительно войти в раздел «Спорт»/«Хоккей» и ограничить условия поиска данным разделом, передвинув соответствующую фишку на пункт «Category». В этом случае автоматически отсеиваются ссылки на многочисленные сайты, в которых встречается термин Bure, но многие не имеют никакого отношения к российскому хоккеисту. В случае успешности поиска в перечне результатов вначале будут указаны категории (рубрики классификации), в которых встречаются искомые слова, а затем названия отдельных сайтов.

Российский вариант Yahoo!, располагающийся по адресу www.yahoo.ru (ru.yahoo.com) не содержит справочника российских ресурсов, функционируя лишь в качестве поисковой системы с интерфейсом на русском языке.

Учитывая большую популярность и, следовательно, посещаемость Yahoo!, в начале 2000-х гг. владельцы портала все большее внимание уделяли побочным сервисам, среди которых элект-

ронные магазины, аукционы, службы новостей, агентства путешествий и т.п. Собственно справочник занимал все более и более скромные места в интерфейсе титульной страницы портала, пока в середине 2004 г. не исчез с нее почти полностью, выделившись в отдельную службу.

С марта 2004 г. за включение сайта в справочник взимается не только разовая, но и ежегодная абонентская плата в 300 долларов. Это не только затрудняет включение новых сайтов мелкого бизнеса, но и приводит к исключению из каталога ресурсов, владельцы которых не уплатили ежегодный взнос. Естественным следствием такого поведения стало постепенное обеднение содержания справочника, который все более расходится с реальным наполнением Интернета. 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, от его былой популярности мало что останется.

Open Directory Project (dmoz.org). Справочник является на сегодня одним из наиболее полных в мире – порядка 4,5 млн сайтов в более чем 590 тыс. категорий. Его главной особенностью является некоммерческая направленность: он практически полностью формируется силами добровольцев из числа пользователей Сети, считающих себя экспертами в различных областях знания. В этом одновременно и сила, и слабость, поскольку степень и качество наполнения разных категорий

Основной интерфейс Open Directory. Маленький зеленый крокодилик составляет серьезную конкуренцию Yahoo!

dmoz.org open directory project In partnership with AOL Search

[about dmoz](#) | [dmoz blog](#) | [suggest URL](#) | [help](#) | [link](#) | [editor login](#)

Search [advanced](#)

Arts Movies, Television, Music...	Business Jobs, Real Estate, Investing...	Computers Internet, Software, Hardware...
Games Video Games, RPGs, Gambling...	Health Fitness, Medicine, Alternative...	Home Family, Consumers, Cooking...
Kids and Teens Arts, School Time, Teen Life...	News Media, Newspapers, Weather...	Recreation Travel, Food, Outdoors, Humor...
Reference Maps, Education, Libraries...	Regional US, Canada, UK, Europe...	Science Biology, Psychology, Physics...
Shopping Clothing, Food, Gifts...	Society People, Religion, Issues...	Sports Baseball, Soccer, Basketball...
World Català, Dansk, Deutsch, Español, Français, Italiano, 日本語, Nederlands, Polski, Русский, Svenska...		

Become an Editor Help build the largest human-edited directory of the web

Copyright © 1998-2010 Netscape

4,528,597 sites - 85,383 editors - over 590,000 categories

справочника полностью определяются наличием, квалификацией и уровнем ответственности редактора конкретной рубрики.

Иерархическая структура Open Directory достаточно научна и логична. В каждой рубрике, помимо англоязычной части, как правило, представлены варианты на нескольких десятках национальных языков, в числе которых – и русский. Все они подготовлены региональными редакторами, являющимися жителями соответствующих стран. При желании можно перейти на русскоязычный вариант с первой страницы, выбрав из основного деления «World» подрубрику «Русский». В ней, однако, на сегодня зарегистрировано немногим более 74 тыс. сайтов, что ни в коей мере не отражает реального наполнения российского сегмента Сети. Подобным же образом обстоит дело и с другими иноязычными вариантами.

В настоящее время Open Directory является динамично развивающимся проектом, который перехватывает значительную долю популярности у коммерциализировавшегося Yahoo!. Увеличение среднего объема справочника более чем на 1000 сайтов в день сулит ему неплохие перспективы, подтверждая мысль о том, что совокупный интеллектуальный потенциал сетевого сообщества может превосходить интеллект одной, пусть даже очень мощной компании. Усовершенствованный вариант Open Directory с успехом используется в качестве справочника в поисковой системе Google.

Глобальные поисковые системы. Поисковые системы (search engines) распространены в большем количестве, нежели справочники, и их число, составляющее сегодня несколько десятков, продолжает неуклонно увеличиваться.

Интерфейс первой страницы Google. Сегодня в Сети у него нет достойных конкурентов.

Профессиональная работа с поисковыми системами требует некоторых специальных навыков, поскольку простой ввод искомого термина в поисковую строку, скорее всего, приведет к получению списка из сотен тысяч или даже миллионов документов, содержащих данное понятие, что практически равносильно нулевому результату.

Google (www.google.com). Поисковая система, запущенная 28 октября 1998 г. 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.* В сутки программы-роботы системы индексируют около 3 млн новых и обновленных страниц, а актуализация базы производится каждые 28 дней.

Второе несомненное преимущество Google – его способность индексировать источники не только в виде HTML-файлов, но также документы в форматах PDF, RTF, PS, DOC, XLS, PPT, SWF 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.

Третьим 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 существенно повышает 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, продвигающихся 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.ru. Длительность процесса поиска в большинстве случаев не превышает одной секунды, несмотря на огромный объем индексного файла системы.

Методика поиска в Google предельно проста. В поисковую строку вводится запрос на естественном языке: на русском, английском или любом другом. Система обладает прекрасным морфологическим аппаратом, поэтому при вводе запроса в именительном падеже единственного числа можно быть уверенным в нахождении всех возможных склонений и чисел употребленных в запросе слов. 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 соответствующий фрагмент текста «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 +И).

Google имеет в своем арсенале множество опций для максимальной конкретизации запроса. Все они доступны через меню «Расширенный поиск». Помимо уже описанных возможностей, добавляются фильтры, ограничивающие язык документа, его формат (например, «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.

Выдаваемые в результате поиска ссылки на документы, помимо традиционных сведений о заголовке документа, контексте искомого слов и данных о размере, содержат функцию «Сохранено в кэше», позволяющую полностью восстановить весь документ, если по каким-либо причинам сайт, на котором он расположен, недоступен или вообще прекратил существование. Еще одна функция «Похожие страницы» позволяет получить перечень страниц, содержание которых схоже с указанным источником. Эта функция, впрочем, пока реализована без особого успеха.

В 2006 г. в Google добавлена возможность перевода найденных web-страниц на наиболее распространенные языки, включая русский, с основных европейских языков. Щелчок мышью по пункту «Перевести эту страницу» в перечне результатов приводит к запуску процесса перевода и через несколько секунд страница на иностранном языке, предстает в переводе на русский. При этом изначальное оформление страницы сохраняется прежним. Качество перевода пока оставляет желать лучшего, однако развитие технологий машинного перевода заставляет смотреть на будущее этой функции с оптимизмом.

Помимо поиска текстовых материалов, Google обладает лучшими на сегодня возможностями поиска иллюстраций с помощью пункта меню «Картинки». В его базе данных учтено более миллиона иллюстраций, разыскание которых ведется аналогично поиску текстовых документов с возможностью ограничения определенным размером, форматом или цветностью графических файлов – все через опцию «Расширенный поиск картинок». Также в арсенале Google находится и модуль

Интерфейс формирования расширенного запроса Google. Максимальные возможности в сочетании с максимальной заботой о пользователе

поиска видеофайлов, размещенных на различных видеосерверах по всему миру.

В качестве собственного справочника ресурсов Интернета Google использует усовершенствованный массив Open Directory Project, что порой позволяет сочетать достоинства обоих поисковых инструментов. В дополнение к стандартному поисковому сервису Google располагает самым внушительным среди глобальных поисковых систем перечнем факультативных услуг. Переход к ним осуществляется кликом по ссылке «Еще» в главном меню над строкой ввода запроса. Среди наиболее популярных дополнительных сервисов значатся поиск по картам местности, поиск полнотекстовых книг, поиск научных публикаций, поиск на постоянно обновляемых новостных сайтах и многое-многое другое.

Yahoo! Search (search.yahoo.com). С момента своего создания в первой половине 1990-х и до начала 2000-х гг. поисковый модуль Yahoo! был абсолютно не развит. Авторитет зиждился прежде всего на мощнейшем справочнике ресурсов. Поисковый модуль выполнял вспомогательную функцию, обеспечивая разыскания внутри гигантского каталога. Финансовая успешность проекта позволила владельцам Yahoo! приобрести ряд серьезных программных разработок в области сетевого поиска, включая Inktomi. При реорганизации портала в 2004 г. этот механизм в доработанном виде лег в основу собственной поисковой системы, получившей название Yahoo! Search.

С весны 2004 г. поисковая система Yahoo! Search функционирует в структуре портала в качестве самостоятельного сервиса и в настоящее время относится к числу лидеров всемирного сетевого поиска. По данным, полученным в результате тестирования, объем ее индексного файла ныне колеблется в пределах триллиона web-страниц. Система обладает возможностью поиска документов в форматах HTML, TXT, PDF, DOC, XLS, PPT и XML. Индексированию подвергаются не только тексты, но и иллюстрации.

Одним из наиболее ценных качеств является способность вести поиск видеофрагментов, в том числе и в русскоязычной части Интернета. Реализованный по классической технологии, этот модуль обеспечивает высокое качество разысканий и выявление большого числа проиндексированных фрагментов.

Как и все профессиональные поисковые средства, Yahoo! Search обладает формой расширенного поиска «Advanced Search», позволяющей формулировать поисковое предписание с помощью многоступенчатых меню.

Также представлена функция перевода найденных страниц на основные европейские языки, включая русский.

Русскоязычная версия Yahoo! Search расположена по адресу <http://ru.yahoo.com>. С ее помощью пользователь имеет возможность произвести поиск по всему Интернету или ограничить разыскания только российским сегментом. Также представлен поиск по российским новостным сайтам и поиск иллюстраций.

Bing (www.bing.com). Эта бета-версия поисковой системы, запущенной компанией Microsoft в середине 2009 г., пока не обладает преимуществами, позволяющими ей на равных конкурировать с Google или Yahoo! Search. Тесты показывают, что объем индексного файла Bing в настоящее время более чем на половину уступает аналогичным показателям обоих конкурентов. Еще в большей степени идет отставание по уровню релевантности выдачи результатов поиска.

Система обладает стандартными возможностями составления запроса, включая поиск иллюстраций и видеофайлов, ограничением поиска на определенном языке или в пределах серверов одной страны. Есть основания надеяться, что Microsoft, проведя тестирования системы и исправив выявленные просчеты, создаст новую вариацию Bing, которая будет в состоянии встать в один ряд с лидерами сетевого поиска.

Помимо перечисленных глобальных поисковых систем, в некоторых случаях по инерции продолжают использоваться устаревшие поисковые сервисы, сохранившие остатки былой популярности со второй половины 90-х гг. прошлого века. К их числу относятся **Alta Vista (www.altavista.com)** и **AlltheWeb (www.alltheweb.com)**. В настоящее время малый объем их индексных файлов и технологические недостатки не позволяют полагаться на сведения, которые они предоставляют.

В то же время высокая доходность поисковых проектов обуславливает перманентное появление новых поисковых машин, разработчики которых непременно заявляют о колос-

сальном потенциале своих детищ. По прошествии времени чаще всего выясняется, что декларированные возможности сильно завышены, и широко разрекламированные на начальном этапе сервисы благополучно отходят в небытие, в лучшем случае сохраняя работу в фоновом режиме. Такая участь постигла запущенные в период 2000–2005 гг. поисковые системы **Teoma** (www.teoma.com), **Wotbox** (www.wotbox.com), **Gigablast** (www.gigablast.com), **Entireweb** (www.entireweb.com), которые так и не смогли завоевать весомый авторитет у широкой аудитории интернет-пользователей.

В 2006–2009 гг. возник ряд новых поисковых проектов, многие из которых до настоящего времени пребывают на этапе бета-тестирования. Идет активный поиск технологий для достижения более высоких результатов. Ирландская поисковая система **Cull** (www.cull.com) декларирует максимальный объем индексного файла, а индийская **Yauba** (www.yauba.com) ставит во главу угла конфиденциальность поисковых запросов.

Создатели целой группы поисковых систем – **Mooter** (www.mooter.com), **Quintura** (www.quintura.com), **Ask.com** (www.ask.com) – активно экспериментируют с кластеризацией получаемых по запросу ссылок. Разработчики **Collarity** (www.collarity.com) главное внимание уделяют способам ранжирования результатов поиска, в том числе принимая во внимание мнения пользователей о сайте, содержащем искомые документы. Сторонники семантического анализа трудятся над созданием систем, способных анализировать смысловое содержание запроса уже на этапе ввода ключевых слов – **Hakia** (www.hakia.com).

Наконец некоторые компании-разработчики идут по пути повышения внешней привлекательности поисковых систем путем визуального представления найденных страниц – **Snap** (www.snap.com). Пока, однако, ни один из недавно запущенных поисковых проектов сколько-нибудь существенного влияния на ситуацию в мировом поисковом сервисе не оказал.

Метапоисковые системы. 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 «метапоисковые машины». К наиболее известным их представителям относятся, в частности, **MetaCrawler** (www.metacrawler.com), **Ixquick** (www.ixquick.com) и **ezZfind** (ezZfind.com).

Метапоисковые 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 **MetaCrawler**, фактически одновременно обращается к пяти поисковым системам. Этим гарантируется «объективность» полученных результатов. Однако, учитывая различия в подходах к обработке терминов разными системами, результат может оказаться не всегда релевантным запросу.

Пользоваться метапоисковыми системами рекомендуется для выполнения запросов, связанных с поиском очень простых и, предположительно, крайне редко встречающихся в Сети предметов (лекарственных препаратов, химических элементов, редких сочетаний имен собственных и т.п.). В этом случае результаты поиска сразу же дают представление о том, насколько широко данные ключевые слова (объекты) встречаются в Сети. Другой причиной обращения к метапоиску является крайний недостаток времени. Однако, поскольку метапоисковые средства часто не в состоянии корректно обработать запрос для различных поисковых систем и правильно совместить полученные из разных источников результаты, их использование не рекомендуется при выполнении сложных запросов. Поисковые предписания, включающие несколько словосочетаний с указанием расстояния между терминами и тому подобными условиями, заведомо не могут принести правильный ответ. Развернутый список метапоисковых систем можно получить в **Русской справочной библиотеке** (library.vadimstepanov.ru).