



Дорожная карта сетевого поискового бизнеса

...Я просто ставил опыты о том, Какая рыба быстрее всех.

Борис ГРЕБЕНЩИКОВ

Дмитрий ЛАНДЭ

Для современного человека именно Интернет становится основным источником информации. Объемы информации, размещаемой в Интернете, поражают воображение. Ориентироваться в этом информационном потоке можно по-разному. Многие читают любимое

интернет-издание, но уже сегодня большинство пользователей обращаются к сетевым информационно-поисковым системам (ИПС), о которых пойдет речь дальше. Казалось бы, ИПС тем лучше, чем больше ресурсов она охватывает. Ведь важно ничего не упустить при поиске.

Однако, как мы увидим ниже, объем индексных баз данных играет пусть важную, но все же не решающую роль. Для пользователя принципиально важны результаты поиска, качество их ранжирования, удобство работы, наличие интуитивного информационно-поискового языка, развитого расширенного языка запросов, удобной навигации — всего того, что называют интерфейсом пользователя и неологизмом «юзабилити».

Определенному типу пользователей (вообще говоря, не составляющему большинства) нужны и так называемые специальные возможности. К ним относятся, во-первых, возможности «расширенного поиска», где обычно задается поиск по определенным доменам, в документах определенных форматов, в определенных полях и т.д. К специальным возможностям относятся также поиск в таких типах документов, как новости, блоги, научные статьи или книги (рис. 1).

Как это ни сенсационно звучит, особенно в контексте данной статьи, но веб-среда сегодня не является крупнейшим информационным ресурсом в Интернете. Основной объем ресурсов

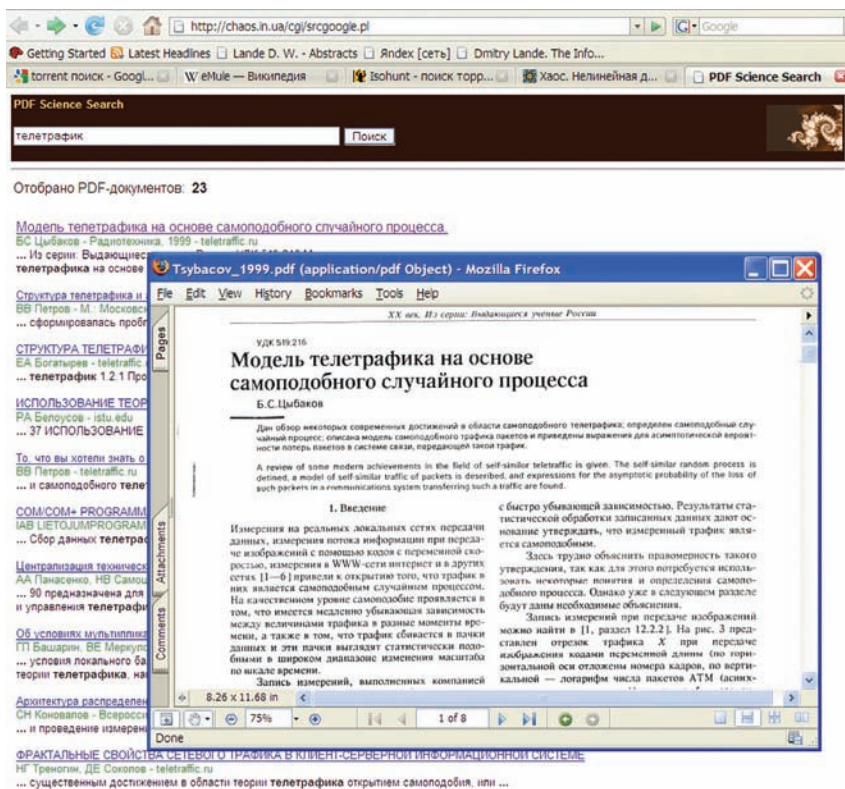


Рис. 1. Сайт chaos.in.ua — метапоисковая система, с которой обеспечивается «научный поиск» — поиск pdf-документов (а не страниц, ссылающихся на них)

сосредоточен в так называемых «пиринговых», файлообменных сетях. Вопрос эффективного поиска в таких сетях пока остается открытым, однако существуют специальные поисковые сайты в веб-пространстве, помогающие решить эту проблему.

По оценкам экспертов, с помощью даже самых крупных глобальных поисковых систем в Интернете сегодня доступно не более 30% открытой информации, присутствующей в веб-среде. Веб-ресурсы, находящиеся в свободном доступе, но не доступные с помощью обычных поисковых систем, образуют так называемый *глубинный веб*. Решение проблемы «глубинного веба», обеспечения доступа к «скрытым» ресурсам, достигается с помощью специальных каталогов и систем, зачастую доступных обычным пользователям Интернета.

В Интернете подавляющая часть информационных ресурсов сосредоточена в так называемых пиринговых, файлообменных сетях.

В последнее время многие из таких технологических решений базируются на идеологии *семантического веба*, основные задачи которого в настоящее время выглядят достаточно химерными. Вместе с тем частные решения, полученные при попытках реализации семантического веба, сегодня широко применяются в информационных технологиях. К таким решениям относятся, например, агрегация новостей или ведение блогов на основе XML.

Несмотря на развитие сопутствующих информационному поиску сервисов, все же основным фактором, определяющим выбор той или иной системы большинством пользователей, является качество информационного поиска. Постоянно проходящие семинары по оценке качества информационного поиска напоминают водопой в период засухи: непримиримые рыночные конкуренты собираются вместе и обсуждают общие на-

О контекстной рекламе

Контекстная реклама – это показ текстовых рекламных блоков и баннеров, которые привязаны определенными ключевыми словами и словосочетаниями (сегодня их принято называть «тегами») на различных сайтах, в частности, на сайтах поисковых систем при вводе результатов поиска. Места для показа контекстной рекламы могут быть различными – от первых результатов поиска (сегодня считается неприличным перемешивать проплаченные и естественные результаты поиска), правых или левых колонок, верхних или боковых баннеров и т.п.

Контекстная реклама чаще всего оплачивается по кликам (переходам на целевые сайты), рекламодатель платит только

за целевых посетителей. Если количество рекламных мест на статических рекламных площадках ограничено, то на динамических страницах, формируемых поисковыми системами, оно не имеет таких ограничений. Контекстная реклама, в отличие от традиционных видов рекламы в Интернете, обеспечивает целенаправленную рекламу (в интернет-сообществе этот эффект принято называть калькой с английского языка – «таргетинг»). Благодаря этому эффекту она не вызывает психологического отторжения у пользователей поисковых систем, а воспринимается как дополнительная полезная информация, повышает лояльность пользователей по отношению к объекту рекламы.

сушные проблемы, находят новые технологические решения.

При этом бизнес на информационно-поисковых системах развивается именно как бизнес, причем

Охват – «объем индекса»

Сегодня информации в Сети появляется больше, чем ее успевают проиндексировать поисковые системы. Это означает, что информационный хаос увеличивается, и существующие подходы не соответствуют требованиям растущего информационного пространства. Вместе с тем, чем больше ресурсов соответствующего профиля включает база данных системы, тем выше должна быть полнота. Именно это соображение объясняет жесткую конкурентную борьбу за объемы баз данных индексов веб-документов, ведущуюся с самого начала возникновения поисковых систем. Такие базы данных в технологиях ИПС принято называть индексами поисковых систем.

Финансовая основа бизнеса на поисковых интернет-системах базируется на рекламе, а точнее, на контекстной рекламе.

свой бизнес? Вариантов правильных ответов много, но, по-видимому, финансовая основа этого бизнеса базируется на рекламе, а точнее, на *контекстной рекламе*. То есть, чем популярнее поисковый сайт, чем больше на него ведет ссылок, чем более он посещаем, тем он прибыльнее. Итак, рассмотрим некоторые свойства поискового сервиса, от которых напрямую зависит поисковый бизнес.

Еще 5 лет назад крупнейшие поисковые системы мира вели ожесточенную борьбу именно за этот показатель. На первых страницах таких поисковых сайтов, как Altavista, Google, Alltheweb, Yahoo! публиковались соответствующие цифры – количество проиндексированных документов или объем индекса. В начале XXI века в лидеры по охвату ресурсов выбилась служба Google. Однако



Рис. 2. Интерфейс системы Cuil

в 2002 году находящаяся сегодня в тени система Alltheweb неожиданно вышла на первую позицию по охвату сетевых ресурсов и, соответственно, была признана лучшей сетевой ИПС в мире по объему индекса, проиндексировав 2,1 млрд. веб-страниц. Затем лидерство вновь вернулось Google — свыше 3,3 млрд. веб-страниц в 2003 году.

Последняя цифра, которая была размещена на титульной странице Google, составляла чуть более 8 млрд. страниц (цифра была приведена в 2005 г.) После этого цифры перестали публиковаться, надо полагать, не по техническим причинам, ведь наивно считать, что владельцы баз данных не знали их объемов. Из официальных пресс-релизов того же 2005 года известно, что объем индекса Google составлял 13 млрд. документов, объем индекса Yahoo! превысил это значение и достиг на то время 20 млрд. документов. Администрация Google была не согласна с этой цифрой, выступая с опровержением. Вместе с тем данные с главной страницы Google были уже сняты, хотя генеральный директор компании Эрик Шмидт одновременно заявил: «Чем больше индекс, тем лучше релевантность и тем полнее обзор». Вместе с тем в заявлении Yahoo! было сказано: «Мы поздравляем Google с изъятием с их главной страницы числа, показывающего размер индекса, и с признанием того, что оно ничего не значит. Как мы уже говорили, важно лишь, чтобы потребители находили то, что они ищут, и мы предлагаем пользователям сравнить результаты поиска наших систем».

Казалось бы, конфликт был исчерпан, и возвращаться к оценке объема индекса никто не будет.

Однако прошло время, и мир поисковых систем облетела очередная сенсационная новость. В конце июля 2008 года появилась новая глобальная поисковая система Cuil (рис. 2) с относительно небольшим бюджетом (33 млн. долларов), содержащая в индексе 121 млрд. веб-страниц, что, по мнению экспертов, в несколько раз превышало индекс Google.

Системы новой поисковой системе ведут к той же Google. Создатели Cuil — Анна Паттерсон, ее муж Том Костелло и еще несколько бывших сотрудников Google (среди которых Луис Моне, один из создателей AltaVista) — специализировались на поиске в сверхбольших базах данных. В частности, Паттерсон, работая в Google, зарегистрировала соответствующий патент (Multiple Based Index Information Retrieval System).

Google сразу же отреагировала на сенсационное заявление Cuil, тут же заявив о том, что успешно проиндексировала триллионную по счету веб-страницу. Понятное дело, кто это может проверить? В общем-то, данное заявление очень расплывчато и означает лишь то, что с момента возникновения системы ею обработан триллион веб-страниц.

В компании говорят, что поисковик научился отыскивать и удалять из индекса дубликаты страниц и страницы с разными адресами. «Старт работы по индексированию начался с того, что поисковый робот начал запоминать содержимое страниц и следовать по гиперссылкам, присутствующим на этих страницах. Система постоянно следует по ссылкам, переходя с сайта на сайт и запоминая содержимое изученных страниц. В реальности Google проиндексировал уже более триллиона страниц, однако далеко не все из них являются уникальными автономными страницами. «Многие из них имеют по несколько адресов, другие являются копиями друг друга», — пишет в официальном блоге ком-

пании Ниссан Хаджай, один из разработчиков поисковой системы. Сегодня, как рассказывают в компании, пополнение индекса не останавливается ни на секунду, а благодаря распределенной системе обчета и оперативному обновлению информации весь поисковый индекс ранжируется заново по несколько раз в сутки.

Несмотря на гигантский размер самой мощной поисковой системы современности, Google, объем ее актуального поискового индекса по каким-то причинам остается тайной за семью печатями. Можно лишь косвенно сравнить показатели Google и Cuil, задавая им простейшие запросы (информации Cuil можно доверять — ее создатели предъявили поисковый индекс внешним экспертам). Как явствует из материалов компаний, обе поисковые системы не используют так называемого стоп-словаря, т.е. запросы по простым, часто употребляемым словам позволяют оценить соотношение объемов индексов. И такую оценку может сделать каждый! Например, введем поисковое слово «the» одновременно двум системам. Получаем:

Google: about 22,550,000,000 for the;

Cuil: 22,883,636,124 results for the.

Результаты вполне сопоставимы — можно сделать вывод о примерно одинаковом объеме поисковых индексов. Введем слово «для» (для проверки русскоязычной части), получаем:

Google: about 546,000,000 for для;

Cuil: 368,508,113 results for для.

Русскоязычная часть индекса Google оказалась несколько большей. О низком качестве (объеме) русскоязычного индекса Cuil свидетельствуют и запросы по другим словам.

Вроде можно было бы остановиться, результаты получены, однако введем еще одно слово «of» для проверки. Получаем неожиданный ответ:

Google: about 22,760,000,000 for of;

Cuil: 121,000,000,000 results for of.

Итак, у Cuil результат более чем в 5 раз весомее. Но, учитывая итоги поиска по слову «the» (и по другим словам, в частности, не только на английском языке), можно сделать иной вывод. Каковы бы ни были результаты подобных сравнений, факт остается фактом: Google — самая популярная поисковая система, самый дорогой бренд в мире, а Cuil — мало кому известный проект с бюджетом региональной поисковой системы. Действительно, можно согласиться, что объем поискового индекса решает далеко не все.

Интерфейс пользователя

Пользователь будет считать систему «своей», если он в значительной степени может управлять результатами ее работы, вводя управляющие воздействия и получая соответствующие отклики. В случае поисковых систем воздействие пользователя — это запрос и набор дополнительных параметров. Можно констатировать, что возможности эффективного управления результатами поиска в общем случае находятся в обратной зависимости от объема индекса системы.

Например, ни одна из известных автору систем не в состоянии сегодня ранжировать найденные документы по длине, как это делала некогда система Alltheweb, в то время, когда ее индекс не превышал миллиарда документов. Система «Яндекс», объем индекса которой в настоящее время — 4,55 млрд. документов, позволяет ранжировать релевантные документы по датам, чего не способна делать Google, индекс которой на порядок больший. А именно последняя возможность кажется автору принципиально важной для аналитической работы — при постоянном отслеживании публикаций по одному и тому же запросу. Для пользователя существенно наличие интуитивного информационно-поискового языка, удобная навигация в результатах поиска, ранжирование результатов, получение

автоматически формируемых понятных кратких описаний документов (сниппетов).

Интерфейсы режимов «расширенного поиска» большинства поисковых систем сегодня, с одной стороны, перегружены, а с другой — не позволяют в полной мере выразить все информационные потребности. Очень часто пользователь буквально тонет в результатах первичного поиска, но, переходя в режим «расширенного», заполняя все прилагаемые поля, приходит к отсутствию результатов (при том что основная проблема веба, как информационной среды, — это избыточность информации). В связи с этим в последнее время получили распространение более гибкие визуальные интерфейсы уточнения запросов, чаще всего реализуемые путем «квазиинтеллектуальной» группировки (кластеризации) результатов первичного поиска. Появилось множество подходов, общее у которых — попыт-

ка представить результаты поиска и соответствующие им кластеры в удобном для пользователей виде.

Например, российская метапоисковая система Nigma.ru по слову «Интернет» представляет кластеры (сгруппированные документы), соответствующие автоматически определенным словам (тэгам) «интернет-магазин», «услуги», «каталог товаров», «доставка», «ссылка» и др., что свидетельствует о коммерциализации российского сегмента Интернета. Nigma.ru также использует технологию отображения результатов кластеризации в виде «облака» тэгов. На каждый тэг можно «кликнуть», в результате чего получить подчиненные ему тэги. Результаты поиска, соответствующие выбранному тэгу, отображаются в нижней части страницы в традиционном списковом виде.

Другая российская поисковая система Qintura (<http://search.qintura.ru/>) обладает «интеллектуальным» интерфейсом, который обеспечи-

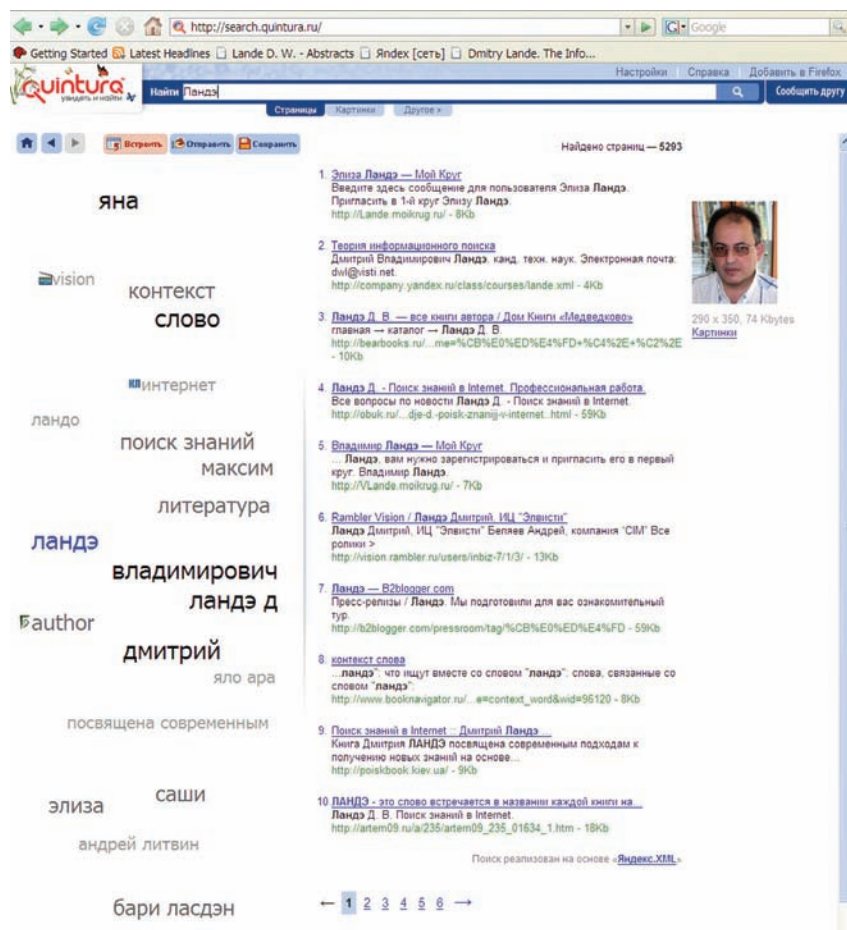


Рис. 3. Результат поиска по запросу фамилии автора в системе Qintura

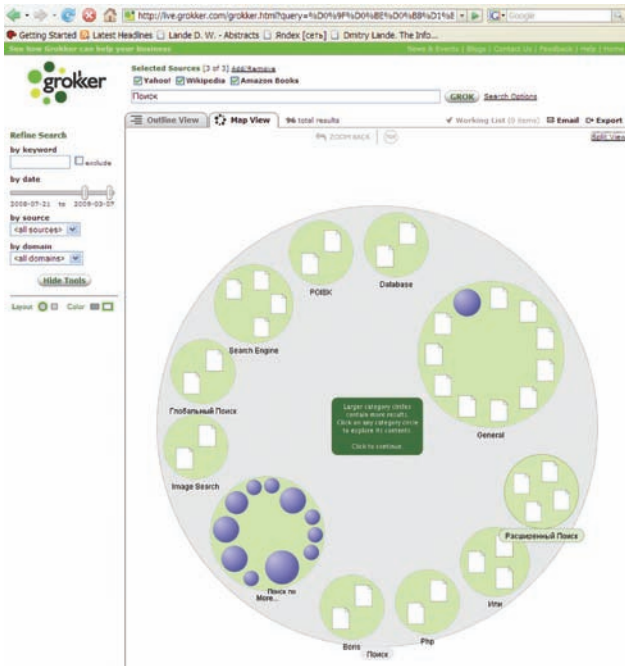


Рис. 4. Визуальный интерфейс системы Гроккер в виде круговых диаграмм

дает получение автоматических подсказок по введенному запросу, помогает динамично управлять процессом поиска (рис. 3). По запросу из одного слова Quintura предьявляет возможные фразы и словосочетания, которыми при необходимости расширяется первичный поисковый запрос.

Развитым графическим интерфейсом обладает система Гроккер (grokker.com), результаты поиска в которой группируются в круговые диаграммы (рис. 4). Система позволяет производить поиск в базах данных Yahoo!, Википедии, Amazon Books. Интересной особенностью

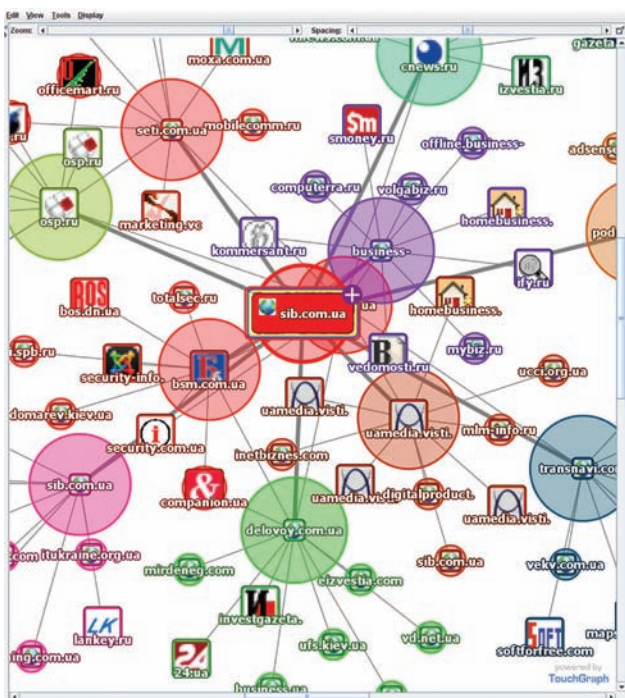


Рис. 5. Отображение сети связей по подобию в системе TouchGraph Google Browser

системы являются такие средства уточнения запроса, как управления временем публикаций (с помощью «ползунка», как и у известного агрегатора новостей NewsIsFree), а также выбор доменов или источников.

Одним из самых динамичных новостных ресурсов Интернета сегодня можно считать и живые журналы (блоги). Компания TouchGraph, в частности, реализовала интерфейс для визуализации социосетей на основе LiveJournal — TouchGraph LiveJournal Browser. Однако самый популярный инструмент компании TouchGraph Google Browser (<http://www.touchgraph.com/TGGoogleBrowser.html>), представляющий собой Java-апплет для визуализации тематического подобию веб-сайтов. Это весьма полезный инструмент при поиске сайтов, связанных с исходным общей тематикой. В интерфейсе TouchGraph Google Browser можно увидеть все сайты, связанные ссылками с исходным заданным сайтом (рис. 5), при этом пользователь может задавать глубину связей и отображать взаимосвязи различных сайтов.

Системы визуального поиска фокусируются на психологических аспектах человеческого восприятия, ориентируясь на методики, которые используют люди в процессе поиска. Именно поэтому визуальные поисковые системы имеют все шансы потеснить на информационном рынке таких гигантов веб-поиска, как Google и Yahoo!, используя базы данных последних. С другой стороны, может быть, более очевидный ход событий заключается в переходе ведущих сетевых ИПС в область визуального поиска? Ведь для пользователей визуально организованные результаты поиска выглядят гораздо привлекательнее и понятнее, чем списки гиперссылок и сниппетов, формируемых традиционными поисковыми системами.

Сегодня основной объем информационных ресурсов сети Интернет сосредоточен не на веб-сайтах, а в так называемых пиринговых, файлообменных сетях. На практике пиринговые сети состоят из рабочих станций, каждая из которых взаимодействует лишь с некоторым подмножеством узлов сети (из-за ограниченности ресурсов). Для реализации протокола P2P используются клиентские программы, обеспечивающие функциональность как отдельных рабочих станций, так и всей пиринговой сети в целом.

Достаточно часто пиринговые сети дополняются выделенными серверами. Чаще всего именно такие серверы позволяют решать вопросы поиска по запросам, так как именно эта проблема для пиринговых сетей не может считаться решенной. Вопрос эффективного поиска в таких сетях средствами самих сетей остается пока открытым (поскольку большинство из них не предполагает жесткой централизации, а напротив, они по определению являются децентрализованными), однако существуют специальные поисковые сайты в веб-пространстве, помогающие решить эту проблему.

При поиске в пиринговых сетях тема полноты поиска отодвигается на второй план, главная же задача — быстрое и эффективное нахождение наиболее релевантных откликов на запрос, передаваемый от рабочей станции всей сети. В частности, актуальная проблема — уменьшение сетевого трафика, порождаемого запросом (например, пересылки запроса по многочисленным рабочим станциям), и в то же время получение наилучших характеристик выдаваемых документов, т.е. получение качественного результата.

Приемлемое качество поиска в пиринговых сетях на сегодняшний день обеспечивают лишь специализированные, централизованно наполняемые, поисковые веб-сайты, естественно, работающие по протоколу HTTP. Например, для файлообменной сети eMule таким поисковым сервером является сайт Figator.com (рис. 6), а для сети BitTorrent — сайт isoHunt.com (рис. 7). Как и для файлообменных сетей, для этих серверов особо актуальными и критичными являются проблемы качества и достоверности предоставляемого контента, фальсификация файлов и распространение фальшивых ресурсов, вирусов, «троянских коней», возможность фальсификации ID рабочих станций.

«Глубинный» веб

Кроме видимой для поисковых систем части WWW существует огромное количество веб-страниц, которые ими не охватываются. При этом доступ пользователя к таким ресурсам вполне возможен без ввода логинов и паролей. Как правило, эти веб-страницы доступны в Интернете, однако выйти на них трудно, а порой невозможно, если не знать точного адреса (или особого правила доступа). Эти ресурсы уже много лет как имеют собственное название — «глубинный» (*deep*) веб, которое ввел Джилл Иллсворт (Jill Ellsworth) в 1994 году, обозначив им документы, недоступные для обычных поисковых систем.

Сегодня такие ресурсы называют также «невидимым», или «скрытым» (*invisible*), вебом. Они чаще всего охватывают динамически формируемые веб-страницы, содержание которых хранится в базах данных и доступно лишь по запросам пользователей. Иногда для доступа к подобным страницам используется так называемый тест Тьюринга (или тест на разумность): предлагается решить арифметическую задачу, загадку или попросту ввести в определенное поле последовательность символов, изображенную графически. В 2000 году американская компания BrightPlanet (www.brightplanet.com) опубликовала сенсационный доклад, в котором утверждается, что в веб-пространстве в сотни раз больше страниц, чем их удалось проиндексировать самыми популярными поисковыми системами. Эта же компания разработала программу LexiBot, которая позволяет скани-

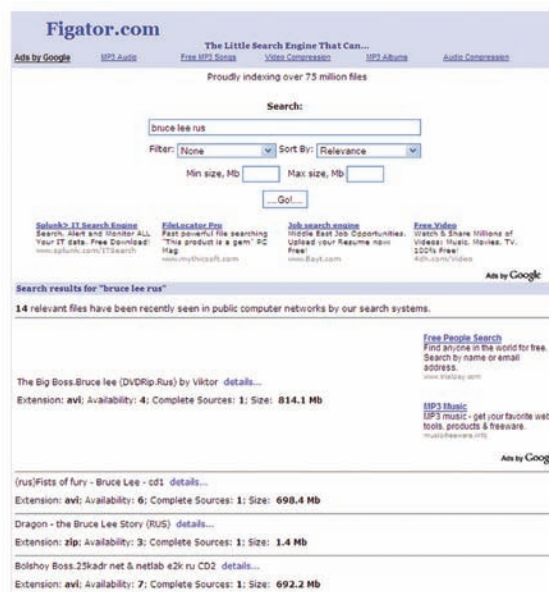


Рис. 6. Поиск в файлообменной сети eMule — сервер Figator.com

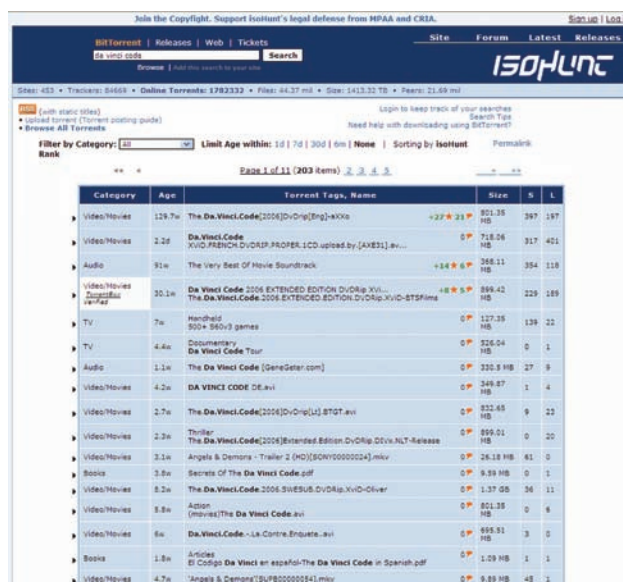


Рис. 7. Поиск торрентов в сети BitTorrent — сайт isoHunt.com

ровать некоторые динамические веб-страницы, формируемые из баз данных, и, запустив ее, получила неожиданные данные.

Основатель BrightPlanet Майкл Бергман (Michael K. Bergman) выделил 12 разновидностей «скрытых» веб-ресурсов (www.leidenuniv.nl/ub/biv/specials.htm), относящихся к классу онлайн-баз данных. В списке оказались как традиционные базы данных (патенты, медицина и финансы), так и публичные ресурсы — объявления о поиске работы, чаты, библиотеки, справочники. Бергман причислил к «скрытым» ресурсам и специализированные поисковые системы, которые обслуживают определенные отрасли или рынки, базы данных которых не включаются в глобальные каталоги традиционных поисковых служб.

К «скрытому» вебу также относятся многочисленные системы интерактивного взаимодействия с пользователями — помощи, консультирования, обучения, требующие участия людей для формирования

динамических ответов от серверов. К ним также можно отнести и закрытую (полностью или частично) информацию, доступную пользователям только с определенных адресов или групп адресов, иногда городов или стран.

К «скрытой» части веб-пространства многие причисляют и веб-страницы, зарегистрированные на бесплатных серверах, которые индексируются, в лучшем случае, лишь частично — поисковые системы во избежание рекламного спама не стремятся обходить их в полном объеме. Целая категория так называемых серых документов, размещенных в среде динамических систем управления контентом (Dynamic Content Management Systems), также относится к «глубинному» вебу. В поисковых системах обычно ограничивается глубина индексирования таких сайтов во избежание возможного циклического просмотра одних и тех же страниц. И, конечно же, «скрытыми» оказываются веб-ресурсы, создатели которых не оповещают кого-либо о создании этих ресурсов.

Информационные ресурсы могут рекламироваться на поисковых сайтах различными способами, самый популярный из которых — оплачиваемые результаты поиска.

В «глубинном» вебе существует множество альтернатив коммерческим базам данных типа Dialog или Lexis-Nexis. К примеру, базы данных с законодательными документами Украины или России (системы «Рада» или «Кодекс», соответственно) вполне можно отнести к такой категории, ведь размещенные в них сотни тысяч документов, доступные для свободного просмотра, не попадают в индексы глобальных сетевых информационно-поисковых систем.

Характеристики качества поиска

Качество работы информационно-поисковой системы — понятие достаточно объективное. В рамках теории информационного поиска

уже давно определены основные характеристики, ставшие де-факто стандартами в этой области. Конечно, такие сервисы, как поддержка бесплатных почтовых ящиков, бесплатный хостинг, передача SMS привлекают пользователей, но, похоже, никогда самый популярный портал, предлагающий все эти возможности (и не только их), не станет популярнее поисковых систем, обеспечивающих нахождение необходимой пользователю информации.

Существует много характеристик поиска, из которых две признаны основными — это *полнота* (recall) и *точность* (precision). Много внимания в настоящее время отводится также такой смысловой характеристике, как *pertинентность*. Она означает соответствие найденных документов информационным потребностям пользователя, а не формальному соответствию документа запросу. Для пользователя этот параметр имеет решающее значение. При этом следует учитывать, что формальный запрос к системе является предметом творческого осмысления информацион-

ной потребности и не всегда точно отражает последнюю.

Неумение большинства пользователей правильно формулировать запросы и получать приемлемые объемы откликов породило в конце XX столетия мысль относительно веба, как об огромной информационной свалке. Достижение высокой пертинентности — основное поле конкурентной борьбы современных поисковых систем. Именно для максимального удовлетворения информационных потребностей пользователей сетевые информационно-поисковые системы сегодня максимально интеллектуализируются — получили широкое применение технологии и методы семантических и нейронных сетей, Text Mining.

Деньги поисковых систем

В 2007 году бренд компании Google, стоимость которой достигла к тому времени \$66,43 млрд., возглавил рейтинг самых дорогих брендов года BrandZ Top-100, составленный исследовательской компанией Millward Brown Optimor для газеты Financial Times. Стоимость этого бренда в предкризисный год возросла на 77%, что позволило Google подняться с 7-го места в 2006-м сразу на первое. На втором месте в списке находится бренд компании General Electric, стоимость которой — \$61,88 млрд., а на третьем месте — Microsoft (\$56,9 млрд.). Coca-Cola заняла четвертую позицию этого списка.

Естественно, возникает вопрос: каким образом поисковые системы могут достигать финансовых успехов, чем определяются их основные финансовые потоки? Ответ очевиден — деньги бесплатных поисковых служб приходят от оплаченной рекламы, напрямую зависят от посещаемости, используемости этих служб. В свою очередь, посещаемость определяется полезностью поискового сервиса — релевантностью, оперативностью, удобством интерфейса, охватом и фильтрацией огромного количества информации, знаний.

В частности, коммерциализация большинства современных поисковых служб базируется на четкой бизнес-модели: клиенты бесплатно пользуются различными информационными услугами (основная из которых — поиск), а доход поступает от рекламодателей за рекламу, предоставляемую потребителям этих услуг.

Информационные ресурсы могут рекламироваться на поисковых сайтах различными способами, самый популярный из которых — *оплачиваемые результаты поиска*. В этих случаях владельцы информационных ресурсов платят деньги за то, что их ресурсы, если они окажутся релевантными запросу пользователя, *будут показаны на первой странице* вывода

результатов поиска. Такая политика порой может противоречить потребностям конечных пользователей, ведь, казалось бы, естественные результаты, которые обеспечиваются алгоритмами поисковых систем, являются более релевантными в отношении запросов большинства пользователей, чем результаты оплаченного поиска. Но в действительности во многих случаях оплаченный поиск является фактически более pertinentным, обеспечивая более *полезные результаты*. Сей феномен является одним из удивительных совпадений стремлений: поисковых служб — увеличивать свои доходы и конечных пользователей — получать полезные документы в результате поиска по запросу.

С другой стороны, рекламодатели наказываются за попытки привлечь трафик, который является нерелевантным, который не превратит пользователей поисковой службы в потребителей. Рекламодатели, ищущие способ развить свой бренд, при использовании оплаченного поиска получают гораздо больше за свои доллары, если их контент является релевантным и интересным пользователям.

В качестве рекламных инструментов в Google используются сервисы AdWords и AdSense. Рекламодатели, использующие Google AdWords, создают объявления для привлечения целенаправленного трафика на свои сайты и получения перспективных клиентов. Издатели-партнеры Google распространяют эти объявления, ориентированные на релевантные результаты поиска, с помощью программы Google AdSense, которая позволяет издателям получать свою долю прибыли от активного просмотра объявлений читателями (рис. 8).

Для владельцев сайтов, которые хотят иметь большую степень контроля над своей корпоративной информационной системой или поиском по сайту, Google разработал масштабируемое программно-техническое решение Google Search Appliance, позволяющее получать точные результаты поиска по любому объему документов.

Кроме того, компания пошла по пути интеграции в своей системе разнообразных инструментов — начиная от почтового сервиса и «мессенжера», заканчивая блогами, фотоальбомами, документами, автоматическими переводчиками, системой электронных платежей и веб-аналитики. Известно, что у Google есть другие источники дохода, например, поставка средств для корпоративного поиска, но их доля практически не влияет на основную бизнес-модель.

В крупнейшей поисковой системе по Рунету — «Яндексе» — практически не применяется в явном виде технология оплаченных результатов поиска, широко используемая в Google. Роль такой рекламы играют другие системы — «Яндекс.Директ» и «Медийно-контекстный банкер».

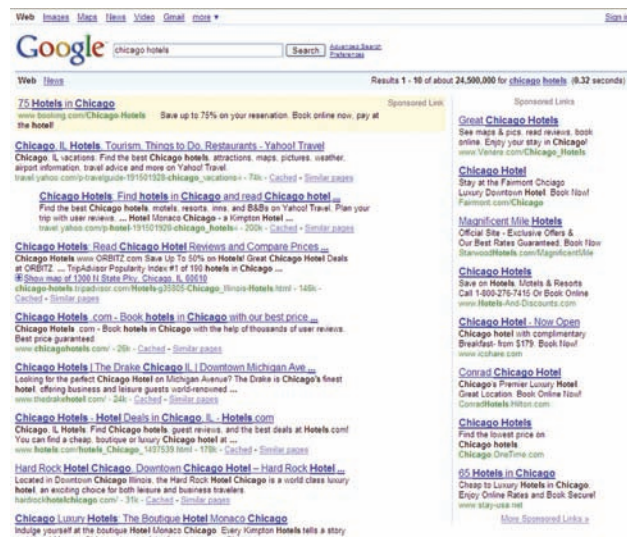


Рис. 8. Страница вывода результатов Google. Перед выдачей результатов — светло-желтый блок — оплаченный результат поиска (Sponsored Link). В правой колонке — несколько оплаченных ссылок, которые могут размещаться даже на нескольких страницах

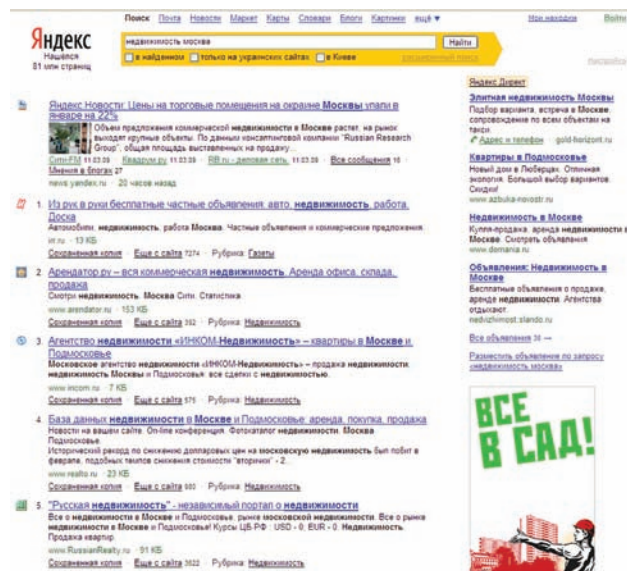


Рис. 9. Страница вывода результатов «Яндекса» — правая колонка — «Яндекс.Директ» и «медийно-контекстный баннер»

На рис. 9 можно видеть результаты поиска по запросу «недвижимость Москвы». Оба рекламных инструмента представлены в правой колонке. В первом случае рекламодатель платит поисковой системе за каждый переход по размещаемой рекламе, а во втором — по так называемой СРМ-модели (оплата за показы). Кроме того, для рекламы интернет-магазинов в «Яндексе» существует еще один инструмент — «Яндекс.Маркет».

«Яндекс.Директ» — это инструмент для размещения контекстных рекламных объявлений на страницах «Яндекса» и на сайтах-участниках его рекламной сети. Контекстная реклама показывает посетителю лишь такие объявления, которые прямо связаны с тем, что его в данный момент интересует. Причем эти интересы пользователя им самим явно сформулированы в запросах к поисковой системе.

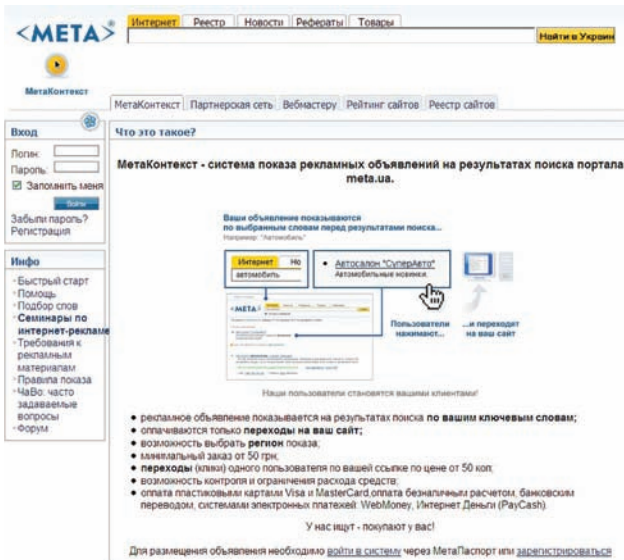


Рис. 10. Система показа рекламных объявлений в результатах поиска на портале meta.ua

«Медийно-контекстный баннер» показывается тем пользователям, которых в данный момент интересует тема рекламного предложения (это определяется по контексту запроса). Соответствие тематики каждой страницы той или иной теме определяется автоматически, на основе заданного набора слов и словосочетаний. Медийно-контекстный баннер размещается по наборам ключевых слов и словосочетаний по определенной тематике (так называемым тематическим пакетам). Кроме того, рекламодатели могут использовать географический таргетинг, чтобы баннер увидели пользователи того региона, в котором заинтересован рекламодатель. Показы баннера организованы таким образом, чтобы охватить максимальное количество пользователей, интересующихся заявленной темой, в результате чего баннер совмещает в себе преимущества и контекстной, и медийной рекламы. Кроме того, рекламный баннер показывается на сайтах — партнерах «Яндекса».

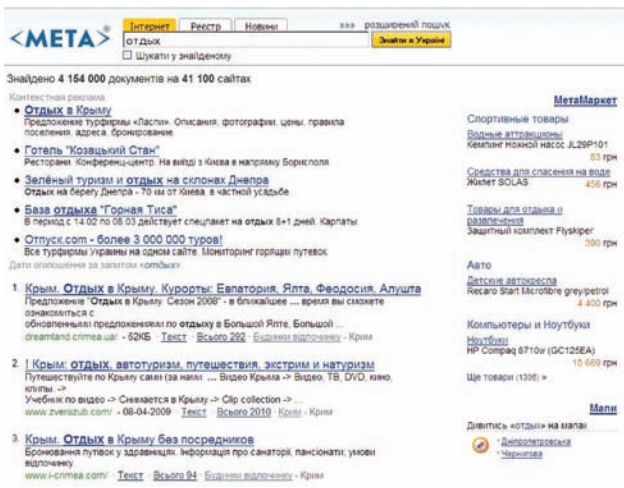


Рис. 11. Пример контекстной рекламы по запросу «отдых» на портале meta.ua

Многие украинские поисковые системы также основаны опцией контекстной рекламы. В частности, крупная украинская система «Мета» предоставляет возможность рекламы в результатах поиска, соответствующих ключевым словам, вводимым пользователем в рамках сервиса «МетаКонтекст» (рис. 10). При этом развернута целая партнерская сеть (рис. 11), которая позволяет владельцам веб-сайтов, размещающих контекстные блоки, предоставляемые «МетаКонтекстом», зарабатывать до 50% от стоимости кликов (переходов на веб-сайты рекламодателей).

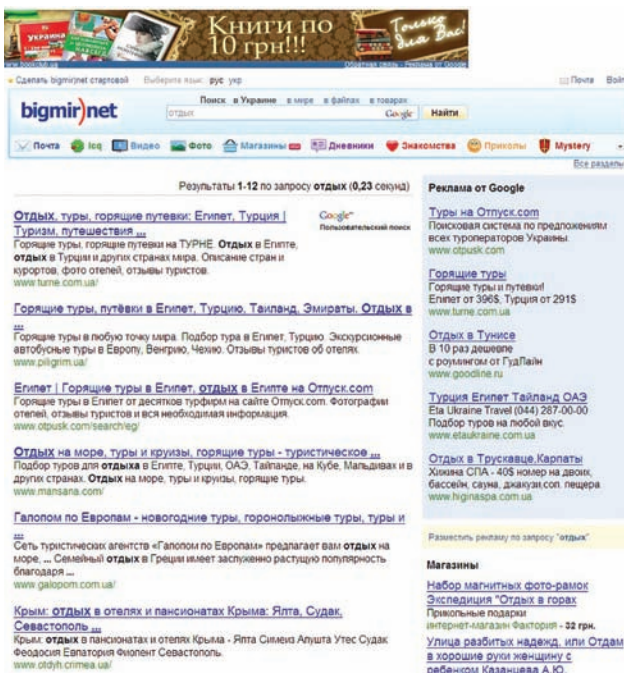


Рис. 12. Результаты поиска на сайте bigmir.net с контекстной рекламой от Google

Один из крупнейших украинских порталов «Бигмир» (www.bigmir.net) также в какой-то мере можно считать поисковой системой. Действительно, среди его функций есть информационный поиск, построенный на технологии Google (рис. 12). Контекстная реклама на «Бигмире», также «позаимствованная» у Google, реализована в двух основных представлениях — в виде контекстных ссылок и контекстных баннеров, которые формируются при вводе указанных рекламодателем ключевых слов. При этом показы ссылок бесплатны, оплачиваются только переходы на целевые сайты, при этом цены кликов формируются на аукционной основе.

Поисковый бизнес, модели которого не укладываются в представления обычных бизнесменов, становится вполне реальным и ощутимым. Сегодня поисковые службы превращаются в значимые бизнес-проекты, доходы от которых оказываются лидирующими в мире.

Дмитрий ЛАНДЭ,

доктор технических наук, заместитель директора
Информационного центра «ЭЛВИСТИ»
dwl@visti.net