

Проблемно-ориентированные страницы на Internet: экономика и финансы

Бухвалов А.В.

Европейский университет в С.-Петербурге
С.-Петербургский государственный университет

С.Петербург, Россия

E-mail: bukh@por3.com.ru

1 Введение

Создание электронных библиотек, электронных коллекций, баз знаний и поисковых систем в различных предметных областях является насущной задачей развития Internet. Каждая предметная область требует своего подхода к проектированию подобной системы, начиная от общей идеи и кончая программистскими средствами. Я рассмотрю специфические проблемы по созданию таких систем в случае **экономики и финансов**.

2 Состояние дел

Академические экономика и финансы богато представлены на Internet. Практически все информационные сайты в данной предметной области поддерживаются иностранными центрами и, без единого известного мне исключения, являются англоязычными. Факт англоязычности не является препятствием для исследователя, но составляет трудность для многих отечественных студентов. С последним обстоятельством мы не можем не считаться при наших собственных разработках.

Наибольший интерес составляет система сетей NetEc, BibEc, WoPEc, CodEc, WebEc, HoPEc:

<http://netec.mcc.ac.uk/NetEc.html>

и сеть по академическим финансам на сайте *Journal of Finance*:

<http://www.cob.ohio-state.edu/~fin/journal/>.

Указанные сайты (и многие другие) предлагают рабочие версии статей (working papers, discussion papers) ведущих экономистов, их домашние страницы, курсы лекций и программное обеспечение, работающие в режиме online (часто, написанные на языке Java).

Бизнес-сайты обычно предоставляют либо платные консультационные услуги, либо — проведение финансовых операций в режиме online (таковы, в своем большинстве, сайты и зарубежных и отечественных финансовых агентств). Многие из таких сайтов содержат и серьезную

Первая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
19 - 21 октября 1999 г., Санкт-Петербург

теоретическую информацию, а иногда и первичные данные — см., например, великолепную группу сайтов компании JP Morgan, Inc, путешествие по которым можно начать с узла

<http://www.jpmorgan.com>.

3 Недостатки существующих подходов

Объем информации, который можно получить, используя только упомянутые сети, столь велик, что уже в нем легко заплутаться. По моему мнению, традиционные системы типа каталога, тезауруса или поисковой машины (с необходимостью угадать заложенную в базу комбинацию терминов) неудовлетворительны для представления знаний в таких разветвленных областях как экономика и финансы.

Системы, спроектированные наподобие **желтых страниц**, или **каталога**, в принципе, полезны как источник первичной информации, которым, в частности, активно пользовался и автор этих строк. Такого рода система, отчасти, напоминает предлагаемую мной ниже систему **проблемно-ориентированных страниц**. Обе системы состоят из нескольких уровней рубрик, уточняющих сферу интереса пользователя. Однако в системах типа желтые страницы эти рубрики, в известных автору случаях, недостаточно детально разработаны, так как существующие системы носят универсальный характер. Что еще более важно, на последнем уровне каталог дает, как правило, ссылки на домашние страницы сайтов, многие из которых содержат сотни мегабайт разнообразной информации. Это делает такого рода подход неудобным для конечного пользователя, который после локализации своей проблемы в каталоге должен опять начать поиски сначала, хотя и в рамках некоторого ограниченного пространства. При этом нет гарантии, что ответ на его вопрос вообще имеется на данном сайте или, вообще, на одном из сайтов, упомянутых в каталоге на конечном шагу.

Типичным примером описанного подхода являются универсальные каталожные системы

<http://www.yahoo.com> и <http://www.list.ru>.

Отметим, что близкий прообраз того, что мы имеем в виду под проблемно-ориентированными страницами, можно найти на втором из упомянутых сайтов в каталоге Компьютеры. Особенно выделим сайт iXBT Hardware. Содержательный успех в создании этой части каталога связан с тем, что сотрудники, создававшие эти страницы, являются специалистами в данной предметной области.

В остальных случаях это, очевидно, было не так. К осуждению этой проблемы мы еще вернемся ниже.

Тезаурус и поисковая машина неудобны, так как пользователь должен, необходимым образом, угадать заложённую в базу комбинацию терминов. При этом подходящие комбинации терминов сплошь и рядом могут доставлять, с точки зрения рассматриваемой задачи, лишь информационный шум. Такого рода системы также полезны скорее для проектировщика предметной среды, чем для конечного пользователя.

4 Проблемно-ориентированные страницы

Автором предлагается принципиально новая концепция построения системы обработки информации с использованием **проблемно-ориентированного подхода**, где поиск информации производится с помощью дерева задач с уточнением деталей запроса при углублении поиска и выходом на конкретную страницу в Internet.

Например, ключевое слово *валюта* может быть связано с различными запросами пользователя, ответы на которые надо искать в абсолютно разных местах на Internet. Пользователь может интересоваться *текущим обменным курсом* двух конкретных валют (иногда с уточнением места обмена). Пользователь может интересоваться также получением *временного ряда* обменных курсов для аналитической работы (это возможно, но уже в другом месте сети). Пользователь может интересоваться различными теоретическими *моделями* формирования и динамики обменных курсов и т.п. Здесь мы выделили дерево понятий глубины три, где на последнем, третьем, уровне происходит уже конкретная отсылка на страницу Internet.

Приведем более формальную блок-схему для первичного входа *Валюта*:

- *Обменный курс*
 - *Текущее значение*
 - * *Обменный курс рубля*
 - *Россия, ММВБ*
 - *Россия, ЦБР*
 - ...
 - * *Универсальный центр*
 - * ...
 - *Временной ряд*
 - * *Выбор пары валют*
 - *Сайты...*
- *Теория и моделирование*
 - ...

Работа по созданию проблемно-ориентированных страниц распадается на аналитическую работу по проектированию дерева задач, что требует понимания предмета, и нахождения конкретных страниц с необходимой информацией (это требует затем постоянного оперативного мониторинга). Будем называть лицо, осуществляющее проектирование и мониторинг такой системы, *модератором*. Очевидно, что требования к квалификации модератора и текущие затраты на проект являются достаточно высокими. Как мы отметили выше, в предметной области компьютерного железа задача поиска модератора вполне реалистична. Значит, эта задача может быть решена и в

любой другой предметной области, хотя в такой области, как экономика и финансы, она требует больших капиталовложений.

Очевидно, что разные предметные области требуют различных решений и средств при проектировании веб-сайта. Системы, посвященные искусству или дизайну, будут выглядеть совершенно иначе, чем системы, посвященные позитивным наукам. Абсолютно различно выглядит и программистский инструментарий. Так как оставшаяся часть статьи посвящена исключительно предметной области экономики и финансов, то мы не будем касаться мультимедийных эффектов, вращающейся трехмерной графики и тому подобных вещей, которые не имеют прямого отношения к специфическому представлению экономической и финансовой информации.

5 Проблемы представления экономической информации

5.1 Содержание, роль модератора

Сайт, посвященный предметной области, представляющей определенную науку, должен курироваться модератором, являющемся специалистом в данной предметной области. Такого человека трудно найти в случае экономики и финансов — наук, заново становящихся в России после 1991 г. Для подсчета российских сайтов по экономике, где есть, что положить на Internet, достаточно пальцев одной руки. Не один из этих сайтов не претендует на академически широкий охват представленной информации.

В данной статье я вообще не касаюсь информации, коммерчески предоставляемой государственными и частными агентствами на Internet. Это относится к области электронной коммерции.

Сайты, посвященные разного рода бизнес и финансовой информации, как правило, весьма неоднородны, а иногда и просто низки, по качеству содержания. Очень симпатичный проект Корпоративные финансы

<http://www.cfin.ru>

с Николаем Чувахиным (одним из наиболее опытных российских модераторов) не дотягивает пока до уровня выше популязаторского. Основной причиной частичной неудачи представляется эклектика авторов в понимании финансов, где смешиваются вместе финансовый менеджмент, бухучет, АСУ и математические методы. При этом забывается экономика, частью которой являются финансы.

К сожалению, многие предметные сайты становятся местом для самиздатского самовыражения неудовлетворенных литературных амбиций. Такие сайты содержат крайне мало позитивной информации, вмещая, одновременно, десятки мегабайтов журналистики. Это нельзя запретить, но нельзя и одобрить.

Примером неудачного сайта такого рода является проект Планета КИС (КИС — новое название АСУ)

<http://kis.pcweek.ru>.

Могу рекомендовать авторам посмотреть на сайт компании SAP — мирового лидера в области АСУ и крупнейшей программистской фирмы в Европе

<http://www.sap.com>.

Как и положено, разработчики из SAP приводят массу примеров удачного внедрения практически во всех странах мира. Сделано это, однако, по-деловому, а не в стиле

— Мне вспоминается пример одной ФПП... — без указания, о чем вообще идет речь. Я надеюсь, что АСУ внедрялось не в ФПП, а на предприятии, и это предприятие имело название и адрес.

Таким образом, задача создания сайтов, специализирующихся на экономике и финансах, остается актуальной и открытой. Важным является также создание системы, ориентированной на информацию об экономике переходного периода и, прежде всего — об экономике России. Многие данные рискуют быть навсегда потеряны, так как люди, ими обладавшие, использовали их в сиюминутных целях, не обращая внимания на их фундаментальную значимость для исследования переходных процессов в России.

5.2 Жанр web-сайта

Каждому web-сайту, посвященному предметной области, необходимо определиться со своим жанром, хотя и эклектическое смешение жанров не всегда противопоказано. В качестве основных возможных жанров я имею в виду

- машина ссылок;
- интерактивный дискуссионный клуб;
- депозитарий материалов (книги, статьи, данные).

И система типа желтые страницы и проблемно-ориентированные страницы являются примерами машины ссылок. Дискуссионный клуб интересная, но трудоемкая форма организации профессионального интереса к предметной области. Электронные конференции и группы новостей являются примерами этого жанра. Депозитарий предполагает, что данные хранятся непосредственно на данном сайте.

Реально мы часто имеем смешение жанров, так как крупный депозитарий, как правило, дает отсылки (links) на другие сайты — всю информацию нельзя хранить у себя не только из-за ограниченности места, но и из-за авторских прав.

Смесь депозитария с дискуссионным клубом не всегда удачна. Сайт проекта Московский Либертариум

<http://www.libertarium.ru>.

является, содержательно, одним из двух лучших в России (по экономике). На сайте сосредоточены оригинальные современные работы наших соотечественников и переводы классических книг Людвиг фон Мизеса и Фридриха фон Хайека, виднейших представителей экономического либерализма.

В мае 1999 г. скромный деловой дизайн сайта был сменен на другой скромный дизайн, но уже с внедрением дискуссионного клуба на каждом шагу. Каждая ссылка на материал сопровождается парой кнопок для интерактивного общения. В качестве одного из модераторов выступает Борис Львин, один из наиболее интересных отечественных экономистов (сотрудник российской дирекции ВМФ, а затем — Всемирного банка). Дискуссии, однако, пока не получилось, а соединение с сайтом настолько замедлилось, что поставлена под угрозу главная содержательная задача сайта — продвижение либеральной идеи. По моему мнению дискуссия не получилась из-за искусственного привязывания тематики дискуссии к помещенным на сайте материалам. Не попробовать ли разделить на составляющие?

Российский пользователь отрезан от многих материалов, широко доступных на Западе. Это касается, прежде всего, электронных библиотек. До 1992 г. мы не имели в России журналов по экономике и финансам по идеологическим соображениям, навязанным нам сверху. Теперь мы отрезаны от мировой науки из-за системы финансирования нашего образования. Система JSTOR содержит online практически все журналы (кроме последних выпусков). Система ECONbase от Elsevier содержит текущие выпуски многих основных экономических и финансовых журналов. Все это нам недоступно. Подписка на услуги этих депозитариев должна производиться университетами, которым необходимо изыскать спонсоров под программу подписки. После этого каждый преподаватель и студент, подключенный через университетскую сеть к Internet, сможет переписать любые статьи для личного пользования (в PDF-формате).

5.3 Online/Offline. Электронные библиотеки

Для российского пользователя важнейшей проблемой является оптимальное разделение между работой online и offline. Ясно, что желательно иметь возможность переписывать материалы для offline работы. Я пользуюсь для этой цели программой WebZIP от SpiderSoft

<http://www.spidersoft.com>.

К сожалению, проектанты супермодных российских сайтов сделали многие свои страницы недоступными для классических HTTP-протоколов, реализованных в WebZIP. Примером является, в частности, новый дизайн сайта Московский Либертариум. Обязаны ли мы читать Хайека online? Сделано ли это намеренно? Сайт проектировался профессионалами Internet из Института Коммерческой Инженерии, поэтому сделанное вызывает удивление. Разумным и доступным приемом является архивирование материалов депозитария.

Раньше представлялся целесообразным прием разбиения документа на небольшие web-страницы, что сокращало время загрузки. После появления у MS Internet Explorer 5.0 опции *Сохранить как* целую web-страницу (со всей связанной графикой) такое разбиение стало большим неудобством. Кстати, Netscape Communicator (v.4.61 на момент написания этой статьи) не имеет этого удобства.

Хочется отметить новый российский сайт Института экономики переходного периода (институт Гайдара)

<http://www.iet.ru>,

который предоставляет библиотеку публикаций Института в виде ZIP-файлов, что реально позволяет пользоваться ими.

5.4 СУБД и форматы данных

Необходимо остановиться также на вопросе об адекватности программных решений задачам организации экономических данных.

Российские программисты привыкли ассоциировать экономику с торговлей, а финансы — с бухгалтером. И то и другое программируется в виде базы данных, организованной с помощью Fox Pro. Экономика и финансы имеют, однако, мало общего с торговлей и бухгалтером. В связи с этим надо поставить важную задачу о формате данных предоставляемых конечному пользователю.

Системы управления базами данных (СУБД) являются преуспевающей областью программирования. Крупнейший поставщик СУБД — фирма Oracle — является второй по величине, после Microsoft, корпорацией в мире среди всех фирм, специализирующейся в области создания software. Поэтому сказанное ниже, вовсе не против баз данных, а против их применения там, где они мешают.

Массовое использование в отечественной практике нелегальных копий пакетов различных СУБД привело к неправильной мысли о бесплатности и общедоступности любого программного обеспечения для любого пользователя. Массированное издание в конце 1970-х годов переводов книг Дж.Мартина о роли СУБД также дезориентировало отечественных специалистов. Читателями было незамечено, что речь исключительно идет о применении СУБД в управлении информационными потоками розничных (торговых) операций коммерческих и финансовых учреждений. Это действительно важная сфера применения СУБД. СУБД полезны при необходимости мониторинга информационных потоков, поступающих в локальную сеть из различных источников. СУБД полезны также для обработки первичной статистической информации, поступающей из разных источников (часто несинхронно). Здесь СУБД могут быть использованы (и используются) от уровня сети магазинов до уровня сбора макроэкономической статистики. Однако сами базы данных не являются приемлемой формой представления экономической и финансовой информации для конечного пользователя, а тогда и на Internet.

Предоставление пользователям Internet файлов непосредственно в формате конкретного СУБД нецелесообразно, так как предполагать легальное владение пользователем какого-либо программного обеспечения, кроме Microsoft Office, нереалистично. Использование ODBC и SQL стандартов в целом полезно, но едва ли удобно для массового пользователя экономической информации на Internet. К счастью, эти приемы на практике не применяются.

Распаковка баз данных в HTML-страницы в режиме online является популярным приемом среди российских программистов. Кажущимся плюсом является возможность увидеть результат в виде таблицы, соответствующей запросу пользователя (например, данные о ставке процента за определенный период). Подобного рода таблицы, как правило, имеют мало интереса для профессионального пользователя. Пользователь хочет не посмотреть на данные, а работать с ними. Для этого таблица в виде HTML-файла малоудобна — универсальным форматом, используемым статистическими (в том числе, MS Excel) и финансовыми пакетами, является формат CSV — Comma Separated Values. В силу неоднозначности HTML как языка перекодировка таблиц в CSV-формат малореальна (и нереальна для общего пользователя; хотя, в целом, хочется подчеркнуть полезность языка AWK для модификации HTML-файлов в пакетном режиме MS DOS или UNIX; в последнем случае обычно используют более развитый язык скриптов — Perl).

Примером неудачного использования технологии представления данных, описанной выше, является Центральный Банк России

<http://www.cbr.ru>.

Эта технология используется и другими агентствами, например, Росбизнесконсалтинг

<http://www.rbc.ru>,

но в этом случае такая технология оправдана — online можно получить информацию бесплатно в неудобном формате, но зарегистрированные пользователи имеют доступ к полноценной информации на ftp-сайте (это нормальное давление на потребителя).

Провайдеры данных, заботящиеся о потребителях, обычно предоставляют, как возможность посмотреть на данные в режиме online, так и возможность переписать всю информацию в заархивированном виде. Положительный пример подают здесь американские федеральные сайты, см., например, сайт Совета Директоров Федеральной Резервной Системы

<http://www.bog.frb.fed.us>,

предоставляющие богатейшую информацию бесплатно в обоих форматах (сразу предлагается переписать *All historical data files* в виде ZIP-файла).

5.5 Файлы сценариев

В качестве общего принципа для предметных отраслей, связанных с наукой, я выдвинул бы скромность программистских средств при реализации проекта. Например, использование сценариев (CGI-файлы) может сослужить плохую службу для провайдера информации. В качестве примера приведу знаменитое издательство MIT Press

<http://www.mitpress.com>.

Издательство должно быть заинтересовано, чтобы потенциальному покупателю было легко составить таблицу для заказа нескольких десятков наименований книг со всеми необходимыми данными. Естественным путем было бы использование HTML-файла. Однако CGI-технология поиска дает возможность получить информацию только по каждой книге в отдельности. По всей видимости, мы опять имеем дело с тем, что провайдеру программисту удобно работать с базой данных, тогда как провайдеру-маркетологу должно быть ясно, что такое решение только мешает процессу заказа.

Некорректная работа с CGI-файлами типична как раз для компаний, рекомендующих себя в качестве профессионалов в области Internet. Так, компания Numa Financial Systems

<http://www.numa.com>.

представляет себя как специалиста в области инвестиций и Internet. Тем не менее, попытка переписать список литературы по производным ценным бумагам с помощью WebZIP (пять коров на TUCOWS) приводит к получению тысяч файлов с заголовками типа

[cgi-bin/numa/contact.pl?form=DerivativesBookshop&sub=Ref:3670](http://www.numa.com/cgi-bin/numa/contact.pl?form=DerivativesBookshop&sub=Ref:3670).

Длина их имен, зачастую, превышает предел MS Windows (и длину строки, при выбранном формате данной публикации, поэтому пришлось сделать перенос). Хуже того, использование метода HEAD (наряду с GET) HTTP-протокола, привело к запуску программы электронной почты на сервере компании Numa, о чем компания немедленно пожаловалась моему провайдеру Internet (февраль 1999 г.). Причина же в излишествах при проектировании оформления сайта, который, в содержательном смысле, содержит вполне симпатичные, работающие online, финансовые калькуляторы, реализованные на языке Java.

6 Заключение

В настоящий момент автор работает одновременно над проектами **Проблемно-ориентированные страницы: экономика и финансы и Экономика перехода в России**.

Благодарности. Значительная часть аналитической работы на Internet была проведена автором в Институте переходных экономик Банка Финляндии (VOFIT). Данная публикация была частично финансово поддержана в рамках гранта НВС-812 Института Открытое Общество (фонд Сороса). Автор выражает также благодарность сотрудникам университета Тампере (Финляндия), осуществляющим техническую поддержку web-страниц, спонсированных автором и посвященных курсу *Russian Economic Developments*

<http://www.uta.fi/laitokset/iss/bukhvalov/>.