

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ INTERNET*

Ю. И. Шокин, А. М. Федотов

Институт вычислительных технологий СО РАН

Новосибирск, Россия

e-mail: shokin@adm.ict.nsc.ru, fedotov@adm.ict.nsc.ru

The leading edge techniques of the presentation of informational resources in the world informational environment (space) emerging due to the Internet are discussed.

1. Введение

Информация была и остается одним из важнейших стратегических ресурсов государства, и наибольший экономический и социальный успех сопутствует тем, кто активно использует современные средства компьютерных телекоммуникаций и сетевые приложения (электронную почту, информационные системы, визуализацию, моделирование, компьютерную графику, дистанционное обучение, мультимедиа, телеконференции и многое другое). Все эти приложения и связанные с ними прикладные задачи требуют создания нового класса телекоммуникационных сетей и систем управления информационными ресурсами. Перенесенные на электронные носители информационные ресурсы приобретают качественно новое состояние, т. е. становятся активными, и в этом качестве являются доступными, в том числе и дистанционно с помощью компьютерных сетей общего пользования. Доступная для оперативного воспроизведения средствами компьютерной обработки информация превращается в важнейший фактор социального развития общества на основе применения новых технологий. Индустрия средств передачи информации, баз данных, системных и прикладных программ, информационно-аналитических услуг формирует динамичный рынок, объединяющий широкие слои населения, промышленность, науку, образование и административные структуры в единую информационную среду. Финансовые вложения в сферу информационных технологий и услуг в ряде стран превысили по объемам вложения в развитие энергетики, благодаря, как правило, рекордным темпам их косвенной окупаемости.

В современных условиях развития общества, усложнения технической и социальной инфраструктуры информация становится таким же стратегическим ресурсом, как традиционные материальные и энергетические ресурсы. Современные информационные технологии, позволяющие создавать, хранить, перерабатывать и обеспечивать эффективные способы предоставления информационных ресурсов потребителю, стали важным фактором жизни общества и средством повышения эффективности управления всеми сферами

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, грант № 97-07-90372.

© Ю. И. Шокин, А. М. Федотов, 1997.

общественной деятельности. Уровень информатизации становится одним из существенных факторов успешного экономического развития и конкурентоспособности региона как на внутреннем, так и на внешнем рынке. Осознание мировым сообществом роли информации как стратегического ресурса стимулировало разработки новых информационных технологий как для получения и переработки больших объемов информации, так и для ее хранения и предоставления пользователям. Первое место среди новых технологий занимают сетевые информационные технологии.

Идеология сети Internet, разрабатывавшаяся первоначально по инициативе министерства обороны США на случай ядерной войны, в настоящий момент стала всепроникающей идеологией, на которой строятся большинство современных информационных систем, предназначенных для широкого использования. Сегодня Internet объединяет множество разных сетей, миллионы компьютеров, сотни миллионов пользователей всех континентов, и, по разным оценкам, число таких пользователей увеличивается на 50–80 % ежегодно. Internet — уникальное средство поиска лиц, способных разделить с вами любые интересы и поддержать общение на любую волнующую вас тему. Internet обеспечивает оперативный доступ к поистине необозримым кладям информации по любой тематике (на десятках тысяч информационных серверов).

Информационные возможности SQL-баз данных и WWW-серверов весьма велики. В течение последних трех лет в мире наблюдается постоянное повышение интереса к этим технологиям (особенно к гипертекстовым) и ежегодный многократный рост поставщиков и пользователей информационных ресурсов. Естественные свойства организации информации через WWW-серверы позволяют систематизировать и в удобном виде представить пользователям очень разнообразную информацию, включая графические образы и звук, а также организовать системы поиска и оперативного обновления информации. WWW (аббревиатура от слов “World Wide Web”, что переводится как “Всемирная паутина”) — это мировая виртуальная файловая система типа hypermedia — “широкомасштабная гипермедиа-среда, ориентированная на предоставление универсального доступа к документам”. Основным достоинством этой технологии является организованная с достаточной простотой для пользователя система предоставления информации.

Сервис World Wide Web (WWW-технология) возник в 1989 году в европейском исследовательском центре CERN — Европейской лаборатории физики элементарных частиц (European Laboratory for Particle Physics in Geneva, Switzerland — <http://www.cern.ch/>). В настоящее время WWW-сервер исследовательского центра CERN (<http://www.w3.org/>) является интегрирующим сервером по поддержке WWW-технологий Internet.

Отметим, что WWW-технология является ровесником российских сетей Internet. В 1992 г. началось практическое применение этой технологии за пределами CERNa, а с конца 1993 г. начался колоссальный рост информационных ресурсов WWW, который привел к тому, что сегодня этот вид ресурсов Internet является самым популярным, наиболее динамично развивающимся. Отметим, что количество WWW-серверов удваивается в среднем каждые три месяца, причем количество серверов коммерческих организаций растет еще быстрее. Темпы роста WWW, проявляющиеся в увеличении как числа WWW-серверов, так и количества доступных через WWW документов и баз данных, превышают даже весьма высокие темпы роста общего количества компьютеров и пользователей Internet.

Однако, с точки зрения конкретного пользователя, например специалиста в конкретной предметной области, сама по себе Internet-технология недостаточна даже при наличии мощных средств поиска информации по образцам. Необходимы соответствующее структурирование информации, обеспечивающее удобство ведения поиска, и организация инфор-

мации способом, отвечающим функциональным запросам специалиста или лица, принимающего решения, а также предоставление возможности оперативного обновления данных и, кроме того, регламентация доступа к данным (поддержка авторизованного доступа) и обеспечение должного уровня информационной безопасности для каждого субъекта, решившего войти в информационное пространство.

2. Информационные технологии

Одним из важнейших аспектов развития информационно-сетевых технологий является создание и эксплуатация информационных ресурсов. Интеграция в мировое информационное пространство, помимо существования надежно функционирующей телекоммуникационной среды, предполагает в первую очередь предоставление как для мирового сообщества, так и для внутреннего использования собственных информационных ресурсов. Развитие собственных информационных ресурсов повышает оперативность управления и является основой систем поддержки принятия решений, кроме того, существенно снижает нагрузку на внешние каналы связи, поскольку при нормальном функционировании сети основные информационные потоки должны быть внутренними, а не внешними.

Информационные ресурсы представляют собой отдельные документы и массивы документов в информационных системах (библиотеках, архивах, фондах, банках данных, других видах информационных систем). К ним относятся рукописные, печатные и электронные издания, содержащие нормативную, распорядительную, фактографическую, справочную, аналитическую и другую информацию по различным направлениям общественной деятельности (законодательство, политика, демография, социальная сфера, наука, техника, технология и т. д.).

Информационные ресурсы могут использоваться для решения разнообразных научных и прикладных задач: от поиска необходимой информации до задач принятия управленческих решений. Для эффективного решения задач управления и поддержки принятия решений необходимо создание адекватной информационно-аналитической системы, моделирующей состояние управляемого объекта и его поведение при тех или иных управляющих воздействиях либо при их отсутствии.

С учетом складывающихся мировых тенденций в рассматриваемой предметной области предлагается при реализации различных задач, связанных с использованием информационных ресурсов, опираться на современные информационные технологии:

- базовые технологии Internet (WWW, IRC, E-mail и т. д.);
- информационную службу X.500;
- идеологию информационных хранилищ и архитектуру “клиент-сервер”;
- SQL-ориентированные инструментальные системы управления базами данных (СУБД ORACLE, INFORMIX, MS SQL и др.);
- CASE-технологии проектирования информационных систем и баз данных;
- ГИС-технологии;
- технологии создания и распространения информации на CD-ROM-носителях;
- развитие системы графических и текстовых редакторов;
- мультимедиа-технологии и технологии создания виртуальной реальности;
- Intranet-идеологии при создании корпоративных и учреждений информационных систем.

Перечисленные технологии позволяют обеспечить сопряжение различных типов информационных серверов (SQL, WWW и т. д.) как в составе одного информационного узла, так и находящихся в различных информационных узлах региональной компьютерной сети. Использование Internet/Intranet-технологий обеспечивает технологические предпосылки для построения территориально распределенных, проблемно-ориентированных и корпоративных информационных сетей и систем в виде логических надстроек над единой информационно-телекоммуникационной средой, а также органической интеграции в единое информационное пространство России и мирового сообщества.

Основу обслуживания пользователей информации составляют услуги, связанные с передачей информации (информационный транзит) и с предоставлением в пользование информационных ресурсов. Пользователи — граждане, органы государственной власти, органы местного самоуправления, организации и общественные объединения — обладают равными правами на доступ к информационным ресурсам и не обязаны обосновывать перед владельцем этих ресурсов необходимость получения запрашиваемой ими информации. Исключение составляет информация с ограниченным доступом.

Владельцы информационных ресурсов обеспечивают пользователей информацией из информационных ресурсов на основе законодательства, уставов указанных органов и организаций, положений, а также договоров.

В современных условиях, при недостаточно развитой системе телекоммуникаций и относительной дороговизне удаленного доступа для рядового пользователя, особенно перспективными представляются издания на компактных дисках. Однако, с точки зрения дальнейшего развития, следует основное внимание уделять поддержке и развитию удаленного доступа через компьютерную сеть к различным информационным ресурсам с использованием Internet/Intranet-технологий.

Для организации единого информационного пространства большое значение имеет разработка пакета нормативных документов, регламентирующих отношения, возникающие при формировании, передаче и использовании информационных ресурсов, защиты информации:

- о владении, распоряжении и пользовании информационными ресурсами, в первую очередь массивами государственной и муниципальной информации;
- о соблюдении прав интеллектуальной собственности и охране авторских прав в условиях открытых информационно-вычислительных сетей;
- о правилах формирования и ведения баз и банков данных, реестров информационных ресурсов;
- о сертификации информационных услуг и лицензировании информационной деятельности;
- о регламенте функционирования и доступности информационных ресурсов;
- о правах пользователей телекоммуникационных и информационных ресурсов.

Другой важной составляющей нормативно-правовой базы является пакет нормативно-технических документов. Он направлен на стандартизацию и унификацию процедур информационного обмена между различными информационными системами, внедрение концепции открытых систем, обеспечение совместимости открытых банков данных, организацию распределенной обработки информации на прикладном уровне, построение специализированных профилей и т. д. Этот пакет документов должен обеспечить функционирование региональной системы классификаторов экономической и статистической информации, включающих порядок использования общероссийских, межотраслевых и отраслевых классификаторов, их увязку и перекодирование, а также порядок создания и ведения

областных классификаторов, реестров, регистров и справочников.

3. Internet & Intranet

Начиная с середины прошлого года, с легкой руки фирмы Microsoft, в информационный обиход вошел новый термин Intranet. Intranet-технология — это идеология построения систем информационной поддержки внутри одной организации (не обязательно расположенной в одном здании, в том числе и для таких “больших организаций”, как, например, Сибирское отделение РАН или отдельный регион страны), основанная на Internet-технологиях.

Intranet-технология основывается на:

- организации технологической среды поддержки различных протоколов передачи данных (основной TCP/IP);
- использовании для представления и хранения информации WWW- и SQL-серверов;
- использовании технологии “клиенты — серверы” с минимальными техническими требованиями к клиентской части.
- поддержке распределенных БД (UNIX, OS/2, DOS, WIN32-платформы) с использованием СУБД типа Oracle, Informix, Delphi, MS SQL, а также поддержке очень популярных форматов СУБД, таких как dBase, FoxPro, MS Access и др.

Использование Internet/Intranet-технологии при построении информационных систем как общего назначения, так и специализированных в ближайшее время станет доминирующим в мировом информационном пространстве по следующим простым причинам:

- достаточно простая для пользователя система поиска нужной информации;
- минимальные требования как с технической стороны, так и со стороны программного обеспечения к рабочему месту клиента (клиент работает со стандартным программным обеспечением и единственным требованием является поддержка работы стандартного WWW-просмотрщика — браузера);
- поддержка распределенной системы хранения информации и множественных методов ее хранения;
- работа с практически неограниченным объемом разноплановых данных (текст, графика, изображение, звук, видео, векторные карты и др.);
- технологически простой способ администрирования информационных систем с одного рабочего места;
- поддержка удаленных методов редактирования и пополнения информации.

Основной принцип построения информационных систем с использованием Intranet-идеологии заключается в организации системы доступа к информации через WWW-сервис Internet.

Основным и единственным ограничением на рабочее место клиента является возможность подключения последнего к Internet. Подключить к Internet можно практически любой компьютер, но от мощности машины и от скорости связи зависит, какие услуги сети будут доступны и, что самое главное, качество предоставляемых услуг.

Существующая технология Internet позволяет получить практически все услуги (за исключением доступа в реальном времени) через электронную почту. А использовать электронную почту можно даже не имея компьютера: есть возможность работать на терминалах — принимать и отправлять текстовые сообщения. Электронная почта является начальным, базовым сервисом Интернет. Через электронную почту в режиме отложенного запроса можно получить любую информацию, а также передать любой компьютерный

файл. К письмам можно прилагать фотографии, аудио- и видеофайлы и др. Для работы с электронной почтой достаточно компьютера с 286-м процессором, работающего под MS DOS (хотя здесь могут возникнуть трудности с кодировками русского языка), однако лучше использовать MS Windows 3.1 (здесь нужен компьютер с 386-м процессором, 4 Мб оперативной памяти). Подключение к сети для электронной почты может обеспечить модем со скоростью не ниже 2400 bps.

Используя MS Windows 3.1, вы уже можете организовать и доступ к информационным ресурсам в реальном времени. Хотя это и не очень удобно в связи с тем, что большинство современных клиентских программ для работы с сервисами Internet разработаны для операционных систем MS Windows 95/NT, MacOS и UNIX (следует отметить, что для операционных систем фирмы Microsoft они являются свободно распространяемыми). Для старой системы Windows 3.1 хотя и существуют программы для Internet, но так как в самой операционной системе нет встроенной поддержки Internet (IP) и нет истинной многозадачности, программы работают хуже. В последнее время все меньше современных программ поддерживают устаревшую операционную систему. Для установки MS Windows 95 вам достаточно скромной 386 машины с 8 Мб оперативной памяти. Доступ к сети в режиме реального времени может обеспечить модем со скоростью не ниже 14 400 bps.

Однако увидеть во всей красе изящные страницы с фотографиями и видео-фрагментами можно лишь на современном компьютере (процессор не ниже 486 с оперативной памятью 16 Мб) с хорошей видеокартой и монитором.

В заключение несколько слов о требованиях к серверам (WWW-серверам и SQL-серверам баз данных). Минимальные требования к компьютеру, на котором можно организовать сервер: для OS UNIX это компьютер с 486 процессором и 16 Мб оперативной памяти (лучше P5-100 с 32 Мб), для MS Windows NT это компьютер с процессором P5-133 и 64 Мб оперативной памяти (оптимально PentiumPro-200 с 128 Мб) для WWW-сервера и 2–4-процессорная машина для сервера баз данных. Отметим, что Intranet-технология не требует установки мощного центрального компьютера для работы информационных систем. Данная технология является распределенной и позволяет размещать информационные ресурсы на различных компьютерах (WWW-сервер на одной машине, серверы баз данных на других), выбирая для каждой информационной задачи оптимальную конфигурацию вычислительной техники.

Наиболее существенной проблемой, возникающей при организации рабочего места клиента, является проблема множественности кодировок русского языка. Несмотря на то, что существует международный стандарт на кодировку русского языка, принятый международной организацией по стандартизации (ISO), — кодировка ISO-8859-5, в нашей стране он практически не используется. Исключение составляют большие компьютеры фирм Sun и DEC, в которых фирмами-производителями предусмотрена национальная локализация программного обеспечения для русского языка. Кроме кодировки ISO-8859-5, в настоящее время достаточно широко применяются четыре другие кодировки символов кириллицы (KOI8-R, CP866, CP1251, MACOS). Поэтому, если в информационной системе применяются различные кодировки русского языка, а это неизбежно при создании распределенных информационных систем с использованием рабочих станций, работающих под разными платформами, то необходима тщательная настройка рабочего места клиента на понимание различных кодировок.

Основной принцип, заложенный в Intranet-технологии построения информационных систем, заключается в разделении вычислительных ресурсов как между серверами, так и между сервером и клиентом. Реализация этого принципа основывается на использо-

вании HTTP-SQL интерфейса для формирования запросов пользователя на получение информации. Использование этого интерфейса обеспечивается установкой в сети HTTP-SQL-сервера, который переводит HTTP-запросы клиента к WWW-серверу в SQL-запросы к различным базам данных. Рассматриваемая технология позволяет удачно сочетать возможности гипертекстового оформления информации и использование современных СУБД. Причем со стороны клиента вы имеете возможность полностью унифицировать запросы как на поиск и представление информации, так и на получение аналитических справок из информационных систем.

С другой стороны, данная технология позволяет использовать в сетевом режиме уже имеющиеся базы данных без дополнительных затрат на их унификацию и приведение к единому стандарту. Основные затраты здесь будут направлены только на соответствующие описания баз данных и запросов для HTTP-SQL-сервера. Причем базы данных могут находиться на различных машинах, расположенных на произвольном расстоянии друг от друга. Использование данной технологии позволяет решать весь спектр задач, присущих интегрированной информационной системе, включая удаленный ввод и редактирование данных.

Соответствующее математическое обеспечение для организации HTTP-SQL-сервера является свободно распространяемым как для MS Windows, так и для UNIX-платформ, а СУБД можно использовать как уже имеющиеся (например, MS Access для MS Windows или mSQL для UNIX), так и сетевые (Informix, Oracle, MS SQL и др.), которые нужно приобретать отдельно. Для несетевых СУБД необходимо устанавливать HTTP-SQL-сервер на той же машине, на которой работает используемая вами СУБД.

В качестве примера использования данной технологии и организации HTTP-SQL-интерфейса укажем информационную систему Сибирского отделения РАН, установленную на WWW-сервере Отделения <http://www-sbras.nsc.ru> и WWW-сервер Института вычислительных технологий <http://www.ict.nsc.ru>, страницы которого полностью формируются из запросов к базе данных.

4. Электронные журналы

Поддержка электронных версий научных журналов, издаваемых Отделением, невозможна без использования реляционных баз данных, в которых легко организовать на WWW-сервере поиск нужной статьи или информации об авторах.

В настоящий момент подготовлена к работе информационная система для ведения полнотекстовых баз данных публикаций, электронных журналов и данных по персонам. Система работает с MS SQL-сервером в среде MS Windows и mSQL-сервером в среде UNIX. Ведение баз данных осуществляется через WWW-сервер с использованием просмотрщиков WWW-страниц Netscape Navigator 3.0 (или выше) и MS Internet Explorer 3.2 (или выше).

В качестве примера использования информационной системы рассмотрим электронную версию журнала "Вычислительные технологии". Относительно каждой статьи в базе данных хранится следующая информация (на русском и английском языках):

- название статьи;
- авторы;
- аннотация (содержащая математические формулы либо в Т_ЕХовской нотации, либо

графический файл, либо в системе WebEQ¹);

- ключевые слова;
- коды классификаторов (ВИНИТИ, РФФИ, AMS);
- статья в виде Postscript file, Compressed Postscript file, L^AT_EX file (на языке оригинала).

Кроме базы данных статей ведется база данных авторов следующего содержания:

- анкетные данные (фамилия, имя, отчество, год рождения; фотография, если есть);
- образование, специальность;
- ученая степень, ученое звание;
- должность, место работы (почтовый адрес, подразделение);
- телефоны, факс, E-mail, URL (ссылка на личную страницу);
- научные интересы (ключевые слова, коды классификаторов);
- краткая справка о научно-технической деятельности (текстовый файл);
- участие в научных советах, редколлегиях и пр.;
- дополнительная информация (текстовый файл).

От представления статьи в виде DVI-файла, как это дополнительно делается в зарубежных электронных журналах, мы сознательно отказались, чтобы не создавать читателям трудностей, связанных с множественностью кодировок русского языка. Текст статьи в виде L^AT_EX-файла может быть получен читателем вместе с T_EXовскими макросами форматирования и для статей на русском языке в любой из выбранных читателем кодировке (DOS CP866, MS Windows CP1251, KOI8-R, ISO-8859-5). По умолчанию читателю будет передан файл в той кодировке, в которой он обратился к WWW-серверу. Сервер работает в режиме динамической перекодировки текстов через проху-сервер² и в настоящий момент поддерживает кодировки русского языка MS Windows CP1251, KOI8-R и ISO-8859-5. В системе организован поиск авторов и статей по ключевым словам, кодам классификаторов, по датам и по образцам текстов и различных полей.

Поступила в редакцию 24 апреля 1997 г.

¹См. статью О. В. Олейника, Е. М. Толкачевой, А. М. Федотова "Электронные издания и представление математических текстов на WWW" (с. 60).

²Proху-сервер описан в статье А. В. Дорофеева и А. М. Федотова "Электронные публикации в среде Internet и множественность кодировок русского языка" (с. 31).