

СОЗДАНИЕ СЕТЕВЫХ ПРИКЛАДНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ
НА ОСНОВЕ СОЧЕТАНИЯ РАЗЛИЧНЫХ ТЕХНОЛОГИЙ

Е.Ю.Зыбарев, Ю.М.Зыбарев, Е.Н.Фаддеенков, А.А.Шейна

В в е д е н и е

За последние несколько лет компьютерные сети в России стали реальностью. Снятие ограничений на поставки телекоммуникационного оборудования и вычислительной техники из-за рубежа позволило активно включиться российским специалистам в создание единого информационного пространства. Однако следует отметить, что создание телекоммуникационной среды (каналов, компьютеров и т.д.), собственно компьютерных сетей - это создание инструмента для формирования единого информационного пространства, тогда как конечная цель - это информационное (содержательное) наполнение создаваемых компьютерных сетей собственными информационными ресурсами с возможностями полноценного доступа (в режимах on-line и off-line) из российских и зарубежных сетей (Internet и т.д.) для широкого круга пользователей.

Обеспечить высокий уровень и возможность промышленной эксплуатации прикладных информационных систем и информационных ресурсов, развивать полноценные механизмы предоставления информационных услуг на их основе в компьютерных сетях возможно только опираясь на реализацию современных сетевых информационных технологий, идеологию открытых систем и сети телекоммуни -

каций. Учитывая необходимость и актуальность интеграции России в мировое информационное пространство, рассматриваемую проблему, на наш взгляд, трудно разрешить без ориентации на стандартные, отработанные в мировой практике решения и технологии.

Особо важное значение данный тезис имеет для научно-образовательной среды, специфика которой определяется в России работой в очень жестких финансовых и соответственно технических условиях. Эти условия, в свою очередь, диктуют необходимость проработки такого набора решений, которые делают возможным поэтапное создание и развитие информационных ресурсов и предоставление информационных услуг, позволяющие без коренной ломки наращивать их номенклатуру.

Цель данной статьи - рассмотреть на примере информационного узла компьютерной сети Новосибирского государственного университета ряд современных технологий и проектных решений, позволяющих обеспечить доступ к структурированной и гипертекстовой информации, а также к другим видам информационных ресурсов в локальных и глобальных (Internet) сетях. Нами предлагается некоторый вариант формальной модели работы с информационными ресурсами, которые создаются в виде комплекса взаимосвязанных информационных серверов различного типа (FTP, News, Gopher, WWW, SQL) и образуют информационный узел сети на основе идеологии "клиент-сервер" в сетях архитектуры Internet. Более подробно в данной статье будут изложены вопросы реализации прикладных информационных систем на основе двух типов информационных серверов, SQL и WWW, реализующих идеологию "клиент-сервер".

1. Модель информационного узла сети семейства Internet

Основные виды информационных ресурсов в сетях Internet создаются в виде информационных серверов с соответствующими технологиями организации обработки данных и доступа к ним:

- News-сервер (сервер телеконференций) обеспечивает использование крупнейшей в мире системы телеконференций USENET/NEWS, а также возможность организации собственных конференций (внутривузовских, региональных);

- FTP-сервер (файловый сервер) - это возможность использовать емкое хранилище файлов, разделяемых по доступу пользователями в сети. С использованием протокола FTP файловые серверы ВУЗа могут быть доступны любым абонентам сети Internet, что обеспечивает возможность широковещательного распространения необходимой информации и данных;

- Gopher-сервер обеспечивает более комфортный доступ к файловым архивам в Internet за счет организации системы иерархических меню, содержащих в себе описание нижележащих разделов. Различные Gopher-серверы в сети могут быть связаны ссылками и образуют распределенную информационную систему. Над сетью Gopher-серверов построена система поиска информации Veronica. Сеть Gopher-серверов является одной из наиболее широко распространенных в университетской среде на западе;

- WWW (World Wide Web) - одна из наиболее интенсивно развивающихся служб в Internet, успех которой заложен в идее объединить в единой оболочке возможность доступа ко всем наиболее распространенным сервисам Internet и добавить к этому элементы multimedia. Пользовательский интерфейс представляет собой средства, включающие возможность просмотра гипертекста, ссылки которого могут указывать на любой информационный ресурс как на данной машине, так и на любом другом сервере в сети. Язык описания HTML позволяет включать в состав документа изображения, звук, видео, экранные формы, гипертекстовые ссылки на файлы и исполняемые программы. Список доступных в Internet WWW-серверов очень широк и включает в себя информационные источники как образовательной, научной сферы, так и множества коммерческих рекламных и информационных компаний. Документы библиотек, раз-

личных фондов и научных программ доступны в наиболее удобном для обозрения виде;

- SQL-сервер обеспечивает создание и работу со структурированными базами данных на основе развитых СУБД (Oracle, Informix, Interbase и т.д.) с языком запросов SQL, ставшим стандартом де-факто.

Для описания модели информационного узла воспользуемся языком сетей Петри [1], который позволяет в удобном и наглядном виде отобразить структуру и механизмы функционирования объектов описания.

Основные понятия сети Петри в данной модели интерпретируются следующим образом:

- переход сети (прямоугольник) - программный (или программно-технический) комплекс, выполняющий заданные функции;
- позиция сети (окружность) - место (средство) хранения информационных ресурсов (данных); файловые системы и т.д.;
- маркер сети - собственно набор данных;
- дуги сети - направления потоков информации.

На рис.1 в виде варианта сети Петри представлена модель информационного узла в сети семейства Internet и механизма взаимодействия с ним клиентов. Концептуально, как это отмечалось выше, предлагаемая модель реализует идеологию "клиент-сервер". Исходными компонентами моделируемой информационной системы служат:

- <программа-клиент> - программный комплекс, установленный на рабочей станции пользователя в сети, реализующий функции пользовательского интерфейса и осуществляющий связь с сервером (когда это нужно). В предлагаемом варианте сети Петри они представлены переходами: <K1> - News-клиент, <K2> - Gopher-клиент, <K3> - FTP-клиент, <K4> - WWW-клиент, <K5> - SQL-клиент. В позиции (A_i) и (B_i), где $i = \overline{1,5}$, маркеры интерпретируются

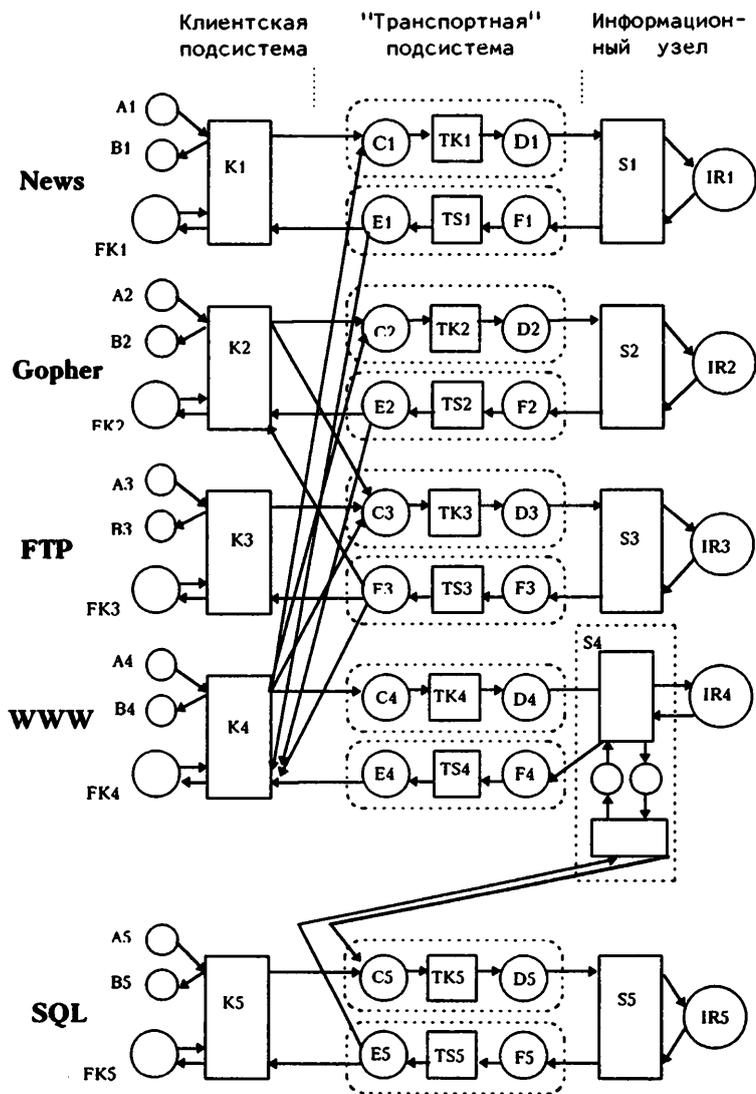


Рис.1. Модель функционирования прикладных информационных систем, создаваемых в идеологии "клиент—сервер" и работающих в сетях Internet.

ются как информационные сообщения, поступающие на вход и на выход <программы-клиента> соответственно;

- <программа-сервер> - установленный на компьютере в сети программный комплекс, основной функцией которого является обработка и выполнение запросов (команд), поступаемых от клиента, обеспечивая ему доступ к информационным ресурсам. В предлагаемом варианте сети Петри информационные серверы представлены переходами: <S1> - News-сервер, <S2> - Gopher-сервер, <S3> - FTP-сервер, <S4> - WWW-сервер, <S5> - SQL-сервер;

- <локальная файловая система> - файловая система компьютера, на котором установлена <программа-клиент>. В данной сети Петри представлена позициями (FK \bar{i}), где $i = \overline{1,5}$;

- <информационные ресурсы> - совокупность данных, организованных согласно соответствующим технологиям, характеризуются в данной сети Петри следующими позициями:

(IR1) - для News-сервера (файловая система);

(IR2) - для Gopher-сервера (файловая система);

(IR3) - для FTP-сервера (файловая система);

(IR4) - для WWW-сервера (файловая система);

(IR5) - для SQL-сервера (база данных);

- <транспортная компонента> - передача запроса/ответа по сети; характеризуется переходами: <TK \bar{i} > - передача информации от клиента к серверу, <TS \bar{i} > - передача информации от сервера к клиенту, где $i = \overline{1,5}$. В качестве сети выступает сеть передачи данных, функционирующая по протоколам TCP/IP, в соответствии с REC 1-15xx;

- информация, которой обмениваются между собой клиент и сервер, содержится в данной сети Петри в позициях (C \bar{i}), (D \bar{i}), (E \bar{i}), (F \bar{i}), где $i = \overline{1,5}$. Формат передаваемых сообщений определяется: для News-сервиса в соответствии с RFC-977 (NNTP), для Gopher-сервиса - по RFC-1436, для FTP-сервиса - по RFC-959. Для WWW-сервиса запрос передается в формате HTTP (Internet Draft

HTTP, Cern), а данные возвращаются в стандарте MIME (RFC-1437,1521-1523). SQL-клиент Oracle осуществляет запрос на языке SQL (ANSI), данные возвращаются в виде текста.

Набор дуг, входящих и выходящих из перехода <K4>, указы- вает на то, что большинство WWW-клиентов являются также и полнофункциональными FTP-, Gopher- и News-клиентами, обеспечивая доступ к различным типам информационных серверов. Так как WWW-клиенты не поддерживают доступ к SQL-серверам, то эта задача может быть возложена на WWW-сервер, что на данной сети Петри отражено дугами (<S4>, (C5)) и ((E5), <S4>). Более подробно эти механизмы будут рассмотрены ниже.

2. Элементы технологии создания гибридных информационных систем

После того, как нами описана принципиальная схема реализации идеологии "клиент-сервер" на примере базовых технологий и видов серверов сети Internet, перейдем к изложению технологических вопросов реализации прикладных информационных систем, создаваемых в рамках данной модели и использующих механизмы взаимодействия WWW- и SQL-серверов. В качестве SQL-сервера нами используется сервер на основе СУБД Oracle v.6.0.

SQL-сервер Oracle - это объединение базы данных, высокопроизводительной системы управления базами данных Oracle и собственно обработчика запросов. Такой сервер в объединении с клиентскими частями образует распределенную среду обработки информации, в которой клиенты, представляющие собой пользовательские интерфейсы, взаимодействуют с сервером, используя запросы на процедурном расширении языка SQL - PL/SQL.

Сервер принимает запросы на PL/SQL, обрабатывает их и возвращает результат (данные) клиенту, который представляет их

в удобном для пользователя виде. Компонента SQL*Net Oracle позволяет связывать клиентскую и серверную части, находящиеся на разных компьютерах, через локальные и глобальные сети передачи данных, используя большое количество протоколов: TCP/IP, IPX/SPX, Appletalk, Decnet и др. На рис.2 приведена функциональная схема взаимодействия SQL-клиента и SQL-сервера.

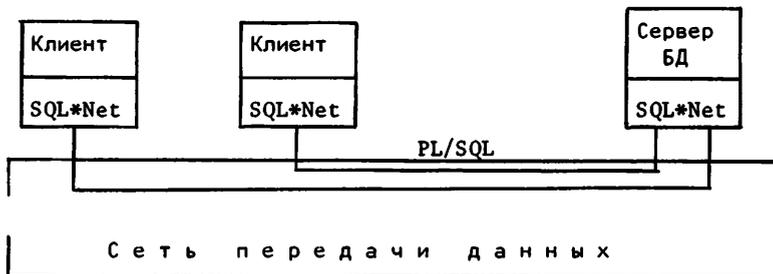


Рис. 2

Технология WWW также использует идеологию "клиент-сервер". Гипертекстовая информация на WWW-сервере хранится в виде отдельных файлов на диске в формате HTML - языке описания гипертекстовых документов, либо в форматах, специфичных для хранения информации (например, jpeg, gif - для графических изображений). По запросу клиента - пользовательского интерфейса - сервер выдает эти файлы, и они интерпретируются (изображаются, звучат и т.п.) клиентской частью. В качестве запрашиваемого файла может выступать стандартный вывод обычной программы, запускаемой на сервере. Такая возможность обеспечивается благодаря реализации Common Gateway Interface (CGI).

На используемом нами сервере установлено программное обеспечение Национального Центра по суперкомпьютерам США - NCSA httpd версии 1-3.

Стандартным языком, используемым при создании гипертекстовых документов для WWW, является HTML (HyperText Markup Lan-

guage). Конструкции этого языка вставляются непосредственно в текст документа, позволяя выделять в нем следующие элементы: имя документа; заголовки различных уровней; параграфы; разделительные линии; выделенные параграфы; предварительно "отформатированные" тексты; шрифты; различные списки (обычные, нумерованные, ...); встроенные (в текст) графические изображения (в форматах gif, xbm, ...); гипертекстовые ссылки по ключевому фрагменту текста или изображению и т.д.

Для обеспечения гипертекстовых ссылок и доступа к сетевым услугам внутри документа HTML использует метод, называемый URL (Uniform Resource Locator). Почти любой файл или сервис Internet можно интерпретировать в терминах URL, используя единую структуру адреса. Такая структура представляется в виде строки, содержащей: метод доступа, адрес сервера, дополнительные параметры (порт для соединения с сервером, путь к файлу и др.).

Приведем перечень основных методов доступа:

file - для локальных файлов и путешествия по файловой системе на локальной машине;

ftp - для доступа к FTP-серверам (файловый доступ);

gopher - для доступа к Gopher-серверам (представление текстовых данных в виде дерева);

http - для доступа к WWW-серверам (представление текстовых, графических и звуковых данных в виде гипертекста);

news - для доступа к серверам телеконференций.

Таким образом, гиперссылка может указывать не только на другой документ или multimedia объект, но и на какой-либо сетевой сервис. В Internet существует большая сеть WWW-серверов, многие из которых содержат ссылки друг на друга.

Последняя, расширенная версия языка HTML, называемая HTML+, поддерживает интерактивные экранные формы, позволяет спе-

цифицировать определенные "горячие" точки в изображениях, со - держит более богатый набор элементов форматирования и компоновки, формируемые таблицы и др.

Готовить тексты в формате HTML можно в текстовых редакторах, которые не имеют соответствующих элементов форматирования. Существуют также специальные инструменты для подготовки HTML-документов. Такими, например, являются HTML-Editor и HotMetal для MS-Windows. В них реализованы основные конструкции языка HTML, а в HotMetal - большинство конструкций HTML+.

В качестве источников для HTML-файлов могут быть использованы уже готовые документы, сформированные в каком-либо редакторе. В этом случае целесообразно применять инструменты, специально созданные для перевода текстов из заданного формата в формат HTML. Так, известны программы перевода из LaTeX и из Word в HTML. Например, с использованием пакета latex2html в зависимости от устанавливаемых в командной строке параметров можно получить заданное число уровней вложенности, графические файлы, переведенные в формат gif из Postscript, оглавление текста в соответствии с разбиением на файлы, навигационные кнопки для движения по тексту. При этом гипертекстовые ссылки расставляются программно.

Для ASCII-файлов также могут быть разработаны подобные программы перевода в HTML, однако к ним предъявляются определенные требования, связанные с форматированием и оформлением текста, которые уже формализованы в более мощных текстовых редакторах. Однако в простейшем случае текст может быть просто выведен на экран в исходном виде при помощи соответствующей команды языка HTML и не содержать шрифтов, заголовков и других конструкций.

Важной проблемой при формировании данных для WWW-сервера в России является проблема выбора кодировки кириллицы, так как различные программные платформы используют различные кодировки

русских букв. Так, большинство Unix-систем использует кодировку KOI-8, MS-DOS в текстовом режиме работает с альтернативной кодировкой, MS-Windows использует кодировку CP1251, широко распространены также стандарт ISO-8859-5, разработанный международной организацией по стандартизации (International Standard Organisation). Существует несколько вариантов решения этой проблемы. Одним из них является хранение нескольких копий информации в различных кодировках. Другим способом решения этой проблемы является динамическая перекодировка данных, которая осуществляется программой, запускаемой WWW-сервером через CGI в зависимости от параметра, переданного WWW-клиентом. В информационной системе о НГУ, на примере которой рассматривается предлагаемая технология, первый вариант используется для текстовых файлов в формате HTML, а второй - для информации, извлекаемой из базы данных.

Средства файловой системы не позволяют эффективно организовывать и обрабатывать большие объемы сложных структурированных массивов данных. С хранением и обработкой такого рода информации гораздо лучше справляются SQL-серверы баз данных, специально предназначенные для этой цели. Организация взаимодействия WWW- и SQL-серверов делает возможным получение доступа к информации, распределенной по различным базам данных, через дешевое клиентское место WWW, а также позволяет на массивах данных, имеющихся на SQL-сервере, создавать различные приложения для WWW, получая при этом возможность комбинирования информации, давая различные ее срезы и позволяя оперативно изменять форму выдачи информации и ее компоновку.

Благодаря функциональным характеристикам интерфейса CGI стала возможной технология создания информационных систем, при которой информация, хранящаяся на SQL-серверах, доступна для WWW-клиентов. На рис.3 приведена функциональная схема взаимодействия WWW-сервера и SQL-сервера.

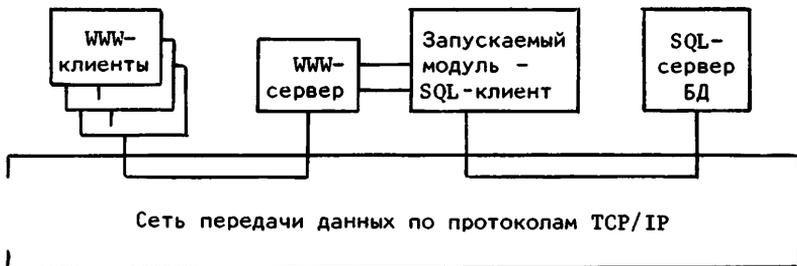


Рис. 3

В этом случае запрос на информацию от WWW-клиента передается серверу, который, в свою очередь, запускает программу, являющуюся SQL-клиентом. SQL-клиент связывается с сервером, получает всю необходимую информацию из базы данных и направляет ее на стандартный вывод. WWW-сервер транслирует эти данные WWW-клиенту.

Такая технология доступа позволяет работать с меняющимися - ся, хорошо защищенными данными, хранящимися в реляционной СУБД из широкого набора клиентских мест WWW, имеющих, в большинстве своем, статус free- или shareware. Кроме этого, корректно соблюдается лицензионное ограничение числа пользователей Oracle, предоставляя в то же время доступ к данным для широкой аудитории.

SQL-приложение для WWW-сервера представляет собой запускаемый модуль, который, используя параметры, переданные ему WWW-сервером, формирует запрос к SQL-серверу, преобразовывает результат выполнения этого запроса в требуемый формат (например, HTML) и отправляет его на стандартный вывод.

В качестве инструментария при создании SQL-приложений для WWW использовались программные продукты фирмы Oracle: SQL*Plus, SQL*ReportWriter, а также программы, написанные на языке C с

использованием Oracle Call Interface и прекомпилятора Pro*C. Из них самым удобным средством для представления данных Oracle в формате HTML является SQL*ReportWriter как средство, специально разработанное для подготовки отчетов, сочетающее средства форматирования текстов с возможностями языка SQL. Наибольшую эффективность этот продукт имеет при формировании отчетов по нескольким сложным образом взаимосвязанным запросам, например, отношения типа "мастер-деталь" ("один-ко-многим"). SQL*Plus уступает SQL*ReportWriter в средствах форматирования текстов, однако он более прост в освоении, требует меньше ресурсов и быстро исполняется, позволяя в то же время использовать все возможности языка SQL и его процедурного расширения PL/SQL. Написание SQL-приложений на языке C позволяет добиться большей гибкости и производительности, однако является достаточно трудоемким и требует определенной квалификации.

3. Пример результата реализации описанных технологий

Фрагмент информационного узла сети Новосибирского государственного университета, содержащий информационно-справочную систему о НГУ, которая реализована на основе описанных выше технологий, представлена на рис.4. Часть информации в этой прикладной информационной системе создается и поддерживается в виде комплекса баз данных на основе СУБД Oracle, другая часть - средствами WWW-технологии (HTML-файлы).

На рис.4 часть информации, расположенная на SQL-сервере, выделена курсивом и между объектами (на уровне их экранного представления) установлены связи. К части информации, которая реализована в виде файлов в формате HTML, также имеется доступ, хотя на схеме не отображены все связи в силу их меньшей информативности; как правило, эта информация в данной системе носит общий обзорный характер.

Необходимо отметить, что в целом информационный узел сети НГУ реализует полную описанную в данной статье (в частности, изображенную на рис.1) модель, включая доступ и к другим видам серверов. Остановимся коротко на аннотации основных информационных ресурсов на вышеупомянутых серверах. На текущий момент, кроме представленной на рис.4 информационной системы, функционирующей на основе взаимодействия WWW- и SQL-серверов, на различных вида серверах содержится следующая информация:

WWW: * На WWW-сервере расположены:

- сведения о структуре и деятельности Областного ЦНИТ при НГУ,
- справочник по телекоммуникационным сетям и информационным ресурсам,
- информация о зарубежных фондах,
- FAQ (сервер часто задаваемых вопросов),
- ссылки на другие WWW-серверы;

FTP: * FTP-архив содержит свободно распространяемое программное обеспечение и документацию, связанные с различными областями применения компьютерных технологий. При этом осуществляется регулярное "зеркалирование" информации с некоторых других FTP-серверов;

Gopher: * Gopher-сервер НГУ содержит положения о деятельности системы региональных ЦНИТов Госкома ВУЗа РФ и их структуре, тематических планах и программах работ; базу данных о мировой системе Gopher- и WWW-серверов в сети Internet и т.д.

News: * Сервер новостей телеконференций. Содержит около 5 тысяч различных тематических групп, причем постоянно происходит обновление информации. Статьи, написанные в телеконференцию, хранятся на сервере 7-9 дней.

Л и т е р а т у р а

1. ЗЫБАРЕВ Е.Ю., ЗЫБАРЕВ Ю.М. Сети Петри как языки спецификации дискретных систем //Настоящий сборник. - С. 95-118.

2. ГОРОДИЛОВ В.В., ЗЫБАРЕВ Ю.М. Создание и развитие университетских информационных ресурсов: проблемы и направления их решения //Новые информационные технологии в университетском образовании: Материалы международной науч.-метод. конф., Новосибирск, 1995. - С. 211-214.

Поступила в редакцию

27 марта 1995 года