

Организация информационного пространства Томского научного центра СО РАН на базе Интернет - технологии

Ахлестин А.Ю., Бабилов Ю.Л., Крутиков В.А.*, Фазлиев А.З.

Институт оптики атмосферы СО РАН, Томск

*Институт оптического мониторинга СО РАН, Томск

E-mail: ylb@iao.ru, faz@iao.ru

1 ВВЕДЕНИЕ

Начиная со второй половины 1998 года в Институте оптики атмосферы СО РАН (ИОА СО РАН) ведутся работы по созданию информационных систем на основе Интернет-технологий для представления научной и деловой информации [1, 2]. В настоящее время разработаны презентационный сайт (<http://www.iao.ru>), сайт конференций, проводимых Институтом (<http://symp.iao.ru>), сайт журнала "Оптика атмосферы и океана" (<http://ao.iao.ru>), инструментальные средства актуализации информации на сайтах, а также интранет Института, в котором собрана справочная и деловая информация для внутреннего пользования.

Результаты этой работы были представлены на заседании Президиума Томского научного центра (ТНЦ СО РАН), где было предложено разработать и внедрить подобные системы других институтах ТНЦ СО РАН, а также создать корпоративную информационную систему для научного центра (корпоративный Интранет). Проект по организации информационного пространства ТНЦ СО РАН получил поддержку РФФИ (грант 00-07-90279).

Актуальность проекта обусловлена состоянием дел в информатизации научных учреждений России. Оно остается неудовлетворительным, что объясняется, прежде всего, экономическими факторами и практически отсутствующим коммерческим программным обеспечением для нужд научных учреждений. Имеющееся бесплатное программное обеспечение бессистемно и разнородно.

Структура информационных систем, применяемых в проекте, является типовой для решения задач такого типа. Основу информационной системы составляют данные, концентрируемые в базах данных или файловой системе (нечетко структурированные данные). Второй составной частью информационной системы является инструментарий для отображения данных. Эта

часть реализована средствами Интернет-технологий. Выбор Интернет-технологий обусловлен возможностью предоставления пользователям единого интерфейса как для выхода в Интернет, так и для доступа к локальным информационным ресурсам. Немаловажным фактором, повлиявшим на этот выбор явилось и то, что реализация клиент-серверных приложений на основе Интернет-технологий существенно проще, чем при использовании других подходов.

2 КЛАССИФИКАЦИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ

Анализ информационного общения в институтах ТНЦ СО РАН показал, что информационные ресурсы, используемые в академическом учреждении, можно разделить по их назначению на научные, организационно-научные, административно-управленческие ресурсы. Такое деление условно, но оно является указателем для пользователя и администратора информационных ресурсов.

К научным ресурсам относятся электронные публикации в предметных областях, персональные каталоги научных публикаций, справочники по фундаментальным константам, предметные информационные системы или оцифрованные экспериментальные данные.

Научно-организационные ресурсы включают в себя информацию о проводимых конференциях, конкурсах проектов, выставках, ежегодные планы и отчеты о деятельности Института и т.д.

Уставы, положения, приказы, нормативные акты, информация о сотрудниках, текущем финансовом положении подразделений и т.д. являются административным информационным ресурсом.

Все прочие информационные ресурсы, например, иностранные словари, руководства пользователя, каталоги ссылок на Интернет-ресурсы, почтовый клиент, чат, свободно распространяемое программное обеспечение, предоставляемое пользователю, относятся к ресурсам общего назначения.

По статусу использования информацию можно разделить на ресурсы текущего пользования и архивные. Такое разделение позволяет, с одной стороны избавиться при отображении от устаревших ресурсов, а с другой, обеспечить пользователю доступ ко всему временному ря-

©Вторая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
26-28 сентября 2000г., Протвино

ду электронных документов.

По уровню доступа ресурсы можно разделить на ресурсы общего доступа, ресурсы научного центра (научные, ведомственные, правовые) и внутриинститутские ресурсы.

Информационные ресурсы общего доступа структурированы следующим образом:

- презентационные материалы;
- информация о научных мероприятиях (конференциях, симпозиумах, рабочих группах), проводимых Институтами ТНЦ СО РАН, в том числе тезисы докладов и труды конференций;
- аннотации и тексты статей научных журналов, издаваемых в ТНЦ СО РАН.

На уровне межинститутского общения можно выделить:

- кооперативные библиотеки избранных электронных публикаций;
- ведомственный документооборот;
- правовую справочную систему;
- доски объявлений и информационные порталы;

Внутриинститутские информационные ресурсы включают в себя:

- внутриинститутский документооборот;
- доски объявлений;
- архив электронной информации;
- каталог научной библиотеки;
- индивидуальную библиографическую информацию научного сотрудника.

3 ОБЩАЯ СХЕМА ОРГАНИЗАЦИИ ИНФОРМАЦИОННОГО ПРОСТРАНСТВА ТНЦ

В соответствии с приведенным выше классификатором информационных ресурсов по уровню доступа информационное пространство ТНЦ разделяется на три крупных блока:

1. Система электронных публикаций в Интернете: материалы конференций, научная литература, презентационные материалы.
2. Интранет научного учреждения (внутренний документооборот, справочная, ссылочная и вспомогательная информация, электронные архивы, библиотечный каталог, библиография). Информация доступна только сотрудникам данного учреждения.
3. Корпоративный интранет ТНЦ СО РАН (документы, справочная и научная информация, представляющая интерес для всех сотрудников ТНЦ). Информация доступна всем сотрудникам ТНЦ.

Каждый блок с той или иной степенью полноты содержит информационные ресурсы всех перечисленных выше типов и в программной реализации состоит из нескольких подсистем, включающих в себя

- базы данных;
- наборы запросов к базе данных;
- наборы скриптов для выполнения запросов и отображения информации в браузере;
- Web-интерфейсы для работы клиента и администратора.

Каждая подсистема реализуется либо как отдельный виртуальный сайт, либо как часть виртуального сайта. Кроме функциональной части сайт может включать типовые компоненты, такие как контекстный поиск по сайту, статистика посещений, раздел новостей, псевдобаннерную систему.

В настоящее время система электронных публикаций в Интернете состоит из

- подсистемы поддержки проведения научных конференций, служащей для информационного обеспечения организации и проведения научных конференций;
- подсистемы поддержки отображения научных публикаций, обеспечивающей проведение полного цикла прохождения издания от поступления в редакцию до выхода в свет;
- презентационных сайтов институтов.

Административное управление сайтами системы электронных публикаций осуществляется через интерфейсы, интегрированные в интранет.

Интранет предназначен для информационной поддержки деятельности Томского научного центра. Он состоит из следующих функциональных компонент:

- тематические доски объявлений (уставы, положения, приказы, ведомственные нормативные акты, отчеты, планы, объявления о конкурсах, грантах, выставках и т.д.);
- справочная информация (структура, сотрудники, телефоны и адреса электронной почты, интернет-ссылки, погода и т.п.);
- средства оперативного обмена информацией (электронная конференция, чат, клиент электронной почты);
- библиотечная и библиографическая информация (каталоги научных библиотек институтов, библиография изданных научных трудов институтов, кооперативная библиотека электронных публикаций);
- юридическая информация (приобретается справочная правовая система, имеющая WWW-интерфейс);
- подсистема поддержки научно-организационной деятельности (научные планы и отчеты);

- поддержка документооборота;
- архив фото- видео- и других мультимедийных файлов;
- архив свободно распространяемого программного обеспечения;
- аксессуаров общего назначения (калькулятор, карта-тека, ежедневник и т.п.).

Структурно Интранет ТНЦ представляет собой набор WWW-сайтов с авторизованным доступом (интранет институтов и корпоративный сайт ТНЦ), позволяющих отображать и изменять информацию из набора баз данных идентичной структуры пространственно распределенной по серверам разных научных подразделений.

4 СИСТЕМА ЭЛЕКТРОННЫХ ПУБЛИКАЦИЙ В ИНТЕРНЕТЕ

За годы своего существования ТНЦ СО РАН накопил значительное количество информационных ресурсов. Так, издано около двухсот монографий, опубликовано несколько сотен препринтов, авторефератов диссертаций, создано несколько десятков баз данных по предметным областям знаний. В год в научном центре проходит около 10 научных мероприятий, издаются два научных журнала ("Оптика атмосферы и океана" и "Физическая мезомеханика").

Система электронных публикаций в Интернете предназначена для решения задачи адекватного отображения этой информации с использованием современных технологий. Она состоит из трех частей: подсистемы поддержки проведения научных конференций, подсистемы поддержки отображения научных публикаций и презентационного сайта.

Первые две подсистемы базируются на аналогичных подсистемах, разработанных для Института оптики атмосферы, которые описаны в работах [1, 2, 3] и далее не рассматриваются. Следует отметить, что при разработке этих подсистем использован опыт реализации подобных систем Институтом вычислительных технологий СО РАН [5, 6].

Разнообразие подходов, используемых при создании презентационных сайтов научных учреждений России, является достаточно обширным. Мы не ставили задачи типизации научных сайтов в масштабах России и ограничились анализом подходов использованных в ТНЦ СО РАН. В настоящее время все институты ТНЦ СО РАН имеют свои презентационные сайты, построенные по традиционной технологии, то есть в виде набора статических гипертекстовых страниц.

Возможно, такое построение сайтов является оптимальным для слабоструктурированной информации, представляемой на сайтах подобного типа. Однако, с другой стороны, как показывает анализ, все презентационные сайты институтов РАН содержат примерно одни и те же разделы, такие как

- основные направления исследований;
- научно-организационная и кадровая структура;

- участие в программах и проектах;
- основные научные достижения;
- материальные и информационные ресурсы;
- основные публикации;
- международные связи.

Значительная часть этой информации изменяется во времени, поэтому актуализация информации на сайтах требует дополнительных затрат и на практике проводится нерегулярно. В то же время информация, отображаемая презентационным сайтом, фактически является производной от информации, содержащейся в базах данных системы Интранет, поэтому удобным является построение презентационного сайта Института как системы отображения информации, представленной в базах данных Интранета. Актуализация баз данных в Интранете обусловлена технологическими процессами внутриинститутской деятельности. Отображение информации можно осуществить несколькими способами:

1. путем генерации статических страниц презентационного сайта, относящихся к определенному разделу, по запросу администратора при изменении информации в исходной базе данных;
2. путем создания отдельной, небольшой по объему, базы данных презентационного сайта, ориентированной на структуру отображаемой информации, динамического отображения информации, содержащейся в ней, и поддержания ее в актуальном состоянии с помощью программного обеспечения Интранета, осуществляющего модификацию информации в БД презентационного сайта по запросу администратора;
3. путем динамического отображения информации непосредственно из баз данных Интранета.

Если добавить к этому инструментарий для подготовки неструктурированной информации, отображаемой сайтом, и библиотеку стилей оформления презентационного сайта, то процесс создания и актуализации информации презентационного сайта научного учреждения может быть существенно облегчен.

В рамках данного проекта планируется применение вышеизложенного подхода для создания презентационного сайта.

5 ИНТРАНЕТ

Интранет является средством интеграции разнородных информационных ресурсов, используемых как на уровне учреждения ТНЦ СО РАН, так и на уровне научного центра в единую информационную среду.

Интерфейс пользователя представляет собой информационный портал, т.е. набор окон, в каждом из которых отображается информация по определенному типу информационного ресурса. Пользователь может изменять набор окон, отображаемых при его входе в Интранет в соответствии с его потребностями. При активизации окна оно

разворачивается на всю рабочую область браузера, а затем может быть свернуто. Общий вид информационного портала приведён на рис. 1.

Доступ в Интранет авторизован. Авторизация доступа производится по обычной схеме, принятой в операционных системах, т.е. пользователи разбиты на группы, каждая из которых имеет разные права доступа к разным информационным ресурсам. В общем случае пользователь может принадлежать к любому количеству групп.

Данные, имеющие отношение к административно-управленческой деятельности, поставляются соответствующими службами института или научного центра. Ввод данных осуществляется через интерфейс пользователя путем заполнения соответствующих форм. Часть баз данных Интранета (доски объявлений, кооперативная библиотека электронных публикаций, архив мультимедийных документов, архив программного обеспечения) функционирует на основе самоорганизации, то есть ведётся сотрудниками, имеющими соответствующие права доступа, без вмешательства со стороны администрации.

5.1 Структура Интранета

Программное обеспечение для функционирования Интранета разделено на две части: ядро и приложения.

5.1.1 Ядро

Ядро - это промежуточное звено между пользователем и приложением Интранет. Ядро осуществляет авторизацию пользователей, запуск приложений и управление окнами, в которых отображается информация приложений.

Существует два режима отображения окон: отображение группы окон, выбранных пользователем - портал, и отображение окна, активизированного пользователем, в рабочей области браузера. При отображении группы окон они строятся в три колонки, две узких по краям и одну широкую в центре. Ядро умеет перемещать окна в портале по колонкам (по горизонтали), и внутри колонки (по вертикали), добавлять в портал или удалить из портала окно. Для добавления окна в информационный портал существует полный список всех доступных пользователю окон - "Главное меню".

Авторизация пользователя происходит по логин-имени и паролю. Логин-имя и пароль заводятся при регистрации пользователя в системе. При входе в Интранет пользователь вводит свои логин-имя и пароль, ядро ищет логин-имя в базе данных и если находит, сравнивает пароли. Если пароли совпали, то формируется уникальная строка символов - идентификатор сеанса, который заносится в базу данных и передаётся пользователю в URL. В дальнейшем, при любом действии пользователя, серверу методом GET передаётся этот идентификатор, по которому происходит аутентификация. Таким образом, пароль пользователя передаётся один раз за сеанс - это увеличивает безопасность системы. Пароли хранятся в зашифрованном виде.

Между группами пользователей и окнами устанавливается связь по правам доступа. Существует три уровня доступа: чтение, запись и администрирование. Пользователь, принадлежащий к группе, имеющей право на чтение, может только просматривать информацию, отобра-

жаемую в окне. Пользователь, имеющий право на запись, может также добавлять информацию в базу данных или ее часть, отображаемую в окне, а также редактировать или удалять информацию, введенную им. Пользователь с правами администратора может редактировать или удалять любую информацию из подмножества баз данных, отображаемых в окне.

База данных ядра содержит следующий набор объектов:

- Пользователи (Users);
- группы пользователей (Groups);
- окна (Windows);
- группы окон (Applications),

а также связи между ними:

- пользователь - группа пользователей, описывающая принадлежность пользователя к группам (Users_Groups);
- окно - группа пользователей (Windows_Groups), устанавливающая уровень прав доступа группы пользователей к окну;
- пользователь - окно (Views), описывающая структуру портала пользователя;
- группа пользователей - окно (Windows_Defaults), описывающая структуру портала по умолчанию для группы пользователей.

Структура базы данных ядра приведена на рис. 2.

5.1.2 Приложения

Логически приложение в Интранете представляет собой достаточно независимую функциональную компоненту системы для отображения информации.

Структурно приложение представляет собой (в зависимости от сложности) таблицу, набор взаимосвязанных таблиц БД, либо отдельную базу данных и набор программных средств, осуществляющих отображение информации в WWW-браузере и манипулирования ей.

С точки зрения пользователя приложение - это содержание отдельного окна Интранета.

Программная реализация приложения представляет собой класс языка PHP. Приложение запускается ядром путем вызова функции start(), которая обязательно должна присутствовать в классе. При запуске приложения ему передается информация о среде, в которой оно работает, в виде набора переменных.

Приложение в общем случае представляется в виде набора окон ядра с различными правами доступа, который пользователю представляется как одно окно Интранета. Рабочая область окна в терминах языка HTML представляет собой ячейку таблицы, что налагает довольно жесткие ограничения на использование приложением выразительных средств этого языка.

Схема взаимодействия пользователей и приложений с ядром системы приведена на рис. 3.

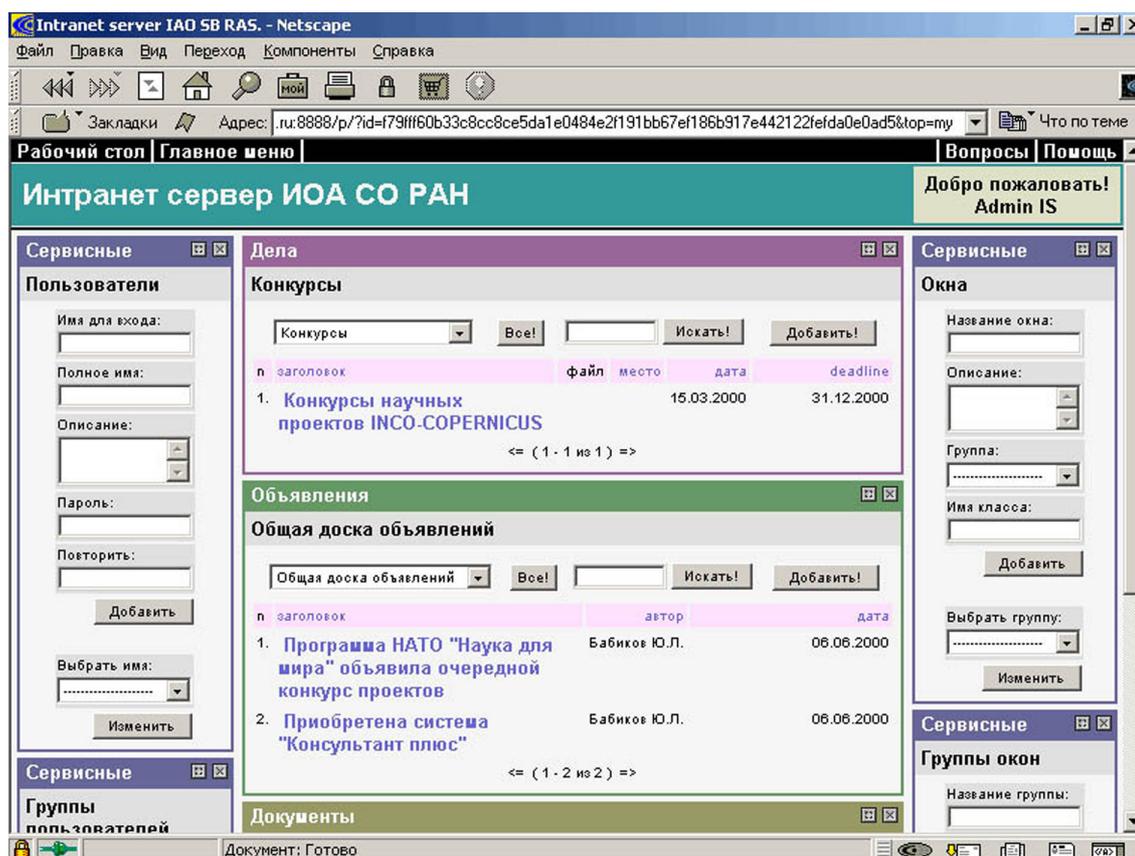


Рис. 1: Общий вид информационного портала Инtranета.

5.2 Приложения, базы данных и типовые классы

Проект находится в стадии реализации. На момент написания работы были созданы ядро Инtranета и набор приложений, работающих с базами данных простой структуры, таких как доски объявлений, архивы и коллекции документов различной структуры. Более сложные приложения находятся в стадии разработки и далее не рассматриваются.

Все информационные ресурсы, необходимые для функционирования рассматриваемых приложений, были структурированы и размещены в базе данных из нескольких таблиц различной структуры (рис. 4). Такой подход позволил упростить работу по созданию однородных по структуре новых информационных ресурсов.

Таблица "Announcements" содержит объявления, помещаемые на доски объявлений и информацию о них - тему, дату и время создания, время жизни и автора объявления. С данной таблицей работает класс board_nstr_news.

Таблицы "Files" и "Software" являются хранилищами соответственно файлов произвольного типа и файлов архива программного обеспечения. Физически файлы хранятся в файловой системе, а в таблицах хранятся ссылки на них и идентифицирующая информация: описание, MIME-тип, дата и время создания, количество связей с коллекциями. Выделение ссылок на файлы архива программного обеспечения в отдельную таблицу обусловле-

но тем, что наряду с доступом через Инtranет к нему обеспечивается доступ по локальной сети как к разделяемому ресурсу, в связи с чем хранилище файлов должно быть иерархической, а не плоской структурой, как для остальных файлов, что отражается в структуре таблицы.

Остальные таблицы представляют собой отображения различных коллекций, построенных из объектов, представленных таблицами "Files" и "Software". Таблица "Documents" представляет собой коллекцию документов (приказы, положения, инструкции, уставы, нормативные акты). Для работы с ней предназначены два класса: board_str_docs - для отдельных документов и boards_str_mdcs - для пакетов документов. Информация о событиях (конференции, выставки, конкурсы проектов) и пакетах документов, относящимся к ним, содержится в таблице "Events". С ней работает класс board_str_events. Таблица "Soft_Collection" и класс board_str_softs предназначены для работы с архивом программного обеспечения. Коллекция мультимедийных файлов управляется классом collection_mmedia путем манипулирования таблицей "MM_Collection". Кроме того, таблица "Custom_Collection" позволяет пользователю строить произвольные коллекции из документов любых типов, организованные в виде иерархической структуры произвольной глубины вложенности, подобной структуре современных файловых систем.

Все перечисленные таблицы имеют связи с таблицами

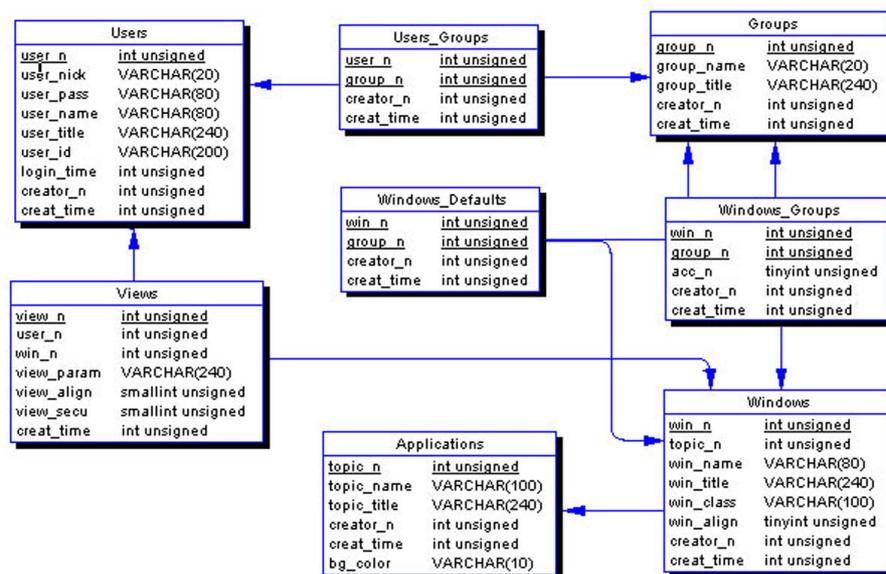


Рис. 2: Модель базы данных ядра Интранета.

ядра Windows, чтобы система знала, в каком псевдоокне Интранета отображать информацию, и Users - для идентификации пользователя, занесшего информацию в базу данных, а в личных коллекциях - ее владельца.

Администрирование Интранета осуществляется с помощью набора сервисных классов, работающих с таблицами ядра.

5.3 Корпоративный интранет

Корпоративный интранет ТНЦ содержит информационные ресурсы, доступные всем сотрудникам центра. Существуют несколько возможных подходов к его организации.

При первом подходе корпоративный интранет представляет собой автономную систему, структурно и функционально идентичную интранетам научных учреждений, со своей системой авторизации, отличающаяся лишь содержащейся в нем информацией. Такой подход несколько увеличивает объем работы администратора, но существенно упрощает его программную реализацию.

Другой подход заключается в том, что корпоративный интранет включает в себя набор баз данных общего пользования, а отображение информации осуществляется приложениями, работающими в интранетах научных учреждений.

При третьем подходе базы данных и приложения корпоративного интранета работают на отдельном сервере, а информация отображается локальным интранетом учреждения.

При двух последних подходах авторизация пользователей производится локальным интранетом, а корпоративный интранет авторизует запросы локального, то есть они отличаются по своей архитектуре. Возможна комбинация этих подходов.

В настоящее время в ТНЦ СО РАН корпоративный Интранет развивается в рамках первого подхода.

6 БАЗОВОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Для реализации Интранета используется только свободно распространяемое программное обеспечение. В качестве базовой платформы выбрана операционная система Linux RedHat, но допускается использование Free BSD.

Базы данных создаются с использованием СУБД MySQL, выбранной из-за ее высоких скоростных характеристик обработки запросов. Ее недостатки (некоторая ограниченность реализации языка SQL, отсутствие механизмов поддержки транзакций и внешних ключей) не являются принципиальными для работы с используемыми структурами данных.

В качестве веб-сервера был выбран Apache. Его структура позволяет создавать гибкие многопользовательские интерфейсы.

В качестве основного инструмента генерации динамических страниц HTML после ряда экспериментов выбран PHP, как наиболее адекватный для решаемых задач. Основной способ реализации - встроенный скрипт.

Обработка событий на стороне клиента осуществляется с использованием языка JavaScript.

Программное обеспечение для подготовки и работы с данными достаточно универсально и может быть адаптировано для использования в других научных учреждениях.

7 ЗАКЛЮЧЕНИЕ

Разрабатываемая система предназначена для информационной поддержки научной и управленческой деятельности академических учреждений Томского научного центра СО РАН, а также для представления о нем в Интернете. Созданное программное обеспечение будет унифицировано и документировано с целью свободного распространения среди научных учреждений.

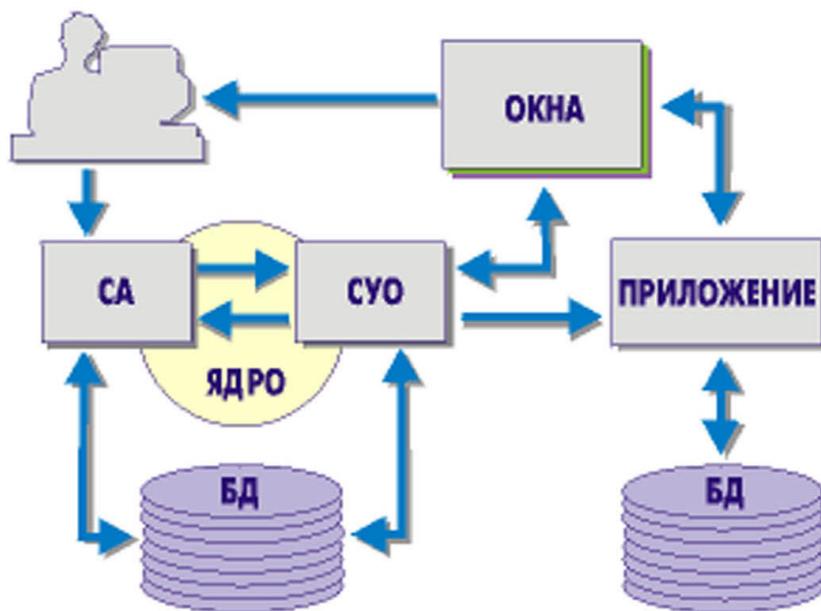


Рис. 3: Схема взаимодействия пользователей и приложений с ядром Интранета (СА-система авторизации доступа, СУО-система управления окнами).

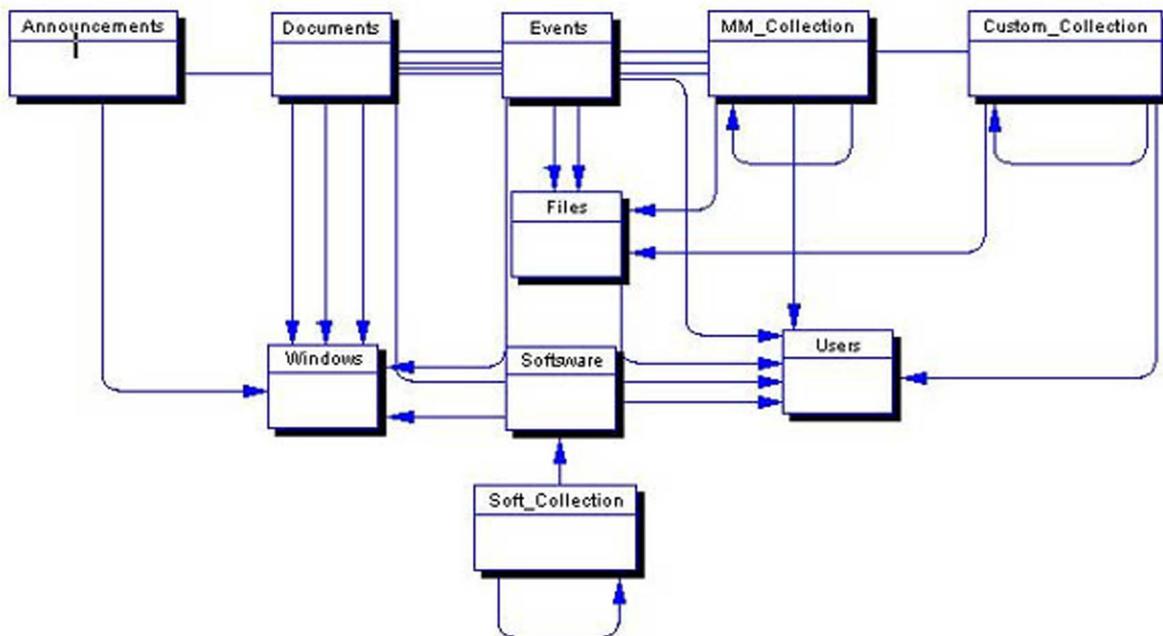


Рис. 4: Структура базы данных "Объявления, архивы, коллекции" Интранета.

Авторы благодарят РФФИ за финансовую поддержку проекта (грант 00-07-90279).

Список литературы

- [1] Ахлестин А.Ю., Бабилов Ю.Л., Карякин А.С., Фазлиев А.З. *Электронное представление информации в рамках академического института* Доклад на Четвертом рабочем совещании по электронным публикациям. Новосибирск, 21-23 апреля 1999 г.
- [2] Ахлестин А.Ю., Бабилов Ю.Л., Карякин А.С., Фазлиев А.З. *Информационные системы Института оптики атмосферы СО РАН: Структура и отображение электронной информации* Доклад на Первой Всероссийской научной конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции" Санкт-Петербург, 19-22 октября 1999г. с. 245-249.
- [3] Карякин А.С., Фазлиев А.З. *Электронное представление материалов конференций в Институте оптики атмосферы СО РАН* Доклад на V рабочей группе по электронным публикациям. Новосибирск, 21-23 июня 2000 г. Опубликовано в Интернете: <http://www-sbras.nsc.ru/ws/abstract.dhtml?1+30>
- [4] Ахлестин А.Ю., Бабилов Ю.Л., Фазлиев А.З. *Организация и отображение информации в Инtranете Института оптики атмосферы СО РАН* Доклад на V рабочей группе по электронным публикациям. Новосибирск, 21-23 июня 2000 г. Опубликовано в Интернете: <http://www-sbras.nsc.ru/ws/abstract.dhtml?1+21>
- [5] П.А.Богомяков, Ю.В.Пузака, А.М. Федотов. *Информационная система конференций*. Доклад на Четвертом рабочем совещании по электронным публикациям. Новосибирск, 21-23 апреля 1999 г.
- [6] П.А.Богомяков, А.М. Федотов, Ю.И.Шокин. *Электронные журналы по математике. На примере электронной версии журнала "Вычислительные Технологии"*. Доклад на конференции "Электронные издательские системы для науки и образования"(EPS4SE-98), 1998.