

Особенности создания корпоративных информационных систем на основе ОСУБД ODB-Jupiter 4.0

Андреев А.М., Березкин Д.В., Самарев Р.С.
НПЦ «ИНТЕЛТЕК ПЛЮС»

Тезисы доклада.

Введение

Внедрение компьютерных технологий в производство и управление предприятиями становится все более насущной задачей. Для хранения электронных документов и эффективной организации поиска предназначены информационно-поисковые системы (ИПС).

Такие системы разрабатываются в “НПЦ ИНТЕЛТЕК ПЛЮС” на основе собственной объектной СУБД ODB-Jupiter. При разработке ОСУБД ставилось несколько целей: обработка неструктурированной информации, хранение документов сложной структуры, встроенная поддержка полнотекстового поиска, возможность расширения набора типов в СУБД. Одна из ключевых особенностей ODB-Jupiter заключается в том, что в ядро ОСУБД встроен модуль полнотекстового поиска, что позволило очень эффективно реализовать индексацию и поиск текстовых данных, сохраняемых в базе.

В ОСУБД ODB-Jupiter встроены мощные средства разграничения доступа к объектам. Для каждого типа объекта и для каждого объекта определяется уровень секретности, группа и область. Каждый объект имеет своего владельца. Пользователи получают доступ к объекту в соответствии со своими правами.

ОСУБД ODB-Jupiter дает возможность создавать новые типы объектов для своих нужд и модифицировать существующие. Для создания новых типов используется специальная программа - *редактор схемы данных*, который позволяет создавать новые типы, а также определять между ними отношения наследования. Т.е. можно, создать тип некоторого общего документа, а затем создать нескольких более конкретных потомков.

Документ - основное понятие в ИПС, построенных на базе ODB-Jupiter. Документ - это информационная единица, состоящая из обязательного раздела - смысловой части (текста), набора реквизитов, которые упрощают поиск, организацию и классификацию и необязательного - ссылок на другие документы.

ИПС, построенные на базе ODB-Jupiter, обладают мощными поисковыми возможностями. Поиск ведется как по каждому виду документа в отдельности, так и по нескольким типам сразу.

Пользователю предоставляется возможность задавать слова не полностью, а лишь указывая основу слова, по которой должен производиться поиск. Для реквизитов, имеющих тип целое число, дата и время можно вводить запрос на точное соответствие, а также задавать отношения и диапазон, в который должно укладываться число.

Однако поисковые возможности ИПС ODB-Jupiter не исчерпываются только формальным поиском. Для более эффективного поиска пользователю предлагается несложный *язык запросов*. ИПС ODB-Text позволяет при формировании запроса разделять слова (или группы слов) пробелами или запятыми. Наличие пробела означает, что поиск будет по «И», запятой - что поиск идет по «ИЛИ». Для правильной интерпретации сложного запроса можно использовать круглые скобки. При поиске по тексту документа пользователю предоставляются дополнительные возможности создания запроса. Помимо всех возможностей фор-

мального запроса по реквизитам здесь можно задаваться *расстоянием между словами* в тексте.

Возможность формулировки запросов на естественном языке - одна из ключевых поисковых возможностей ODB-Jupiter.

Особенности реализации сервера ОСУБД

ОСУБД ODB-Jupiter имеет клиент-серверную архитектуру. Структура сервера приведена на рис 1. Особенностью ОСУБД ODB-Jupiter является то, что она ориентирована на построение информационно-поисковых систем, следствием чего является наличие встроенных средств поддержки Web-клиентов.

Сервер имеет 3 интерфейса, принимающих запросы по протоколу TCP/IP:

- интерфейс для обслуживания запросов от клиентов ОСУБД с полноценным доступом;
- интерфейс для обслуживания запросов от Web-клиентов (запросы от Web-сервера перенаправляются специально программой-переходником);
- интерфейс для подключения программы администрирования.

Основное назначение Web-канала – обеспечить работу Web-клиентов с ИПС, для чего предусмотрен набор стандартных операций и, соответственно, набор стандартных шаблонов, которые могут редактироваться администратором системы. Шаблоны хранятся в БД, поэтому для каждой БД возможен свой способ отображения данных. Канал для подключения Web-клиентов принимает запросы от Web-сервера, производит их разбор и обработку, формирует и высылает сформированную по соответствующему шаблону html-страницу. То, что сервер ИПС для Web-клиентов реализован вместе с сервером ОСУБД, позволяет наиболее оптимальным образом организовать его работу и обеспечить максимальную производительность.

Запросы от всех интерфейсов поступают на селектор БД. Поскольку сервер обеспечивает работу с несколькими БД, для каждой из них ведется схема данных, работает отдельный диспетчер безопасности и диспетчер транзакций. Общий для всех БД диспетчер блокировок обеспечивает возможность работы БД с пересекающимся набором хранилищ.

В функции диспетчера запросов входит журнализация всех операций, выполняемых сервером для выполнения запросов пользователей. Возможна журнализация действий пользователей на нескольких уровнях детализации:

- регистрация имени пользователя, IP-адреса, с которого поступило обращение;
- регистрация выполняемых типов операций (добавление, удаление, чтение, изменение, поиск и пр.);
- регистрация идентификатора объекта;
- применительно к ИПС – регистрация названия объекта.

Модуль доступа к хранилищам реализован таким образом, чтобы обеспечить полноценное функционирование сервера ОСУБД с любыми хранилищами, в т.ч. использующих для хранения данных другие СУБД. Это позволяет обеспечить высокий уровень надежности за счет применения журнализации и двухфазных транзакций не только в собственных хра-

нилищах, но и в том случае, когда применяемое внешнее хранилище (или внешняя БД) не имеют соответствующих средств.

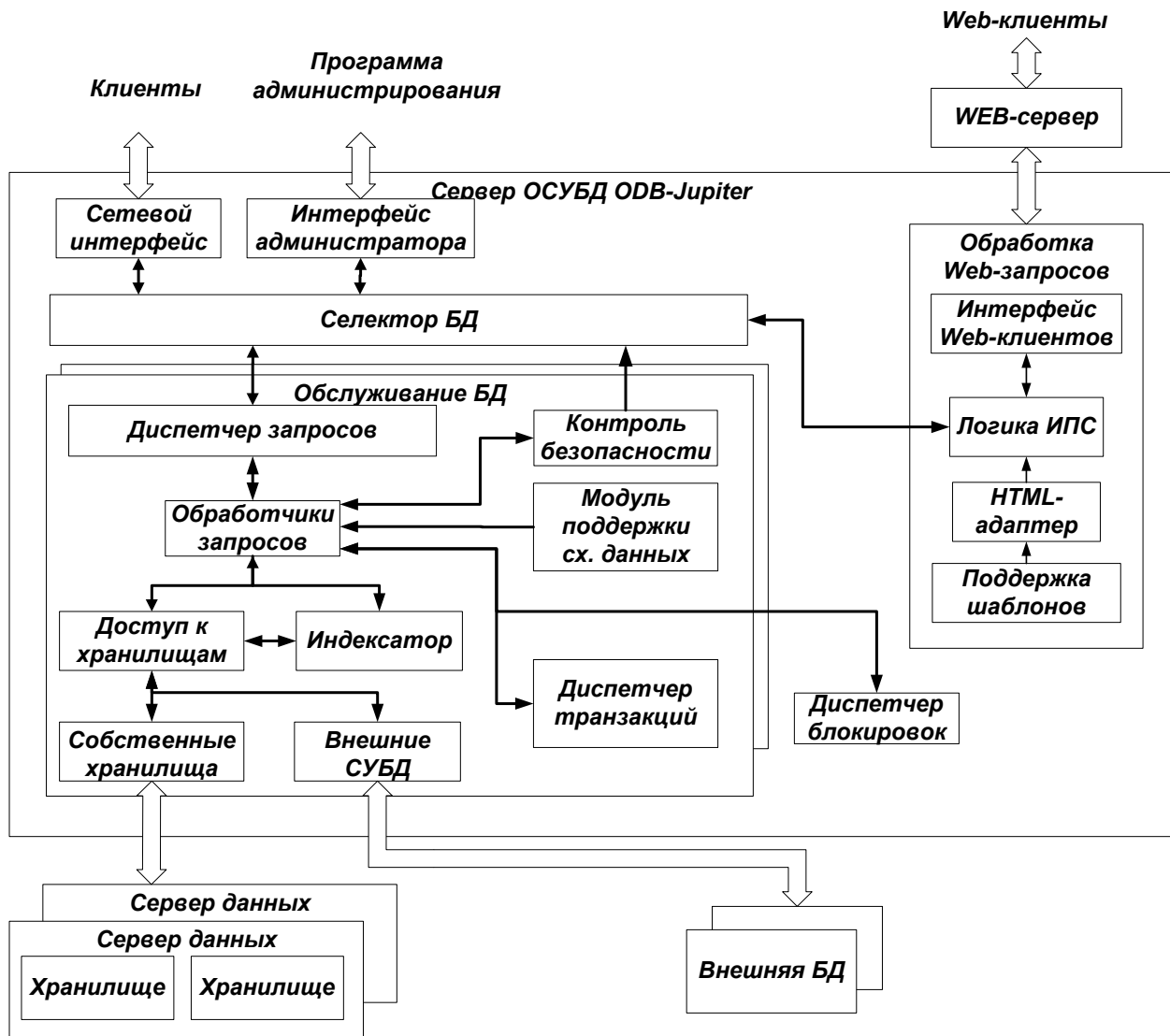


рис. 1 Структура сервера ОСУБД ODB-Jupiter.

Сервер обеспечивает возможность работы одной БД с несколькими хранилищами. Это позволяет для каждого типа объектов использовать своё хранилище, а также организовывать разные БД, имеющие пересекающийся набор объектов одного типа, например общее хранилище безопасности с объектами, определяющими права доступа пользователей, хранилище с метатипами и пр.

Хранилища сервера

Хранилища сервера ОСУБД работают в отдельных процессах, взаимодействие сервера ОСУБД с которыми производится через интерфейс CORBA, что позволяет их размещать в т.ч. на отдельных вычислительных системах (ВС). Это позволяет рассматривать процесс хранилища (или несколько таких процессов на одной ВС) как отдельный сервер хранилищ. Каждое хранилище работает независимо от других. При этом обеспечивается оптимальное использование доступной ОЗУ ВС, кэширование данных, не зависящее от операционной системы и полное отсутствие влияния возможных сбоев на работу других компонентов системы.

Хранилища ОСУБД ODB-Jupiter обеспечивают возможность одновременного использования нескольких разделов. Причем, в процессе эксплуатации возможно добавление новых разделов. Это позволяет, например, при необходимости увеличения объема хранилища, подключать дополнительные дисковые накопители в качестве нового раздела. Хранилище производит оптимальное заполнение разделов, обеспечивает журнализацию изменений, а также автоматическое резервирование данных (в виде копии текущего состояния и копий журнала операций) по заданному расписанию. Хранилище использует 64-битную нумерацию объектов, что теоретически позволяет хранить до 2^{64} объектов объемом до 2 ГБ каждый, подключив до 2^{16} разделов. При этом, возможно использование разных хранилищ для разных типов объектов, что теоретически позволяет еще больше увеличить допустимый объем БД. Предусмотрены средства ручного администрирования хранилищ.

Такая организация хранилищ позволяет обеспечить высокий уровень надежности системы, поскольку сбой любого из модулей системы не может привести к глобальному сбою всей системы. Каждое хранилище имеет средства автоматического восстановления данных. Даже в случае сбоя, данные могут быть восстановлены. Для исключения нарушений логической целостности данных БД в разных хранилищах применяются двухфазные транзакции.

Структура хранилища приведена на рис 2. Для работы с хранилищем формируется объект сеанса работы, через который производится обращение к хранилищу, в т.ч. через CORBA. Блок управления хранилищем обеспечивает диспетчеризацию запросов, блокировку объектов при выполнении всех операций, обеспечивает ведение транзакций для каждого сеанса работы, в т.ч. обеспечивает детектирование ситуации мертвых блокировок.

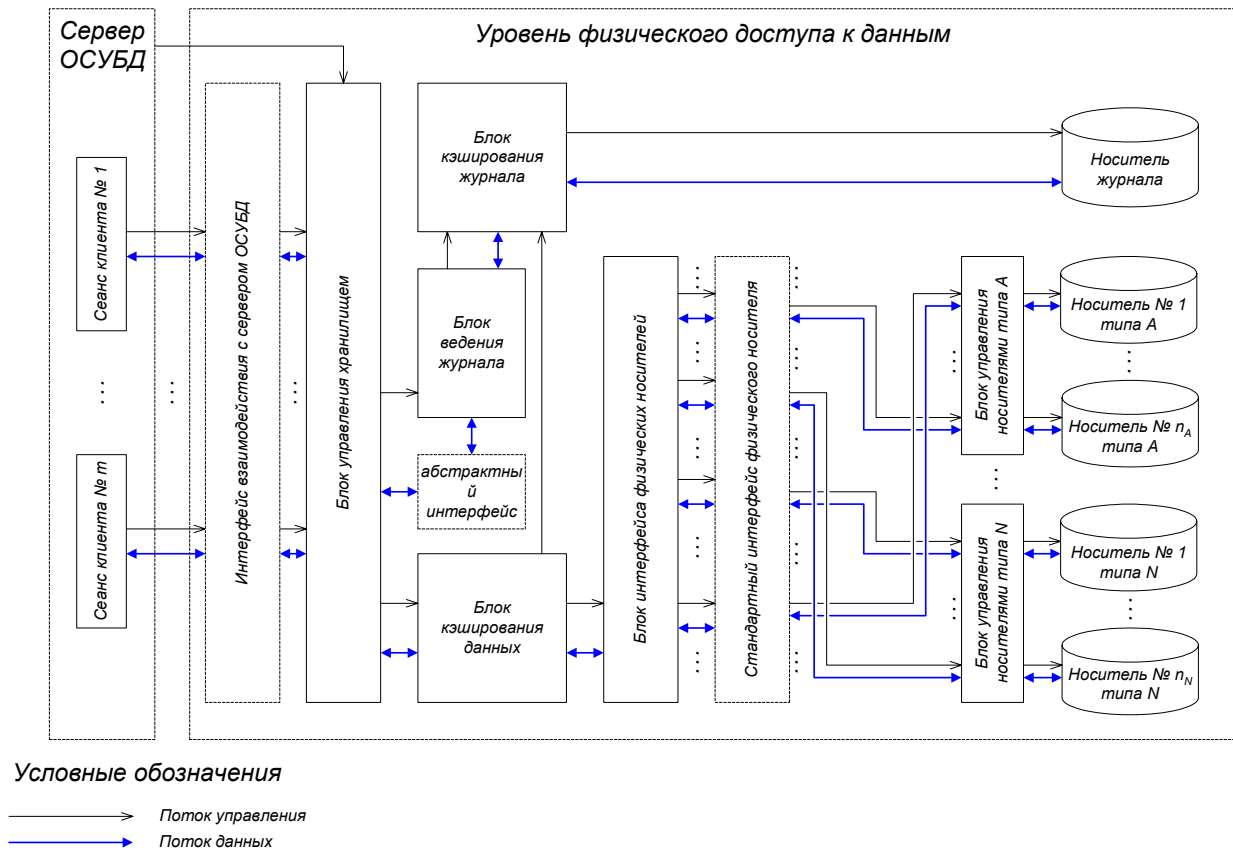


рис. 2 структура хранилища.

Интерфейс физических носителей предназначен для унификации интерфейса различных физических носителей. Для обеспечения большей гибкости работы, предполагается, что хранилище сможет использовать в качестве носителей: файловую систему для хранения объектов как файлов, файловую систему для хранения объектов в одном файле (эмуляция файловой системы на существующей файловой системе), раздел дискового накопителя (специализированная файловая система), а также, другую СУБД как физический носитель, при использовании специального транслятора объектов. Хранилище может работать одновременно с несколькими типами носителей.

Специальные средства кэширования позволяют оптимальным образом использовать доступную ОЗУ для минимизации дисковых операций при типовых запросах сервера ОСУБД.

Блок ведения журнала производит регистрацию всех изменений. Используемый алгоритм самовосстановления при любых изменениях производит синхронную запись в журнал и, лишь потом, асинхронную запись основных данных. Возможна полная регистрация всех операций с хранилищем, включая фиксацию идентификаторов пользователей, запросивших объект.

Схема данных

Каждый объект ОСУБД ODB-Jupiter имеет уникальный идентификатор, состоящий из: 32-битного номера хранилища, определяемого временем создания хранилища, строковым именем типа с максимальной длиной 255 символов и 32-битного малого идентификатора объекта, уникального в пределах своего типа и хранилища. Полный идентификатор объект получает в момент своего создания, сохраняя его в дальнейшем. Это позволяет выполнять перенос объектов из одной БД в другую с такой же схемой данных без опасения встретить другие объекты с таким же идентификатором, поскольку в другом хранилище даже при пересечении имени типа и малого идентификатора объекта, идентификатор хранилища при создании этих объектов будет другим. Следовательно, появляется возможность переноса объектов с учетом всех связей.

Каждый объект содержит запись о своей дате создания и дате модификации.

Сервер ОСУБД выполняет операции с объектами, основываясь на схеме данных. Среди системных типов можно выделить следующие:

Register – тип, являющийся описателем других типов.

%Level, %Category, %Area, %Group, %User – типы для хранения объектов-описателей безопасности.

Для каждого нового типа создается объект типа “Register”, содержащий:

- имя типа и список его алиасов (применительно к ИПС отображаемое имя типа может отличаться от системного);
- список предков (множественное наследование разрешено, но конфликты имен должен устранять пользователь);
- уникальный номер типа в пределах схемы данных;
- список описателей реквизитов.

Описатель реквизита представляет собой следующую структуру:

- название реквизита;
- тип реквизита;
- номер реквизита в типе, который не меняется при любых изменениях типа;
- специальные флаги реквизита (поисковый, множественный и пр.)

Любой реквизит в итоге характеризуется уникальным 32-битным номером, составленным из 16-битного номера типа и 16-битного номера реквизита в типе. Таким образом, при наследовании реквизита, в типе-потомке он сохраняет свой номер, что делает возможным быстрое определение базового типа, содержащего этот реквизит, а также устраняет проблему множественного наследования, заключающуюся в пересечении реквизитов одного общего предка в нескольких ветвях наследования. Администратор системы не должен допускать только наследования одноименных реквизитов от разных предков, что в реальной ситуации легко осуществимо.

Администратору доступны следующие типы реквизитов:

- целое число со знаком 16 бит;
- целое число со знаком 32 бита;
- целое число со знаком 64 бита;
- вещественное число 16 бит;

- вещественное число 32 бита;
- дата;
- время;
- дата-время;
- строка без разбиения переменной длины;
- строка с разбиением переменной длины (при индексации каждое слово строки рассматривается как отдельный ключ);
- двоичные данные (до 2 ГБ);
- ссылка на объект;
- текст документа (отличается от строки иным режимом индексации и поиска).

Администратор имеет возможность изменения типов и набора реквизитов в том числе в процессе эксплуатации системы. Причем, сервер обеспечивает хранение объектов даже с такими реквизитами, которые не заданы схемой данных.

Администратор задает хранилища для хранения типов. На рис 4 приводится пример назначения типов.

Для каждого типа может быть задано свое хранилище или несколько типов хранится в одном хранилище. Кроме того, администратор может вообще не распределять типы по хранилищам, а задать для типа “Unkown” единственное хранилище. Тогда, модуль поддержки схемы данных будет направлять запросы на обращения к хранилищам в соответствии с типом “Unkown”. Возможны любые комбинации с этим типом.

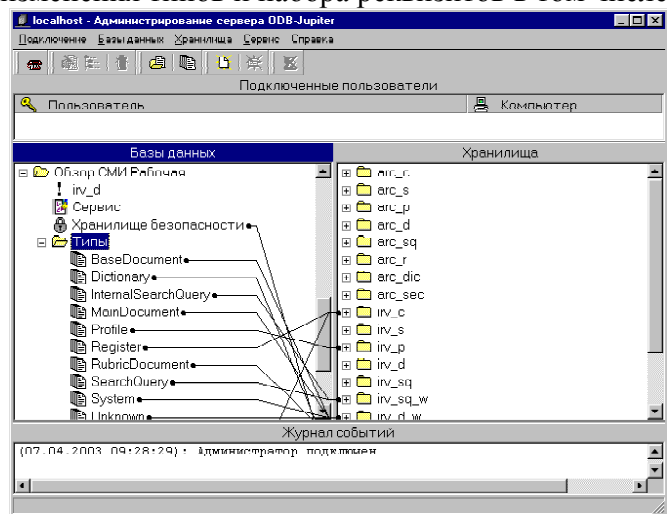


рис. 3 Назначение типов хранилищам.

Разделение прав доступа

Реализация разграничения прав доступа.

Каждый объект в системе наделяется дескриптором безопасности, который содержит идентификатор владельца, дискреционные списки разрешения по каждому виду доступа, дискреционные списки запрета по каждому виду доступа и метки безопасности «чтение» и «запись».

С другой стороны каждый зарегистрированный пользователь имеет свою учетную запись, которая содержит его идентификатор, список идентификаторов групп, в которые он включен, и метки безопасности «чтение», «запись», «база модификации чтения» и «база модификации записи». Список идентификаторов групп используется при проверке условий разрешения доступа механизмами произвольного контроля, а метки безопасности соответствуют уровням безопасности мандатной модели безопасности и используются при проверке разрешения доступа механизмами принудительного контроля.

Таким образом, ОСУБД ODB-Jupiter реализует как дискреционную, так и мандатную модели безопасности.

Произвольный контроль.

Произвольный контроль реализует правила проверки возможности доступа согласно матричной модели безопасности. В системе зафиксированы следующие проверяемые виды доступа:

- Чтение – вид доступа, связанный с операцией чтения объектов и с операцией получения результатов поиска.
- Изменение – вид доступа, связанный с операцией изменения объекта или с операцией создания объекта.
- Удаление – вид доступа, наличие которого позволяет выполнить операцию удаления объекта из системы.
- Изменение дескриптора безопасности объекта (ДБО) – вид доступа по изменению элементов ДБО.
- Смена владельца – вид доступа, связанный с операцией назначения другого владельца объекта.

Непосредственная проверка разрешения доступа пользователя осуществляется по правилам матричной модели безопасности. Сначала проверяется разрешение доступа на уровне типа объекта. Для получения доступа на этом уровне пользователь или одна из групп, куда он включен, должен присутствовать в дискреционном списке разрешений ДБО типа и в то же время пользователь и все его группы должны отсутствовать в списке запретов того же дескриптора. Если на уровне типа объекта пользователь получает разрешение, то выполняется проверка на уровне объекта, а именно, доступ разрешается, если пользователь является владельцем объекта, в противном случае он или одна или несколько его групп должны присутствовать в дискреционном списке разрешений дескриптора безопасности и в то же время он и все его группы должны отсутствовать в аналогичном списке запретов дескриптора объекта.

Принудительный контроль.

Реализация принудительного контроля доступа, соответствует мандатной модели безопасности. Как отмечалось, в дескрипторе безопасности объектов и в учетных записях пользователей имеются метки безопасности «чтения» и «записи», они ставятся в соответствие уровням безопасности, рассмотренным в мандатной модели безопасности. Метка имеет следующую структуру:

- Уровень секретности – элемент упорядоченного множества, такого как например: несекретно, конфиденциально, секретно, совершенно секретно и т.д.
- Категории – неупорядоченное множество, позволяющее разделить информацию базы данных на «отсеки».
- Области – неупорядоченное множество, так же как категории позволяет разделять информацию.

Назначение этих элементов станет понятным после рассмотрения механизма проверки доминирования одной метки над другой.

Доминирование меток

Пусть имеется две метки: А и Б. Операция проверки доминирования позволяет осуществить мнимое отображение меток на элементы множества уровней безопасности L, характерное для модели безопасности Белла-ЛаПадулы, что в свою очередь дает возможность их сравнения. Указанное отображение приведено на рис. 3.

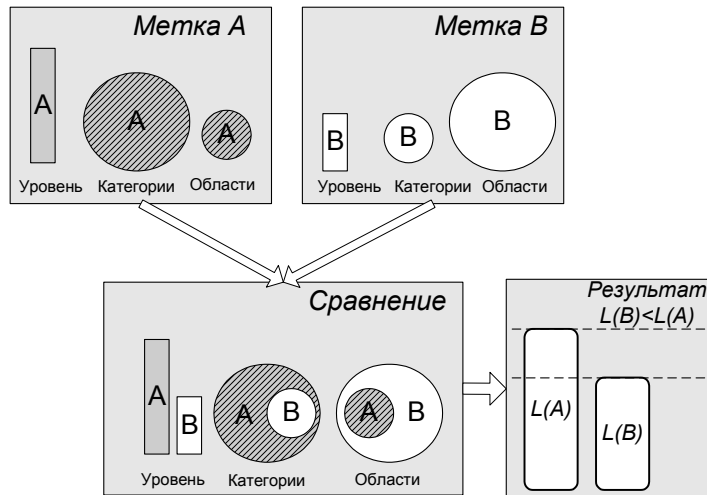


рис. 4 Отображение меток на множество уровней безопасности.

Таким образом уровень безопасности L(A) больше уровня безопасности L(B), то есть метка А доминирует над меткой В, если уровень секретности А выше уровня секретности В, категории метки В включены в категории метки А, и области метки А включены в области метки В.

Принудительный контроль реализуется на основе правила доминирования меток. Как было отмечено, каждый объект БД и каждый пользователь имеет метку «чтение» и мету «запись». Пользователь получает возможность чтения информации из объекта, если его метка «чтение» доминирует над одноименной меткой типа объекта и самого объекта. Аналогично для записи.

Индексация и поиск.

ОСУБД ODB-Jupiter в качестве индекса использует В+-деревья. Для того чтобы имелась возможность поиска по реквизиту, необходимо, чтобы для него был создан индекс (отдельное дерево). При наследовании реквизита, соответствующий индекс этого реквизита должен быть создан у каждого потомка.

В случае поиска по реквизиту предка, сервер ОСУБД по схеме данных автоматически определяет полный список индексов для данного реквизита у всех потомков. Это делает возможным по реквизиту предка автоматически осуществить выборку по объектам всех его потомков.

ИПС ODB-Text

ИПС ODB-Text представляет собой набор инструментов для создания специализированных ИПС. Он включает:

- сервер ОСУБД ODB-Jupiter;

- сервер хранилищ и Ogb-сервер;
- специальные утилиты для управления хранилищами;
- программу администратор ОСУБД;
- редактор схемы данных;
- программу настройки шаблонов Web-сервера;
- универсальный клиентский интерфейс с возможностью настройки профиля пользователя.

Схема данных

Для управления БД и хранилищами (выполнения служебных операций типа индексации, дефрагментации, групповые операции изменения дескрипторов безопасности и пр.), изменения настроек безопасности (изменение категорий, уровней безопасности, областей, списка групп и пользователей, их паролей доступа и прав доступа по типам.) предназначена программа-администратор.

На рис 5 представлены основные этапы настройки БД. Вначале, создается БД и набор необходимых хранилищ. Для БД назначается хранилище безопасности, в котором будут храниться все объекты безопасности. Хранилище безопасности может быть одним на несколько БД в том случае, если все пользователи работают с разными ИПС, но имеют однородные права доступа. Кроме того, создаются другие хранилища, которые будут необходимы в процессе работы.

Определяются категории, области и уровни доступа, а также списки групп и пользователей.

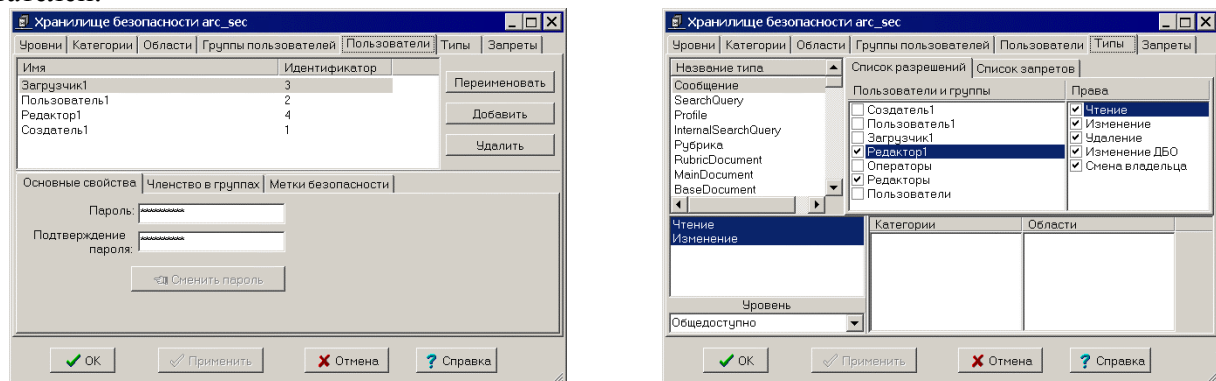


рис. 5 Настройка прав пользователей.

Создание типов производится в программе редактор схемы данных интерфейс которой представлен на рис 6. Она создает набор системных типов, которые необходимы для работы любой ИПС, построенной на основе ИПС ODB-Text.

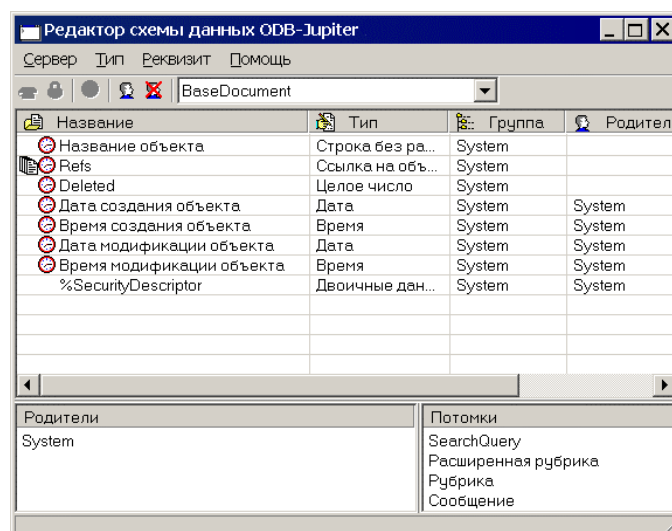


рис. 6 Редактор схемы данных

На основе ИПС ODB-Text возможно построение специализированных ИПС. В некоторых случаях для этого требуется лишь определение дополнительных типов, в некоторых требуется внесение изменений в логику клиентского приложения.

На рис. 7 представлена схема данных ИПС ODB-Text. Типы, выделенные цветом, будут доступны пользователю как рубрики и документы, соответственно. Добавление новых реквизитов возможно на любом уровне. Создание же собственной иерархии типов возможно только по аналогии с типами «Рубрика», «Документ с файлами», «Документ». Любой документ должен быть потомком типов «BaseDocument» и «MainDocument». В случае, если необходимо хранение файлов без поиска по их модержимому, тип должен быть потомком «FileDocument» и «MainDocument». Тип рубрик должен быть потомком «BaseDocument» и «RubricDocument».

Таблица 1. Назначение системных типов.

Наименование типа	Назначение типа
System	Общие реквизиты (дата время создания, дескриптор безопасности)
BaseDocument	Общие реквизиты для типов объектов, имеющих название, предназначенных для размещения в рубрикаторе
MainDocument	Предок всех документов
RubricDocument	Предок всех рубрик
FileDocument	Предок всех документов, содержащих двоичные файлы
Dictionary	Хранение словарей для альтернативных поисковых запросов
Profile	Хранение профиля пользователя
InternalSearchQuery	Предок типа хранилища запросов
SearchQuery	Хранение запросов
SystemFile	Хранение шаблонов для Web-интерфейса

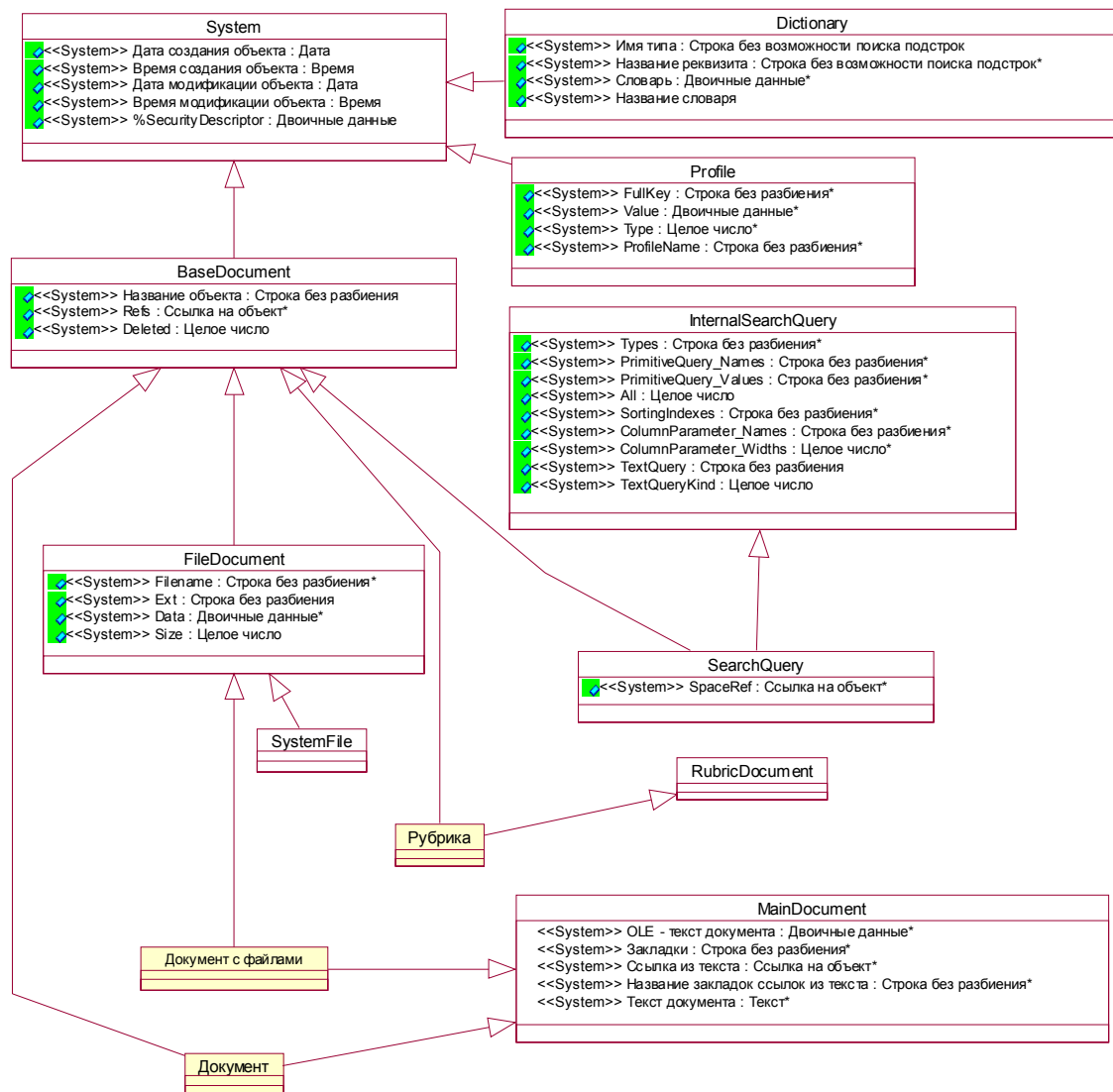


рис. 7 Схема данных ODB-Text.

Рассмотрим особенности организации ИПС, построенных на базе ИПС ODB-Text.

ИПС «Система обработки потоков сообщений электронных средств информации»

ИПС «Система обработки потоков сообщений электронных средств информации» (ИПС Обзор ЭСИ) предназначена для получения, накопления, рубрицирования, фильтрации и анализа новостных сообщений, получаемых из сети Интернет и различных внутренних электронных источников. Структура ИПС Обзор ЭСИ представлена на рис 8.

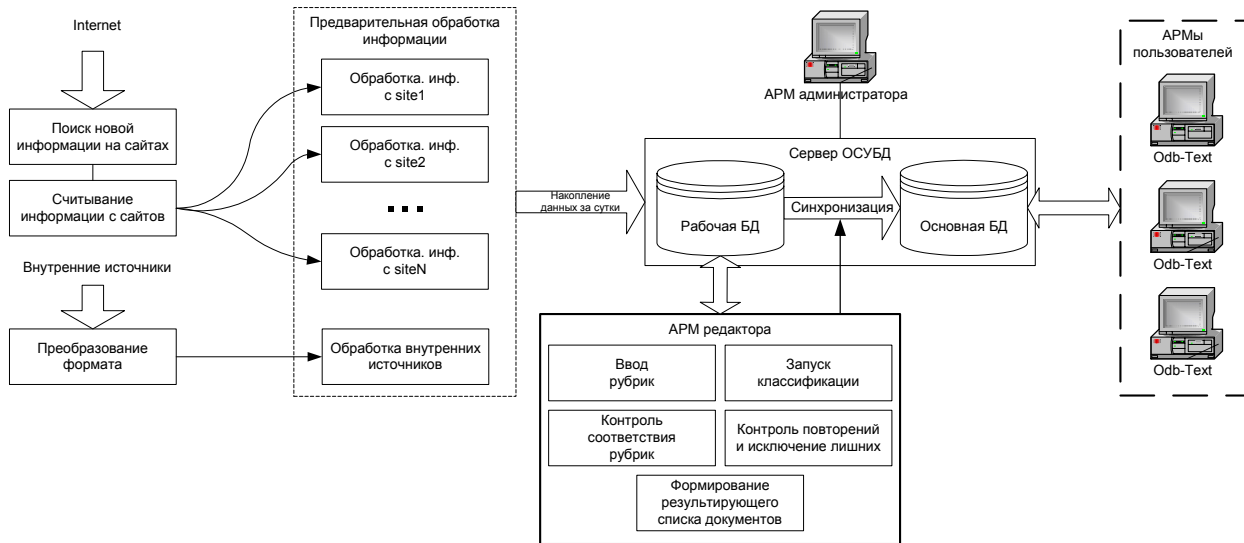


рис. 8 Структура ИПС Обзор ЭСИ.

Основными компонентами системы являются:

- комплекс конвертеров входной информации, предназначенный для преобразования разнородной входной информации в единый формат, пригодный для дальнейшей загрузки в рабочую БД;
- загрузчик, производящий загрузку предварительно подготовленных сообщений в рабочую БД;
- сервер ОСУБД;
- рабочая БД, содержащая все сообщения, когда-либо поступившие в систему;
- основная БД, содержащая сообщения, обработанные редактором, отрубрицированные и разрешенные к публикации;
- АРМ редактора, предназначенный для выполнения рубрицирования сообщений, контроля правильности помещения в рубрику, удаление одинаковых по смыслу, подтверждение готовности к публикации;
- АРМ пользователя, предназначенный для работы пользователя с разрешенным для просмотра набором сообщений.

Схема данных

На рис 9 приводится схема данных ИПС Обзор ЭСИ. Ключевым типом является тип «Сообщение».

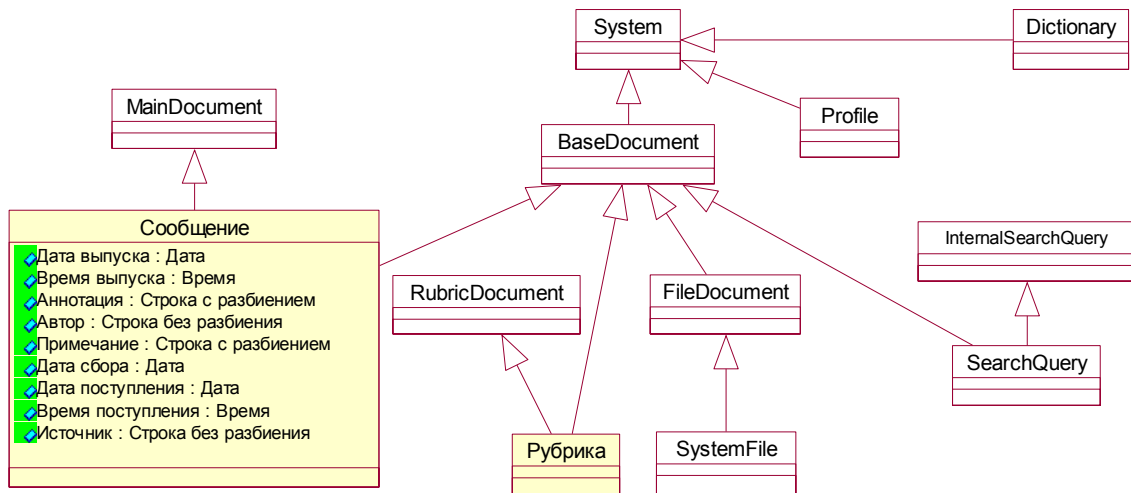


рис. 9 Схема данных ИПС Обзор ЭСИ.

Получение информации из Интернет

В состав ППП Обзор ЭСИ входит модуль сбора информации из Интернет источников. Он предназначен для считывания новостной информации с сайтов, выделения значимой ее части и преобразования в специальный формат для последующей загрузки в БД. По возможности обрабатываются: название статьи, дата публикации, дата написания, аннотация и полный вариант статьи. При этом, удаляется «мусор» типа банеров, различных ненужных оформительских элементов, ссылки на статьи по этой тематике и пр.

Преобразование основано на предположении статичности Интернет сайтов, т.е. основные разделы и размещение блоков новостей в пределах одного раздела остаются неизменными при обновлении новостей. То, что html-формат представляет собой иерархию элементов, позволяет зафиксировать ветвь элементов, в которой находится новость.

Исходная html-страница преобразуется в xhtml для того, чтобы обеспечить возможность поиска с помощью XPath. Запросы для XPath, очевидно, являются уникальными для каждого сайта. Процесс выделения значимой информации строится таким образом, чтобы полученные списки XPath содержали именно те элементы, которые нужны. Этому способствует тот факт, что, часто, новостные блоки помещаются в отдельные таблицы, а сами новости в строки и колонки этих таблиц. Кроме того, часто эти элементы имеют специфические атрибуты типа «class». Следовательно, получение новости заключается в том, чтобы выделить именно такую последовательность вложенных элементов, которая приведет к новостям, рассматриваемым в данном случае, как массив. Например, `<html><body><table class="main"><tr><td><table class="news">` - предполагает, что вложенные на следующем уровне элементы `<tr>` представляют собой список новостей.

Помимо того, что новости имеют фиксированное положение, как правило, такие атрибуты как название, дата публикации, ссылка на статью целиком, имеют фиксированное положение в тексте самой новости. Таким образом, для каждой новости обеспечивается выделение необходимых атрибутов, получение полной версии статьи, удаление «мусора», преобразование html-фрагмента статьи в документ MS Word и получение пары специальных файлов - .xml для реквизитной части новости и .irs для текста статьи в формате MS Word.

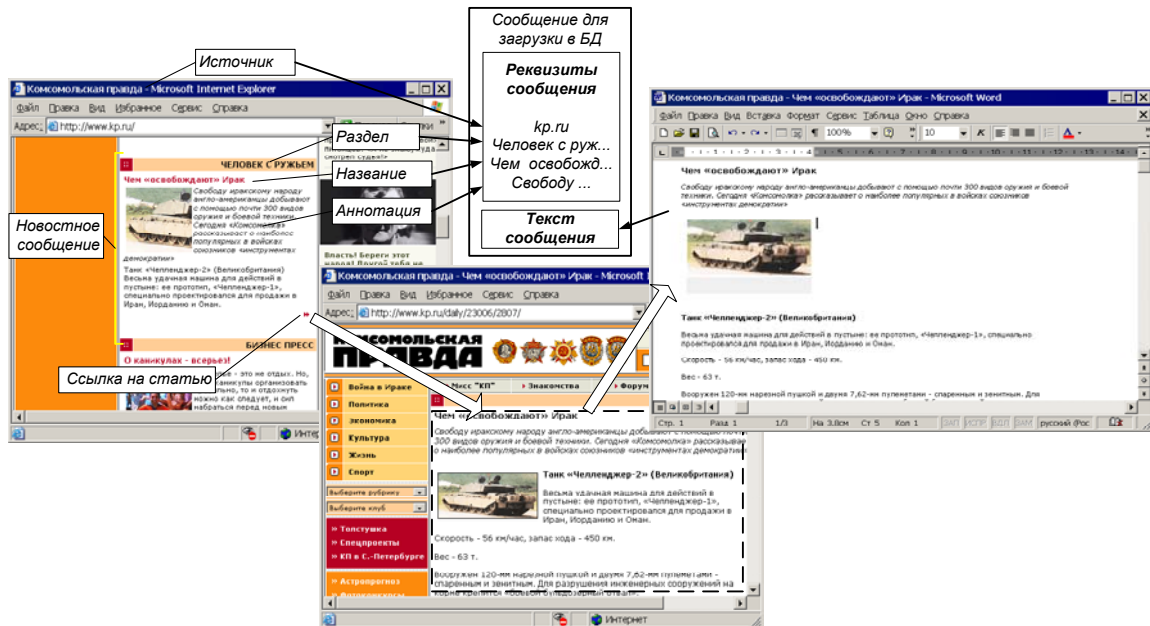


рис. 10 Выделение новостного сообщения с сайта.

АРМ редактора

После загрузки сообщений, они автоматически попадают в рубрику «Нерубрицированные». Задача редактора заключается в том, чтобы настроить рубрицирующие запросы, запустить их на выполнение и включить результат в предварительно созданную рубрику.

Внешний вид АРМ редактора приведен на рис 11 .

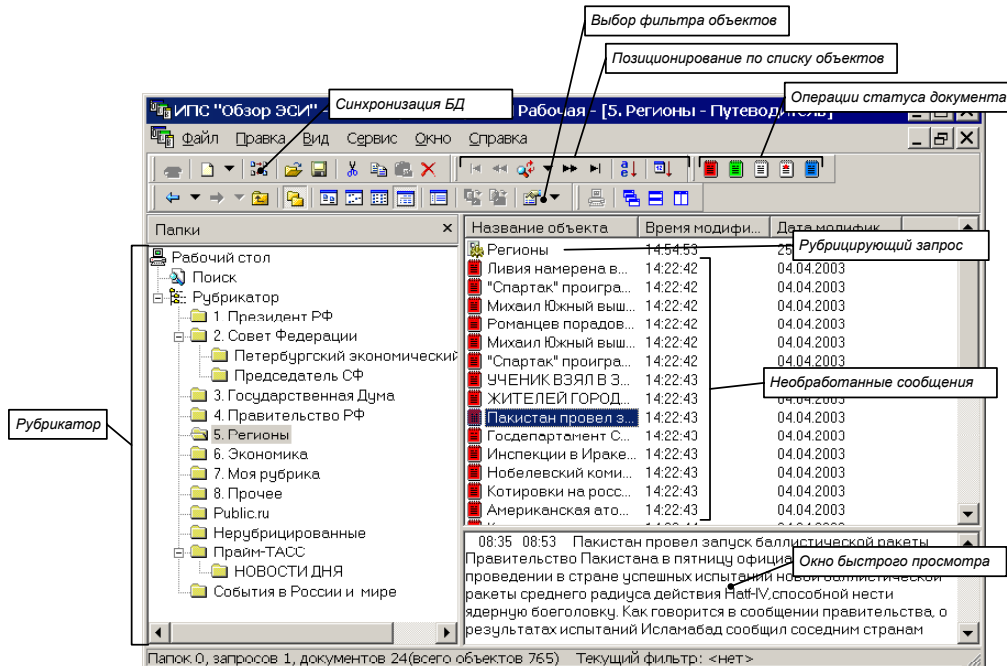


рис. 11 АРМ редактора

Рубрицирующие запросы должны корректируются редактором при необходимости. Редактор может запустить на выполнение и включение результата в рубрики все запросы во всех подрубриках сразу. Итогом включения результата рубрицирования являются помещенные в рубрику документы, отмеченные красным цветом. Редактор должен подтвердить размещение документа в текущую рубрику или исключить документ. Один документ может быть одновременно включен в несколько рубрик. Если одновременно работает несколько редакторов, то обрабатываемые документы можно пометить как рабочие. Это, также исключит возможность попадания документа в другие рубрики при запуске последующих рубрицирующих запросов.

Для просмотра текста документа редактору предоставляется окно быстрого просмотра, в котором отображается текст без исходного форматирования и изображений. Редактор принимает решение о статусе документа и может исключить из рубрики, пометить как опубликованный или актуальный. В случае необходимости, редактор может изменить содержание сообщения. Редактором в ИПС является редактор MS Word.

После завершения контроля документов, редактор должен выполнить синхронизацию рабочей БД, в которой он работал с основной БД, предназначенной для конечных пользователей.

Документы, которые остались в рубрике «Нерубрицированные» являются ненужными и могут быть удалены.

Фильтры

Для удобства работы пользователя (как редактора так и конечных пользователей), в системе предусмотрены возможность использования фильтров объектов. Пользователь может настроить фильтр отображения, отображающий только объекты, удовлетворяющие определенным условиям, например, документы за 3 дня, документы за неделю, все документы и пр. Статус «Актуальный» для документов сделан для того, чтобы временные фильтры не влияли на их отображение.

Поиск документов

ИПС обладает развитыми поисковыми возможностями:

- выбор типа для поиска (документы и/или рубрики). В зависимости от выбранного типа, будет доступен соответствующий набор реквизитов для поиска, например при одновременном поиске по рубрикам и документам, будет возможен поиск по дате создания/модификации и названию.
- Поиск в конкретной рубрике.
- Уточняющий поиск.
- Сортировка по нескольким реквизитам с указанием порядка сортировки.
- Сохранения запросов как объект соответствующего типа.
- Использования ключей в качестве основы запросов по тексту. Под ключевыми словами понимается список слов, помещенных в индекс при индексации текста документов.
- Использование словаря альтернативных запросов.

Словарь альтернативных запросов позволяет для каждого реквизита задать устойчивые слова/словосочетания, характеризующие данную предметную область, т.е. максимально упростить конечному пользователю работу формирование поисковых запросов. Кроме того,

отображаемое словосочетание может скрывать под собой более точный запрос, например, под словосочетанием «регионы России», отображаемым пользователю, может находиться строка запроса, перечисляющая все регионы России. Редактор словаря запросов приведен на рис 12.

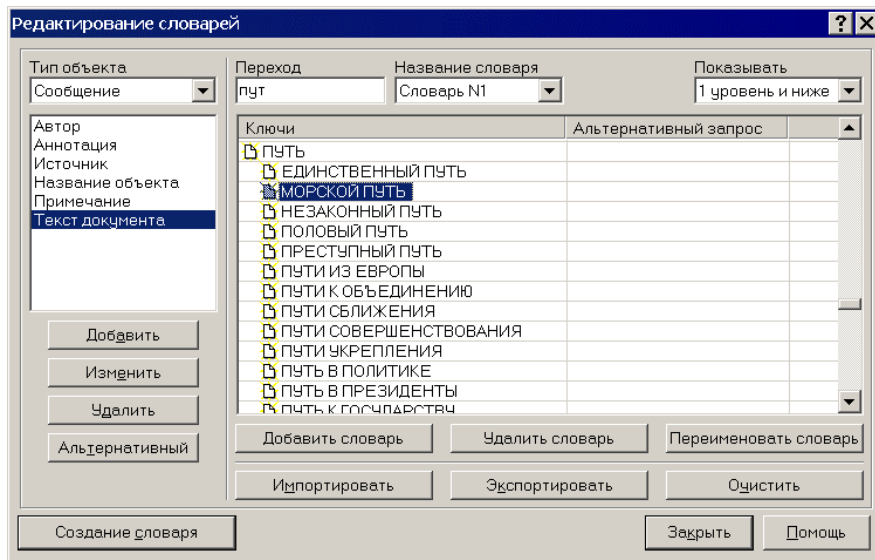


рис. 12 Реквизиты документа

Еще одной особенностью является то, что количество альтернативных словарей запросов не ограничено, т.е. можно создать словарь «Политическая жизнь», «Научные достижения» и пр., в которых одни и те же словосочетания могут иметь различные развернутые строки запросов.

АРМ пользователя

АРМ пользователя подобен АРМ редактора за исключением отсутствия возможности изменения документов и синхронизации БД.

Следует отметить, порядок сортировки документов в рубрике настраивается пользователем. Предполагается, что по умолчанию, пользователь видит документы в порядке убывания их даты поступления.

Пользователь может просматривать документы аналогично редактору, пользуясь окном просмотра или открыть документ, чтобы увидеть документ с иллюстрациями и исходным форматированием. Кроме того, пользователь может просмотреть реквизитную часть сообщения – дату/время поступления, источник, автора сообщения и пр. . На рис. 13 представлены варианты просмотра документа.

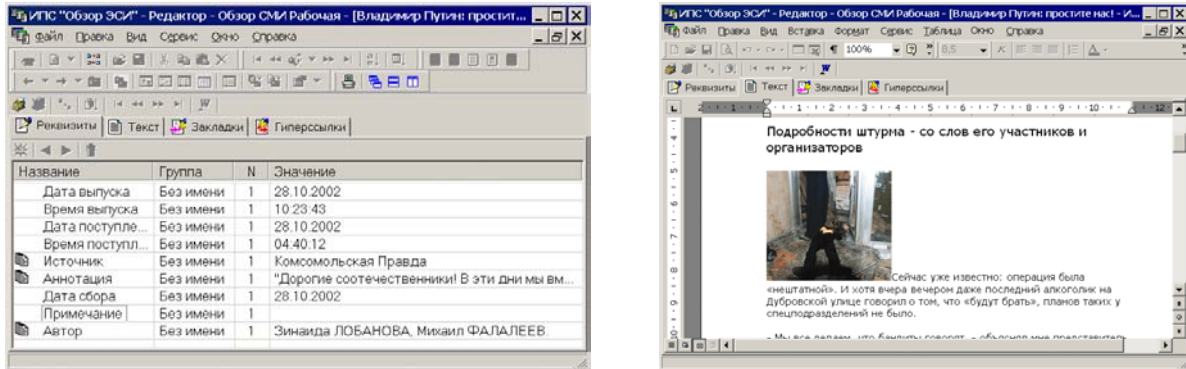


рис. 13 Реквизиты документа

Пользователь может выбирать отображаемые реквизиты документа, просматривая список документов в виде таблицы, столбцами которой будут реквизиты. Также, возможно построение отчетов в формате html в виде таблицы и текстов сообщений в формате MS Word.

АРМ пользователя WEB

Функционально, пользователи Web могут только просматривать документы и выполнять запросы на поиск.

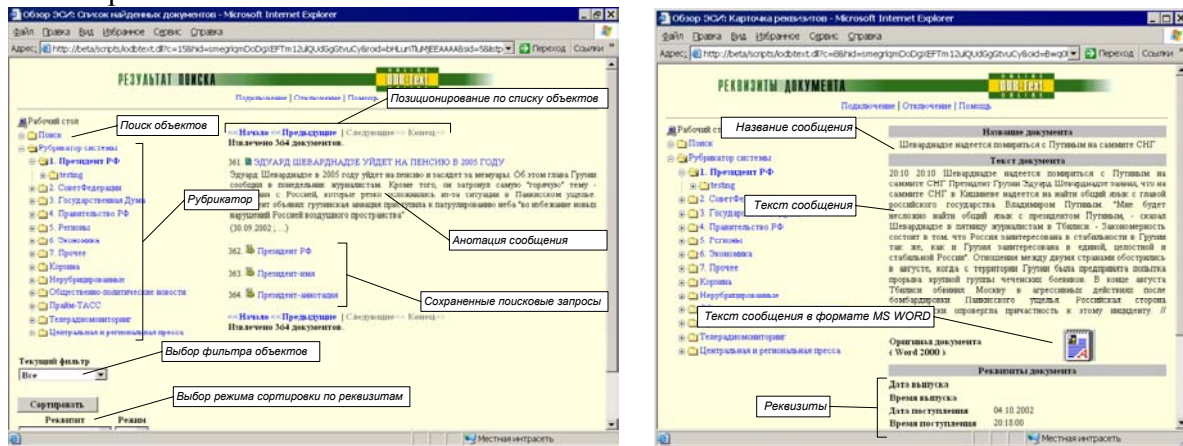


рис. 14 Web-интерфейс.

ИПС Система построения архивов электронных документов

ИПС «Система построения архивов электронных документов» (ИПС Архив) предназначена для ведения архивов электронных материалов, ведения номенклатуры дел, автоматизации ведения дел, построения описи дел, хранения и связывание с документами в БД фото-, видео- и аудио- фрагментов, относящихся к событию, рассматриваемому в документе и пр.

В отличие от ИПС Обзор ЭСИ, ИПС Архив не имеет двух различных БД. Кроме того, основное ее назначение – хранение сложно связанных данных, например материалы, посвященные определенным событиям.

Структура ИПС Архив приведена на рис 15.

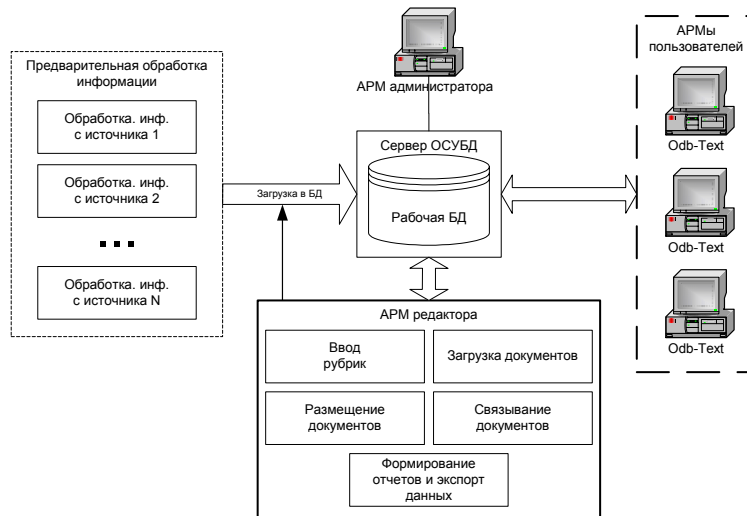


рис. 15 Структура ИПС Архив.

Различие в схеме данных заключается в том, что появляется возможность хранения двоичных файлов без индексации, а также каждый документ может иметь ссылки на другие документы. Такими файлами могут быть графические файлы, видеофрагменты, отчеты в формате MS Excel, ссылки на Интернет-ресурсы и пр., что не нуждается в поиске по содержанию, но должно быть объединено в единый логический блок.

Схема данных

Схема данных ИПС Архив представлена на рис 16. Главное отличие от ИПС Обзор ЭСИ заключается в том, что прямым предком типа «Сообщение» является тип «FileDocument». Это позволяет обеспечить хранение двоичных файлов вместе с сообщением.

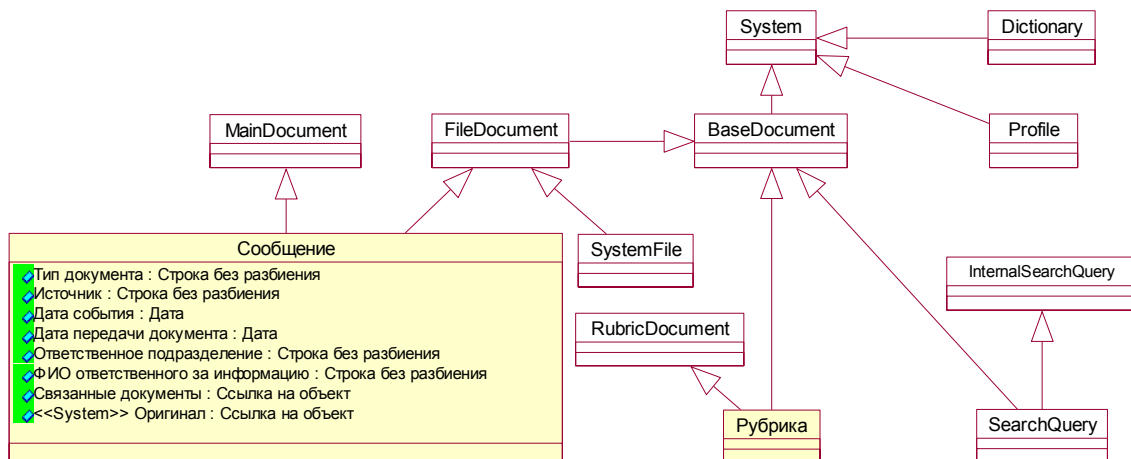


рис. 16 Схема данных ИПС Архив.

АРМ редактора архива и клиента справочной службы архива

Различие АРМов заключается только в том, что редактор имеет возможность редактирования объектов, а пользователь имеет возможность только просмотра.

Отличие от ИПС Обзор ЭСИ заключается в том, что окно предварительного просмотра отображает не только текст документа, но и предложения, соответствующие запросу, если просматривается список найденных документов. Кроме того, редактор получил возможность изменения внедренных в документ файлов и расстановку ссылок между документами.

К возможностям экспорта следует отнести то, что помимо аналогичному в ИПС Обзор ЭСИ экспорту в виде html страницы, возможен экспорт данных в виде запросов MySQL и, соответственно, списка файлов, на имена которых выставлены ссылки.

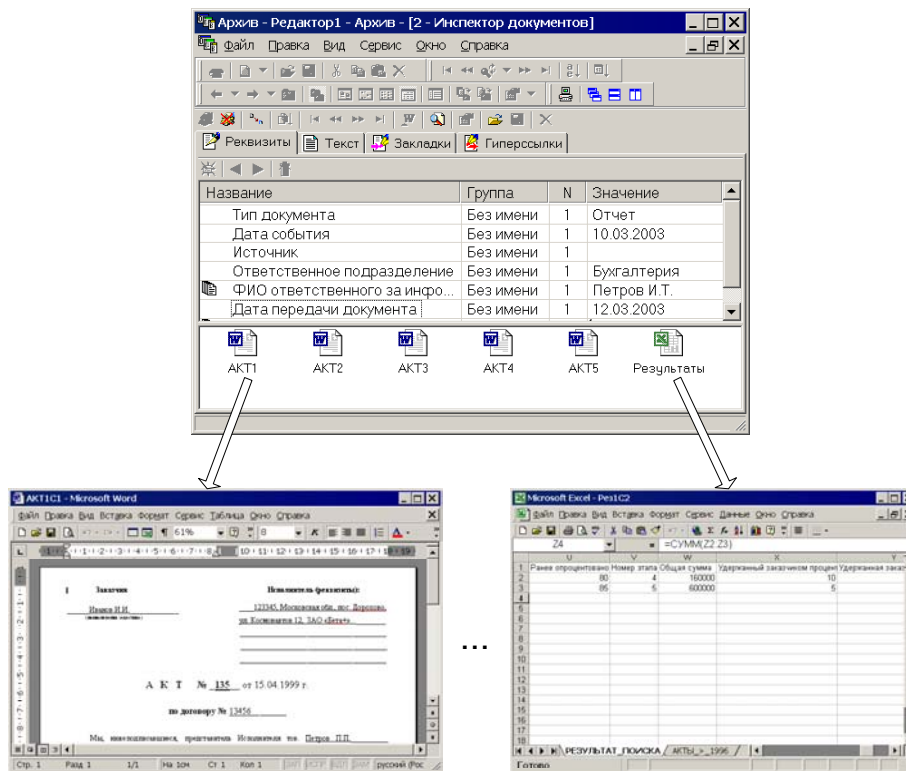


рис. 17 Документ ИПС Архив с вложенными файлами.

Кроме возможности ведения реквизитной части документов, в базовой системе ИПС ODB-Text предусмотрена возможность создания типа, обеспечивающего хранение реквизитов в рубриках. Пример представлен на рис 18.

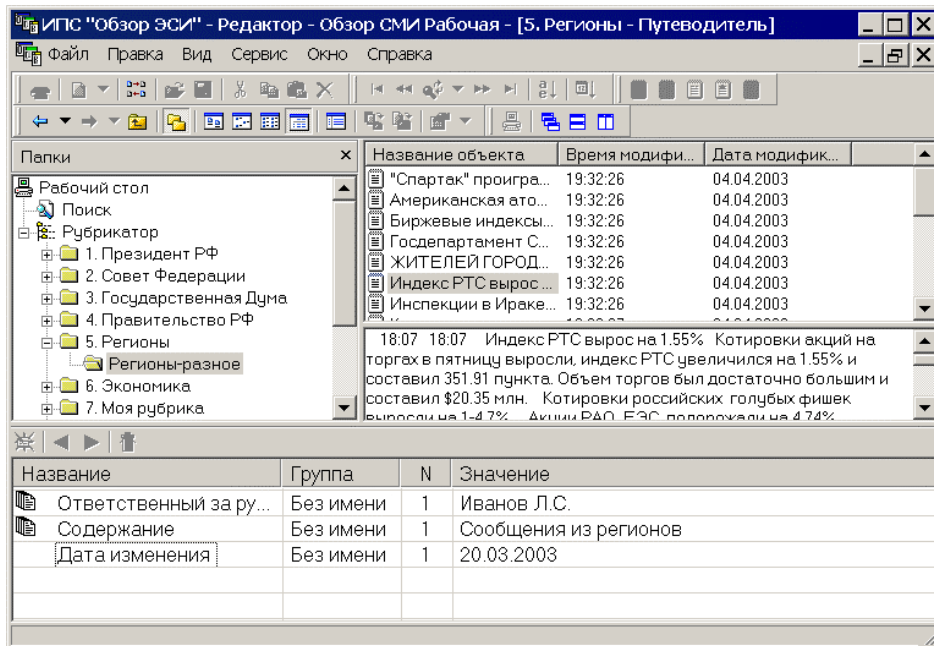


рис. 18 Рубрика «Регионы-разное» с реквизитной частью.

АРМ WEB-клиента справочной службы архива

К АРМ WEB-клиента справочной службы архива также, аналогичен ИПС Обзор ЭСИ, за исключением возможности отображения вложенных двоичных файлов. Пример приведен на рис 19.

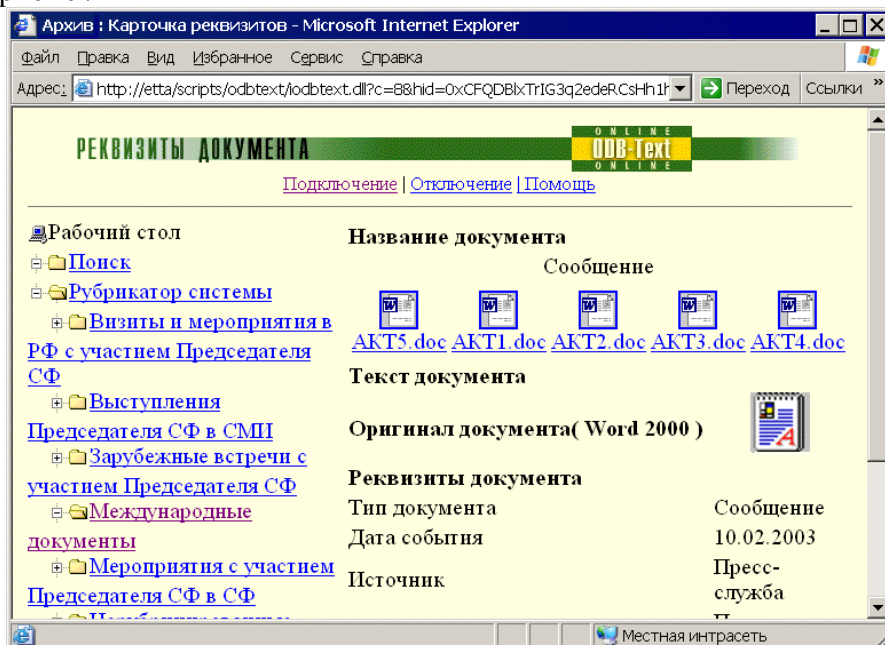


рис. 19 Внешний вид документа в Web-интерфейсе.

Внедрение

В настоящее время, на основе ИПС Обзор ЭСИ разработана система обзора новостных источников, применяемая в Совете Федерации Федерального Собрания РФ. Используется порядка 10 Интернет источников, а также ряд внутренних источников информации. За полгода эксплуатации, в системе накоплено порядка 50 тыс. сообщений.

На основе ИПС Архив, также для СФ РФ разработана система Электронного архива публичных выступлений Председателя СФ РФ.

Заключение

ОСУБД ODB-Jupiter и ИПС ODB-Text являются удобным средством для быстрого построения различных ИПС и архивно-справочных документальных систем, поскольку имеют гибкую систему настройки типов, широкие возможности по объемам хранения данных и характеру хранимых данных. Кроме того, интерфейс ИПС ODB-Text разработан с использованием современных средств быстрой разработки приложений и может быть легко адаптирован разработчиками под конкретного заказчика.