

РАЗДЕЛ I

ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ

Документооборот как информационная база накопления знаний

В. Л. Арлазаров, Н. Е. Емельянов

ИСА РАН активно развиваются три направления автоматизации управления организациями (предприятиями): документооборот, средства генерации информационных систем и управление знаниями. В статье рассматриваются проблемы и возможности их взаимного использования для построения информационных систем, ориентированных на поддержку и автоматизацию процессов управления. Рассматриваются системы обработки документов, такие как Делопроизводство, Документооборот, Управление документами, Управление знаниями, и основные компоненты этих систем. Предлагается целостное понимание процесса обработки документов. Рассматриваются основные средства сбора, хранения и гибкого использования не только самих документов, но и извлекаемых из них знаний. Один из лозунгов подхода: «Создание хранилища знаний о тех информационных объектах, которые описаны в документах». Наряду с обычными компонентами систем управления документами, рассматриваются также системы массовой загрузки потоков документов, автоматической рубрикации, создания БД содержания документов, системы типа OLAP, классификаторы и др.

1. Немного истории

В 1970-х годах, когда начали создаваться автоматизированные системы управления, стало ясно, что одни и те же данные об объекте используются в самых разных задачах. Следовательно, эти данные нужно выделить в отдельную подсистему, которая будет хранить их и обеспечивать доступ к ним всем заинтересованным лицам. Это дало толчок к бурному развитию систем управления базами данных (СУБД). Проектирование баз данных и их организация представляют собой сегодня целую отрасль.

Обычно данные, необходимые для решения задачи (задач), изначально находятся на бумажных документах, вводятся в компьютер и оседают в базах данных, откуда их черпают различные программы. Эта технология, вполне плодотворная, настолько въелась в сознание, что исходные документы стали казаться каким-то ненужным сырьем, от которого не худо бы отказаться, особенно в будущем, когда возобладают безбумажные технологии, и данные будут передаваться по сети от одного компьютера к другому. Существовали, правда, несколько неудобные объекты вроде статей, тексты которых не вполне укладывались в понятие «данного» (реквизита, атрибута, показателя и т. п.), так как не имели ограничений на длину, да и обычные операции (сравнения, поиска и т. п.) с ними производить трудно. Сначала такие объекты просто не воспринимались, потом, скрепя сердце, в базах данных для них стали выделять специальные (так называемые “мето”) поля. Операций с ними не предусматривалось.

Однако, в начале 1990-х годов привычная концепция работы с данными перестала казаться такой уж универсальной, а текстовые документы, напротив, уж очень редкими и специфическими.

При этом прорыв произошел сразу по нескольким направлениям. Первую волну поднял Интернет. Электронная почта, огромное количество доступных сайтов с самой разнообразной информацией потребовали иной, чем в СУБД, системы структуризации данных. «Сообщение», HTML, XML, «поисковая машина» и т. п. — термины из совершенно другой области, чем СУБД. Одновременно с этим начали развиваться системы делопроизводства и контроля исполнения распоряжений. В этих системах понятие документа является основным, даже если в реальности происходит движение только вторичной информации — регистрационных и контрольных карточек. И наконец, фундаментальное обоснование «документного» взгляда на информацию принесли попытки внедрения той самой безбумажной технологии, которая, казалось, должна была с ним покончить. Выяснилось, что для передачи по сети значимой информации она должна быть «заверена» подписью. Но подпись, хотя бы и электронная, ставится не под каким-то данным или набором данных, а только под документом, аналогом того самого, бумажного. Круг замкнулся.

Таким образом, постепенно формируется целая область — системы управления документами (СУД). Можно было бы сказать — документооборот, и это было бы правильно, но, к сожалению, слово это понимается часто в очень узком смысле как некое расширение делопроизводства. Область формируется, хотя далеко еще не оформилась, не стала всем понятной, «переваренной» в своей особенности, с четко выраженной кон-

цепцией, кругом понятий, задач и т. п. Кое-что из перечисленного уже достаточно проработано, другое только начинает осознаваться.

Попробуем обозначить некоторые задачи, отличающие работу с документами от традиционной работы с данными и, соответственно, требующие разработки собственного инструментария.

Прежде всего, речь идет о разработке концепции документа и способов его описания. Понятия структуры, содержания, формы представления для документов существенно другие, чем для «данных», даже в простейших случаях, когда документ может рассматриваться как простой текст или как фрагмент базы данных.

Процедура «создания» документа вообще отсутствует в БД, где документ является понятием внешним. Для СУД это важнейший процесс, который должен поддерживать компоновку целого из частей, версию, возможность выбора вариантов, фиксацию и т. п.

Новым по существу является и понятие пересылки документа. Оно связано с сохранением времени и даты, подтверждением доставки, проверкой электронной подписи.

Создание больших архивов, поиск в них, обработка потока документов, выделение и классификация — типичные задачи при работе с документами. При этом могут использоваться традиционные для БД методы отбора или поиска по комбинациям значений формально выделенных реквизитов, элементы лингвистической обработки, например, полнотекстовый поиск, а также и «искусственно-интеллектуальные» методы определения «близости», содержательных понятий и т. п.

Сюда же примыкают задачи извлечения из документов «знаний» — содержательной структурированной информации — для дальнейшего использования.

Необходимость в новых методах поиска, а также в хранении графического образа документа наряду с его содержанием приводит к специфическим способам организации хранилища документов: дополнительным индексным справочникам по словам, выражениям, понятиям и особым методам сжатия и упаковки самих документов.

По-новому следует подойти и к самой традиционной части работы с документами — вводу и выводу. Когда мы вводим в компьютер «бумажный» документ, во многих случаях необходимо сохранить не только его содержание, но и графическое представление, дабы его всегда можно было воспроизвести в «первозданном виде». Вкупе с технологиями распознавания, это порождает целую систему отношений между описанием документа и различными программными комплексами: экранного ввода и вывода, сканирования — распознавания, печати, передачи в электронной форме. Необходим единый для всех этих комплексов взгляд на соответствие формы представления и содержания. Пока же все

такого рода программы имеют собственные описания форм, проверки целостности, отображения во внутреннее представление.

Однако, в связи с распространением документно-ориентированного подхода, возникает еще одна проблема общего плана. С одной стороны, успешно развиваются системы ввода, вывода, движения документов, архивы и методы работы с ними. С другой — огромное число управленческих программ: бухгалтерия и кадры, планирование и отчетность, заказы и сбыт, а также множество других — полностью ориентированы на работу с базами данных. Возникает разрыв: старые системы, нацеленные на непосредственный ввод документов в БД и вывод их из базы, явно недостаточны, а более современные, учитывающие процессы создания, передачи, хранения, исполнения, идентификации документов, непосредственно с этими базами данных не связаны.

2. Документы — основа управления

Работа с документами является основной частью процессов управления и принятия управленческих решений. Процесс принятия управленческого решения можно разбить на три этапа [1]:

- получение информации,
- переработка и анализ информации,
- подготовка и принятие решения.

Все эти этапы самым тесным образом связаны с документооборотом¹⁾. Если выделить основные компоненты в управлении предприятием (или организацией), то можно их представить в виде схемы (см. рис. 1 из [1]).

В предлагаемой архитектуре системы управления (рис. 1) прикладные компоненты опираются на документооборот предприятия, который является документационным обеспечением управления. Система документооборота должна базироваться на программно-технических платформах предприятия, включать средства и правила создания документов, поддержки их движения, электронный архив и т. д. [2].

Все другие компоненты автоматизированной системы управления (АСУ) предприятием должны опираться на систему документооборота предприятия. Как провести границы между документооборотом и прикладными компонентами АСУ, какие интерфейсы должны быть между

¹⁾ В [1] употребляется термин делопроизводство, так назывались первые версии системы Евфрат. В работе [2] предложен термин документооборот, с более широкой трактовкой, чем делопроизводство.



Рис. 1. Компоненты управления организацией (предприятием)

этими слоями? Этим вопросам посвящен Сборник [3], они представляются актуальными в связи с тем, что старые концепции, которые родились 10 лет назад при разработке Евфрат, отраженные в работе [1], теперь стали насущными при создании современных систем.

В статье [2] отмечено, что современные системы документооборота по существу являются базами знаний о предприятиях. В Сборниках [2] и [3] обсуждаются проблемы интеграции документооборота (Евфрат—Документооборот [4]) с информационными системами, построенными на самых разнообразных платформах (Oracle, SQL, xNika технологии [5] и др.). Принципы интеграции исследуются на примерах управления электронной торговлей [6], маркетингом [7, 8], государственными и муниципальными структурами [9], музеями [10], банками [11], системами защиты интеллектуальной собственности [12, 13], органами исполнительной власти [14]. Ниже делается попытка выделения общих закономерностей такой интеграции.

С юридической точки зрения, **система управления организацией** — это ее документооборот. Любая комиссия (налоговая, экологическая и т. д.) работает только с документами, проверяя их подлинность и соответствие реальности. Поэтому можно сказать: «управление организацией это ее документооборот». С другой стороны, согласно [15]: «**документооборот** — процесс прохождения документов внутри системы управления», то есть только учет поступления документов, их утверждение, исполнение, сдача в архив и т. п. без анализа содержания документов. В статье [19] приводится набор функций, необходимый для расширения понятия документооборота в сторону учета содержания документов²⁾. Около половины статей Сборника [3] посвящены проблеме определения

²⁾ Эта работа получила широкий резонанс, см. [20, 21].

недостающих функций в типовых системах документооборота (например, Евфрат—Документооборот) для того, чтобы на них можно было построить автоматизированную систему управления организацией (АСУ в понимании 1960–1970-х годов). Вернее даже иначе, рассматриваются доработки, которые пришлось внести в типовую систему документооборота для доведения ее до уровня АСУ, потому что документооборот, безусловно, необходим при создании АСУ, и он, естественно, становится ее ядром.

Рассмотрим несколько примеров АСУ. Вопрос проведения электронных торгов в нашей стране в последнее время становится все более актуальным, особенно в части государственных электронных закупок, которые закреплены законодательством [16, 17]. Однако, следует отметить, что «функционирование национального рынка товаров и услуг, его интеграция в мировую экономическую систему сегодня невозможны без развития электронной торговли, понимаемой в широком смысле как использование *электронного документооборота на всех этапах проведения торговых операций*, включая заказ товара, проведение платежей, доставку товара и других операций в торговле» [6]. Таким образом, электронная торговля сведется при этой постановке задачи к электронному документообороту.

Информационная система юридической фирмы «Городисский и партнеры», занятая защитой интеллектуальной собственности (патенты, товарные знаки и др.), построена (см. [12]) на высокой степени интеграции между информационным ядром системы и документооборотом.

В статье [11] описывается архитектура специализированной системы внутренней бухгалтерии банка, интегрированной с системой Евфрат-документооборот. Система включает ведение счетов клиентов и собственных счетов банка, учет операций с клиентами банка, планирование заказов, подробный справочник клиентов с приложенными файлами, с возможностью сканирования, детальное разграничение функций и прав доступа пользователей, создание «электронного архива» сделок с клиентами на основе интеграции с системой Евфрат-документооборот.

Современный музей — сложнейший информационный объект [10]. Музейная информационная система функционально сложнее системы «Операционный день банка» [18]. Правда, к надежности работы банковских систем больше требований, если они останавливаются хотя бы на час — это ЧП. В музеях не так, значит музейные системы пока не являются жизненно необходимыми, и только поэтому они сейчас проще банковских. Хотя, если оценивать системы по стоимости управляемых ими ценностей, то музеи могут оказаться более значимыми, так как стоимость музейных коллекций часто составляет миллиарды долларов.

Чего не хватает информационным системам для управления предприятием?

Лозунг системы Евфрат—Документооборот — Легко управлять! Чего не хватает информационным системам предприятий (например, музейным системам или ядру системы учета патентов [12, 13]) для легкости управления?

Не хватает многого:

- Упорядочения документопотоков.
- Маршрутизации прохождения документов.
- Определения ответственных исполнителей.
- Формирования заданий и сроков.
- Контроля исполнения.
- Анализа и оценки производимых работ.

Другими словами, им не хватает настоящей системы документооборота. В прикладных информационных системах (музейных, банковских, патентных и т. д.) необходимо превратить документы в основную единицу управленческого «хозяйства», поддержать основные функции обработки документов (подготовка нормативных актов и договоров, включение работ в план, выдача справок, отчетов, подготовка и согласование решений и т. п.).

При этом должны быть соблюдены следующие принципы:

- Хранение всех документов (включая версии).
- Возможность увидеть документы со всех рабочих мест.
- Возможность передвижения документов из любой подсистемы в любую (подобно бумажным).
- Информационная безопасность.
- Юридическая значимость.

Суть предлагаемого подхода (см. рис. 2):

- документ — основная единица управленческого хозяйства;
- поддержка основных функций обработки документов (подготовка нормативных актов, договоров, включение в план, согласование, выдача справок, отчетов...);
- прозрачность деятельности организации (взаимосвязи документов и операций).

Документооборот удалось разложить [19, 20] на основные компоненты, их 16 (с учетом подкомпонент — 88). Для организации инструментария при сборке АСУ необходимо разложить на компоненты этап переработки и анализа информации (см. начало статьи). Это трудоемкая

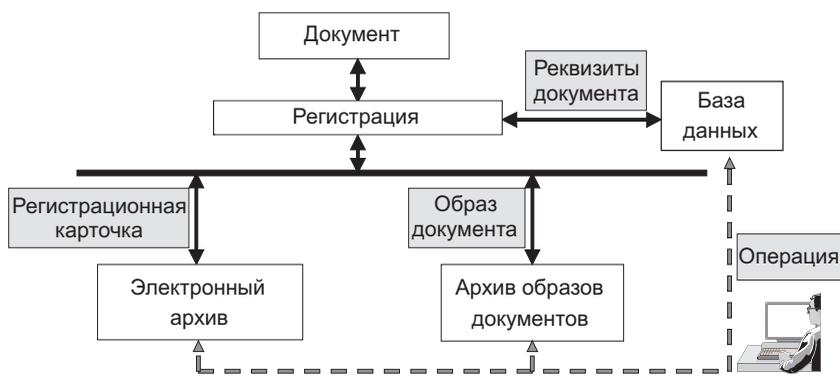


Рис. 2. Современный подход к построению систем обработки документов. Прозрачность деятельности организации (взаимосвязи документов и операций)

и не простая задача. В частности, нужно разложить на независимые компоненты с фиксацией их взаимосвязей (интерфейсов) системы ERP³⁾. Очевидно, что в качестве компонент должны быть аппарат событий, встроенные процедуры переработки информационных ресурсов (подобие встроенных процедур в СУБД), генерация внутренних электронных документов для взаимодействия между подсистемами, ведение протоколов обработки внешних и внутренних заявок (документов) между подсистемами и многое другое.

Документооборот уже разложен на компоненты, и поэтому на его основе можно строить АСУ. Если бы удалось также разбить на компоненты системы класса ERP, то из таких составляющих уже можно было бы конструировать АСУ, подключая и готовые прикладные компоненты (кадры, зарплата и др.).

В сравнении с такими монстрами, как ERP, запуск которых на конкретном предприятии заключается в отсечении множества ненужных возможностей системы средствами настройки, являющейся трудоемкой и требующей высокой квалификации. В перечисленных выше работах — АСУ организаций, построенные на основе системы документооборота путем доработки сравнительно небольшого числа специальных блоков.

Посмотрим, что бы мог дать изложенный выше подход, например, для музеев. По рейтингу журнала «Экономика и Жизнь» № 23, 2003, Евфрат—Документооборот занял первое место по критериям пригодности

³⁾ ERP (Enterprise Resource Planning) — система планирования и управления ресурсами предприятия

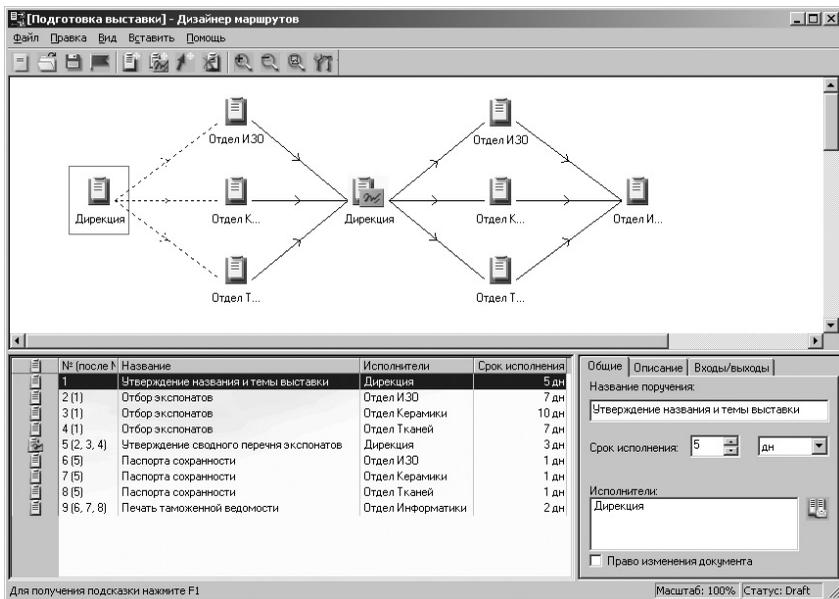


Рис. 3. Представление бизнес процессов в системе Евфрат—Документооборот

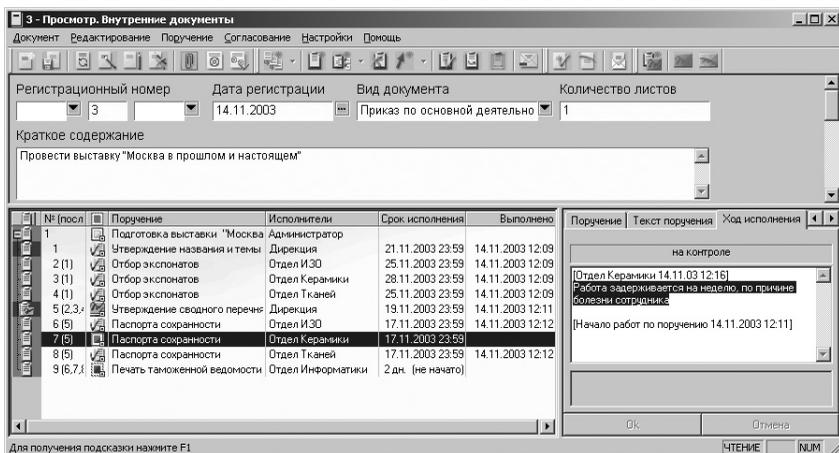


Рис. 4. Отображение состояния дел (выполнению работ) по организации выставки (см. 3) в системе Евфрат—Документооборот

для предприятий малого и среднего бизнеса⁴⁾. К этим категориям относятся все музеи. Евфрат—Документооборот, создан, как и НИКА-Музей, на *x-Nika* технологии, поэтому напрашивается их совместное использование в организации управления музеями. На рис. 3 представлено изображение бизнес-процессов подготовки выставки. На рис. 4 отображается текущее состояние дел по выполнению приказа о проведении выставки. В 3-м столбце специальными знаками отмечены выполненные этапы работ (галочками), просроченные (про них видны причины задержки — правое окно), находящиеся в исполнении и еще не начатые этапы.

Такое представление процессов в виде маршрутов, общее для всех систем документооборота, вовсе не используется в музейных информационных системах. Хотя, конечно, обойти подобные графы невозможно, они существуют на листах бумаги или в виде развернутого описания в проекте инструкции Министерства культуры РФ (МК РФ). Эти графы часто заканчиваются операцией сдачи подготовленных документов в архив, причем архивов возникает столько же, сколько форм документов (а по проекту инструкции их более 80), в то время как при внедрении нормального документооборота они все помещаются в соответствующие папки (дела) единого архива документов организации (наравне с письмами из МК РФ и др.)

Немецкая музейная система MuseumPlus [22] развивается под правильным лозунгом «от управления коллекцией к управлению музеем», но она выросла из музейной системы управления коллекцией, поэтому далека от промышленных систем документооборота. Хотя в системе и появилось упорядочение бизнес-процессов, но так же, как и во всех других музейных системах, нет исполнителей, сроков, контроля исполнения и др.

3. Обычные трудности внедрения документооборота

Есть определенный набор сложностей внедрения систем документооборота.

Первое — это отношение пользователя. Он, как правило, считает, что документооборот — дело канцелярии: «у них любовь к бумагам, к подписям и печатям».

Второе — все любят автоматизировать то, что у них уже есть. Назовем это автоматизацией «неразберихи». Иногда, действительно, очень сложно уговорить сначала все привести в порядок, а уже потом автоматизировать.

⁴⁾ Эта оценка согласуется с оценками: журналов «Мир ПК» 23 мая 2003 г., «Компания» 2002 г. и др

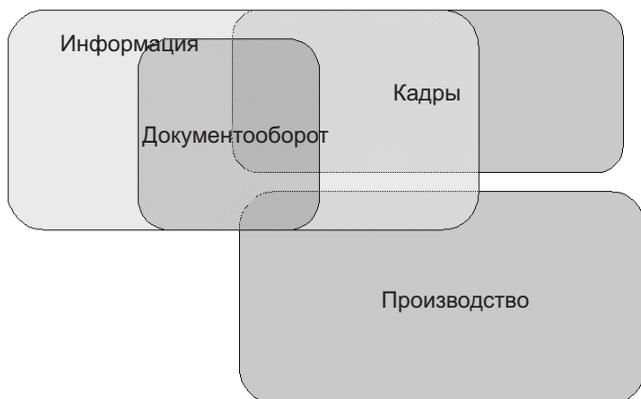


Рис. 5. Место документооборота

Третье — кому это нужно? Например, в банке не стоит вопрос внедрения банковской системы, но под вопросом внедрение документооборота — нужен ли он? Да, наверное, нужен, это хорошо, но он не является системой жизнеобеспечения предприятия, без него жить можно.

И последнее, документооборот (в традиционном понимании) не приносит ничего нового в управление. Например, документы стали передавать по локальной сети вместо курьеров, но для задач управления это мало что дает.

Возьмем произвольную организацию, допустим, что у нее есть подразделения — кадры, производство, и есть некоторое информационное поле. Поле, которое захватывает частично кадры и производство. Есть в организации свои налаженные процессы, и оттуда часть информации идет руководству, есть кто-то, кто занимается СМИ и связью с внешним миром и т. д.

Обычно документооборотом охвачена часть информации (см. рис. 5). Он решает часть задач: обеспечивает движение документов, упрощает контроль исполнения, а также накопление документов и их поиск. Но не решает задачу накопления информации — содержания документов. Если бы было накопление информации (отчетность подразделений, планы, бюджеты, ведомости на получение зарплаты и т. п.), то ее можно было бы использовать, например, для анализа.

Документы (письма и т. п.), содержание которых не может быть структурировано, занимают ~ 30 % (по одной американской работе — 17 %), а остальная информация, курсирующая по организации, является строго структурированной. Сейчас руководителю нужен помощник-управленец для обработки и анализа информации, содержащейся в до-

кументах. Документооборот предлагает набор услуг, а функция, которая нужна руководителю, — это анализ информации (причем, часто довольно простой) отсутствует.

4. Пути преодоления трудностей

Как решают эту проблему. Иногда используют много прикладных систем (ПС) для разных частей, в частности, Excel, Access, Word, отчеты в Crystal или еще что-то (см. рис. 6). Если опять же взять банк в качестве примера, то увидим, что банковские системы пытаются собрать кусочки этой информации для руководства или для других подразделений — из этого, как правило, мало что получается. Очень редко и с большими сложностями конечные данные попадают к аналитику. Получается конгломерат, который нужно собрать в систему.

Мы предлагаем раздвинуть рамки делопроизводства, чтобы оно захватывало всю информацию, курсирующую в организации, чтобы мы могли осуществлять управленческие функции с помощью документооборота. И не только с помощью движения документов, но еще и анализа информации, которая ходит по предприятию.

Как уже мы отмечали, 30 % информации не поддается структурированию, еще около 30 % может быть структурировано, и остальная информация уже структурирована. Мы предлагаем путь создания стандартных форм (может быть, стандартных только для одного предприятия). Например, если ведется график учета рабочего времени сотрудников —

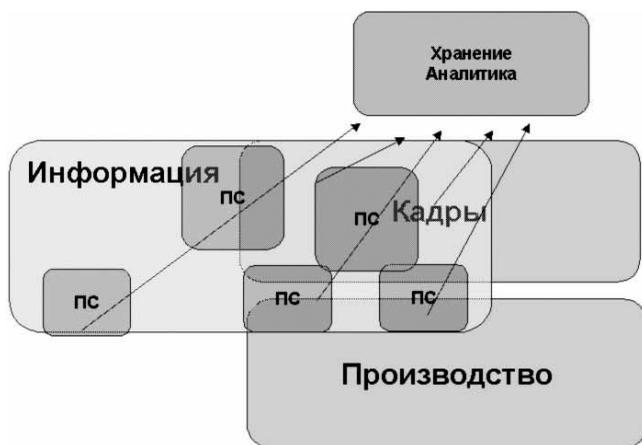


Рис. 6. Обычное решение проблем

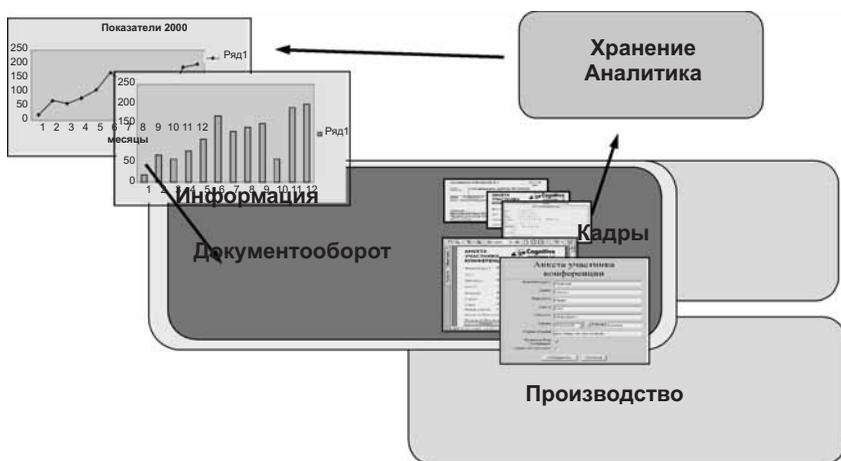


Рис. 7. Единый документооборот для всех структурированных форм

напечатать *бланки прихода на работу*, которые заполняют сотрудники вручную и т. п.

Пусть будет какая угодно форма — либо бумажная структурированная и удобная для заполнения вручную, либо электронная форма для ввода в свой компьютер или на сервер в Интернет. Не нужно ничего нового устанавливать на компьютер — есть готовая форма, которую заполняет человек на рабочем месте, никуда не отходя, и, таким образом он отчитывается перед начальством или запрашивает информацию. В таком случае информация автоматически получается структурированной и может быть введена в базу данных.

Схема документооборота на предприятии приобретает другой смысл. Для всех структурированных документов делаются формы самого разнообразного вида и назначения. После этого данные, собранные по формам, заносятся в единое хранилище данных (см. рис. 7). Они могут использоваться для анализа (в Excel или аналитических системах типа OLAP). После анализа они вновь попадают в систему, но в гораздо более удобном виде — это графики или аналитические отчеты. И уже в системе документооборота циркулирует не та разрозненная, разбросанная информация, а структурированная и проиндексированная. Это все наши *знания* о предприятии, которые можно эффективно использовать.

На рис. 8 изображена общая схема обработки форм. Формы проектируются, далее либо распечатываются бланки для ручного заполнения форм, последующего их распознавания и ввода в БД, либо формы публикуются в электронном хранилище форм для компьютерного заполнения



Рис. 8. Общая схема обработки форм

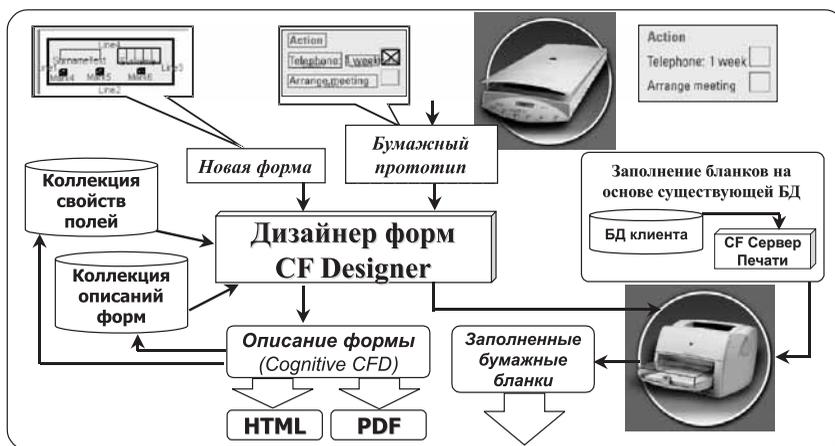


Рис. 9. Проектирование форм

их средствами стандартных интернет-браузеров или специальным редактором структурированных документов с последующим вводом данных в БД.

На рис. 9 представлен процесс проектирования форм документов. Слева показано проектирование новой формы на основе коллекции полей и ранее описанных форм. Отдельные поля из коллекции или целые агрегаты полей (например, паспортные данные) можно использовать для создания новой формы. Справа на рис. 9 показан вариант проектиро-

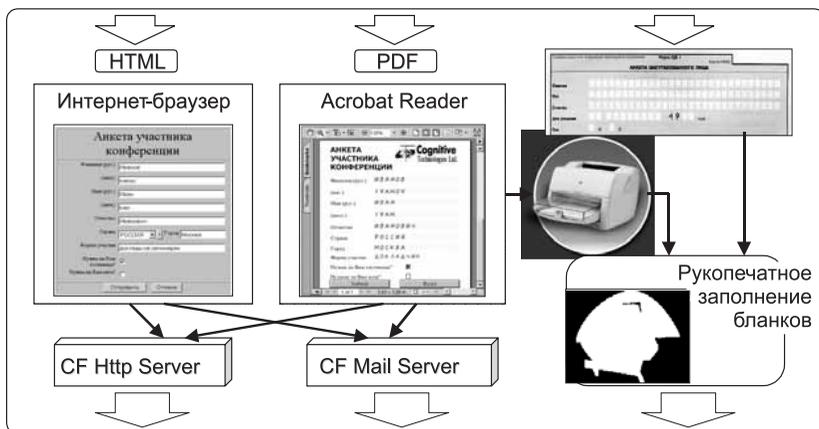


Рис. 10. Заполнение форм

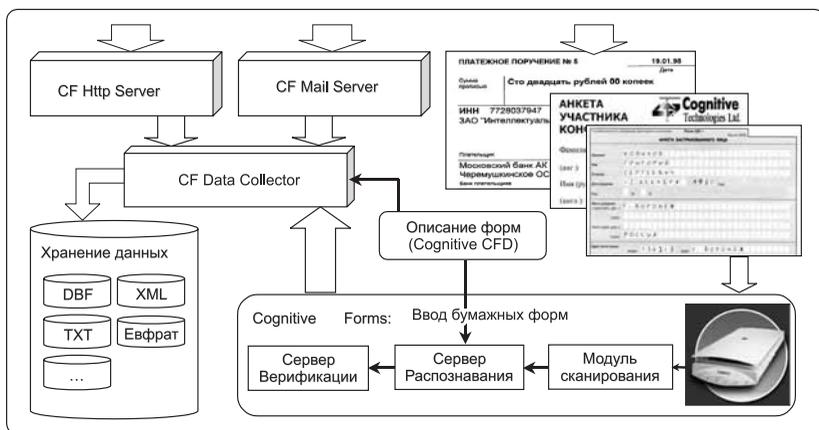


Рис. 11. Сбор и хранение данных

вания формы на основе бумажного прототипа, который сканируется, распознается, обрабатывается дизайнером и размещается в хранилище форм. Эти формы могут быть автоматически заполнены данными из БД пользователя и распечатаны.

На рис. 10 показан процесс заполнения форм. Они могут заполняться средствами интернет-браузеров, редактором структурированных документов, Acrobat Reader или вручную.

На рис. 11 представлены HTTP Server и MAIL Server, которые обрабатывают полученные по интернет и электронной почте документы,

созданные по формам, и (справа) поток бумажных документов, которые сканируются, распознаются, верифицируются и средствами Data Collector направляются на хранение. Система обеспечивает различные форматы хранения: в архиве Евфрат, XML, HTML, текстовые файлы, DBF, запись в любую СУБД, имеющую ODBC интерфейс.

5. Документооборот и управление знаниями. Основная идея реализации подхода

Знания о любой предметной области, любом объекте управления делятся на *декларативные* (знания фактов) и *процедурные* (знание процедур — алгоритмов получения новых фактов из имеющихся). Системы документооборота также содержат эти два вида знаний: *факты*, записанные в документах, и *процедуры* — процессы обработки документов.

Уже общепринятым международным стандартом для представления данных сложной структуры стал XML [23]. В предлагаемом подходе на основе XML вводится стандарт для передачи не только данных, но и форм представления данных, компонент содержания входных и выходных документов, записи типовых запросов, процессов обработки документов и генерации системы. Все это должно храниться в виде XML-документов с обеспечением индексации, быстрого поиска, отображения, корректировки, аналитической обработки таких документов.

Jon Udell в статье [24] пишет: «...ориентация на XML является революцией в управлении знаниями о предприятии. Информационные поля документов можно связать с элементами данных, описанными в схеме XML, и в диалоге вводить документы по схеме. Это революционное достижение.

Мы знаем, что уже многие годы большинство нашей жизненно важной информации живет в документах, не в базах данных. XML поможет нам захватить неявную структуру обычного делового документа (записок, расходных отчетов) и сделать ее явной. Тогда наборы таких документов приобретут вид виртуальных баз данных».

Далее автор отмечает, что эта «великая идея останется мертворожденной» пока не будет стерто различие между XML-документами, которые пишут и читают люди (с использованием программ), и XML-документами, производимыми и используемыми базами данных и WEB-сервисами. Возникающим здесь проблемам посвящены следующие разделы статьи.

6. Технические проблемы реализации

Для того, чтобы идею нового подхода, высказанную в предыдущем параграфе, воплотить в жизнь и на ее основе создать развитую и удобную систему документооборота, необходимо преодолеть технические проблемы.

6.1. Возможность настройки на потребности конкретных приложений

Имеется в виду построение (по возможности, силами пользователей) индивидуальных информационных систем со своим набором входных и выходных форм, типовых запросов, настроенных на потребности пользователей.

Поставленная цель в значительной степени претендует на то, что многие традиционные задачи АСУ — ведение договоров, учет кадров, планирование, отчетность и другие, рассматриваются как профили систем документооборота. Это не удивительно, во всех задачах организационного управления на вход поступают документы (планы и факты деятельности организации, приказы и решения по управлению и т. п.), на выходе — выходные документы (отчеты для вышестоящих и контролирующих организаций, аналитические расчеты для руководства и др.).

Мы не рассматриваем в статье проблем анализа бизнес-процессов в организации, определения набора и функций документов, необходимых для управления. То есть, проблем наведения порядка в организации, о чем мы говорили выше, см. также [25]. Мы предполагаем, что этот анализ уже произведен и составлен проект документооборота для решения частных задач или задач управления организацией в целом, внедрение которого будет поддержано, после технической его реализации, соответствующими приказами и волей руководства.

Рассмотрим проблемы технической реализации проекта.

Первая из них — предоставление возможности настройки на свои потребности. Система должна быть построена на наборе инструментальных средств: дизайнер форм, редактор XML-документов, электронная почта, поисковая система, генератор отчетов, система распознавания и др. (см. [26]). На основе этих средств проектируются и создаются системы для конкретного заказчика. Часто системы документооборота для целого класса объектов управления отличаются только набором входных и выходных форм и некоторой модификацией их обработки. С каждым документом и любым его полем можно связать процесс обработки: от самого простого — сохранить данное под тем же названием, до сколь угодно сложного — образование новых документов, запуск процессов, организация поручений и контроля их исполнения, расчет новых данных и т. п. С классом документов связывается библиотека процессов, есть язык (языки) программирования новых процессов. Имея развитый механизм привязки этих процессов к документам, данным, элементам формы, времени исполнения — так называемый аппарат событий [27], можно создавать новые системы данного класса с учетом требований конкретного приложения, проектируя новые формы и процедуры обработки. Это мы называем «настройкой».

В системе предоставляется возможность конфигурации рабочих мест: наборов БД, с которыми возможна работа, входных и выходных документов (отчетов, справок и т. п.), типовых запросов и словарей, прав доступа и др. По формам входных документов может автоматически генерироваться схема БД, в которую будет загружаться информация из документов. Входная форма определяет не только состав и структуру ее реквизитов, но и отображение их на экране, правила заполнения форм (словари значений ее полей, алгоритмы расчетов, проверки допустимости и целостности и др.).

Пример

Система ведения контрактов (Contracts)⁵. Форма регистрации контракта представлена на рис. 12.

На форме контракта можно указать, за сколько дней или недель до конца срока (End) система должна возобновить контракт (см. рис. 13).

The screenshot shows the 'EditContracts1' window. The 'AutoRenewal' checkbox is checked, and the duration is set to '1 days before expiration'. The table below shows the following data:

#	Product	Quantity	Unit Price	Discount	Price
1	beer				45,20
2	bread				12,80

Рис. 12. Форма описания контракта

The screenshot shows the 'RegisterContracts1' window. The 'AutoRenewal' checkbox is checked, and the duration is set to '1 weeks before expiration'. A red circle highlights the 'AutoRenewal' checkbox and the duration field.

Рис. 13. Указание автоматического возобновления контракта

⁵) Использованы материалы, предоставленные А. Ю. Долгоруковым.

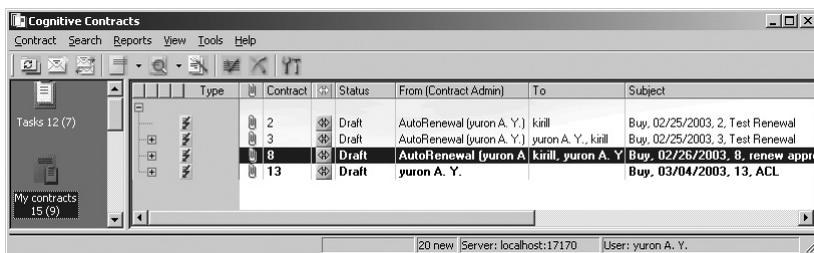


Рис. 14. Появление нового контракта в папке контрактов

При наступлении указанного срока система проверит, что контракт подписан (Executed) и создаст новый контракт, аналогичный исходному контракту. У Администратора Контракта появится новый контракт в папке My Contracts (см. рис. 14).

Срок действия (Duration) нового контракта будет равным сроку действия исходного, а дата начала (Start) — дате окончания исходного контракта (End).

В таблице Associations нового контракта будет упомянут исходный контракт с типом связи (Link Type) genewal.

Система ведения контрактов реализована настройкой системы делопроизводства [28], в которой введены новые формы регистрации документов (контрактов) и отображения их состояния. Введены также новые процедуры автоматического продления срока действия контрактов.

Система поставляет дизайнер форм [29], включающий описание процедур обработки, для проектирования регистрационных и контрольных карточек, что высоко оценили многие пользователи и компьютерные издания. Большинство производителей систем документооборота не предоставляют таких возможностей, но количество спецификаций (подключение процедур автоматического расчета значений полей, проверка целостности, связь с БД и др.) при проектировании форм весьма велико, и превращение дизайнера в «простой» инструмент конечного пользователя пока остается проблемой.

6.2. Организация хранилищ форм

Ориентация на формы и формооборот, как основу автоматизации документооборота, вызывает необходимость определения понятия «форма документа» и поддержки его необходимыми программными средствами. Пока нет такого общепринятого определения. Словарь⁶⁾ определяет

⁶⁾ См. Webster's New Collegiate Dictionary. London. 2000.

форму документа как распечатанный на машинке или типографский документ с пустыми местами для заполнения. Другими словами, форма — постоянная часть документа, содержание — переменная.

Как отделить содержание документа от формы отображения данных? Этот вопрос относится к семантике документа. Если известна семантика документа, то при проектировании формы документа естественно выделяются для ввода содержания документа переменные поля, которые распознаются системой ввода и заполняются системой вывода. Когда мы говорим о содержании документа, мы подразумеваем семантическую модель документа [30].

Необходимы средства проектирования форм и язык описания форм. Язык XForms комитета W3C [31] направлен на решение этих проблем, но пока описывает только их малую часть, в частности, совершенно не учтена возможность использования форм в процессах массового ввода и распознавания бумажных документов.

Кроме того, необходимо создание сети взаимосвязанных хранилищ форм и идентификация форм в хранилищах, которые должны обеспечить получение нужной формы (в зависимости от хранилища, автора, времени, места и других параметров запроса). Решение этой проблемы требует не только создания программных средств, но и организационных решений. Должны быть регламентированы процессы ввода в эксплуатацию новых форм, размещения их в хранилищах, корректировки форм и репликации изменений по соответствующим хранилищам. Определены директивными решениями (приказами, постановлениями — в зависимости от области распространения форм) назначение форм и ответственность за невыполнение правил их применения. Ярким примером внедрения форм являются платежные поручения [32]. С этими формами имеют дело практически все организации России, унификация их и массовое внедрение позволили сэкономить огромные средства на разработку программ обработки данных об оплате через банки. Важно отметить, что для внедрения формы важна не только ее пустографка (бланк или группа бланков), но и правила ее заполнения, словари возможных значений полей и многое другое, что входит в понятие формы.

Аналогично, унификация форм произведена в налоговых инспекциях, пенсионных фондах и ряде других ведомств. Форма может быть принята для использования в рамках организации, корпорации, отрасли, региона, всей страны. Хранилища форм должны обеспечить их распространение в соответствующих сферах.

6.3. Стандарт описания форм

Создание *стандарта описания формы*⁷⁾ преследует следующие основные цели:

- разработка единого понятия формы и сводного тезауруса спецификаций и возможностей разнообразных систем обработки документов;
- создание БД элементов формы;
- разработка механизма и формата сохранения формы;
- упрощение взаимодействия различных комплексов и модулей как собственной разработки, так и внешних между собой;
- объединение программ дизайна форм на уровне концепции, модели формы, ее элементов и формата хранения.

Стандарт должен охватить следующие системы: дизайнеры форм (экранных, печатных, бумажных, ориентированных на последующее распознавание), системы распознавания, генераторы отчетов, документооборот, обмен с внешними системами и др.

В данной статье форма рассматривается как информационный объект, представляющий собой логический образ класса информационных объектов (документов). Форма позволяет порождать информационные объекты этого класса различными способами, гарантируя структурную и логическую непротиворечивость. Форма обеспечивает также взаимодействие информационных объектов, преобразование объектов данного класса в объекты другого класса и отображение информационных объектов для восприятия их человеком. Модель формы — составной информационный объект, однозначно определяющий класс экземпляров формы. Составная природа формы и модели заключается в двух аспектах — во-первых, в том, что и модель и экземпляр состоят из различных блоков, каждый из которых описывает или выполняет свои функции, а во-вторых, в том, что модель может состоять из набора моделей формы, организованного в специального вида структуру.

Модель формы состоит из трех основных компонентов — модели содержания, модели взаимодействия с пользователем, и модели визуализации. В одной модели формы могут присутствовать несколько моделей взаимодействия с пользователем и несколько моделей визуализации. Модель содержания по определению единственна. Например, одна и та же форма платежного поручения на бумаге и на экране может выглядеть различно, при этом модель данных будет одна и та же. Возможно включение в состав основной модели отдельных частей других форм, т. е.

⁷⁾ В этом разделе использованы рабочие материалы, подготовленные В. В. Арлазаровым и Д. Л. Шоломовым

в форму можно включить, например, модель данных другой формы как часть модели данных текущей.

Основные задачи решаемые моделью содержания — описание структуры документа, проверки непротиворечивости, сохранение/получение данных. Модель содержания формы есть дерево именованных объектов данных, см. [29].

Модель формы разбивается на ряд слоев, в которых задаются те или иные логически связанные элементы и атрибуты модели. Слой ограничения данных обеспечивает целостность данных формы. Проверяется «правильность», не противоречивость данных, если данные не корректны — диагностируется ошибка, описываются ограничения на данные формы не типологического и структурного порядка — они наследуются от слоя «модель содержания». Слой не специфицирует, когда проверять те или иные правила или ограничения, время вызова задается в модели взаимодействия слоем «событийная модель».

В слое ограничения данных описываются все правила, необходимые для обеспечения целостности данных. Описание включает: формулы, вычисления, тривиальные логические ограничения типа (обязательность заполнения, алфавит, диапазоны, числа строк и символов, регулярные выражения, формат, автоматическое заполнение, описание словарей, релевантность и др.). Слой также включает стандартные и пользовательские правила проверки данных. Модель не накладывает ограничений на способ реализации тривиальных проверок, четко специфицируется только язык описания таких правил. Каждому правилу ставится в соответствие имя или имя типа для разделяемых правил. Правила могут задаваться в виде ссылок. Способы описания нестандартных пользовательских правил — интерпретируемые языки JavaScript 1.2 и VBScript, импорт из динамических библиотек, ActiveX объекты. Кроме того, в модели присутствуют элементы типа триггер (кнопка, и т. д.).

Имеется промежуточный слой между представлением об информационном объекте как о структурированном наборе характеристик с четкой типизацией и моделями представления средств обмена данными (генераторами отчетов, системами распознавания, экранными формами). В экземпляре формы значениями элементов будут либо строки, либо бинарные данные и ссылка на элемент модели. Для элемента должны быть определены его значение(я) и переменные характеризующие его состояние. Экземпляр слоя может быть сохранен в формате XML и загружен без потери данных.

6.4. Стандарт представления данных

Документ, введенный по одной из форм в одной из систем, должен быть узнаваем другой системой. Для этого необходимо кроме стан-

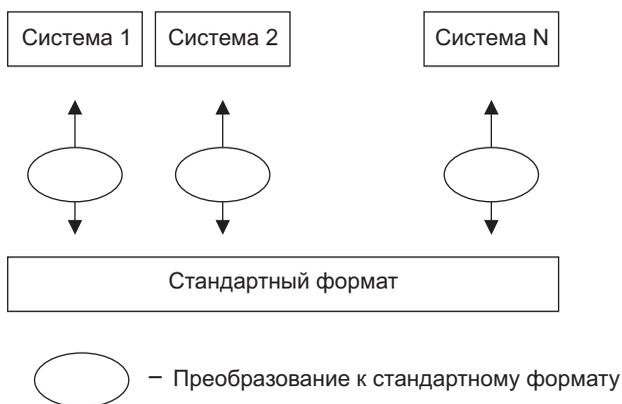


Рис. 15. Движение документов из одной системы в другую

дарта описания форм ввести стандарт на представление данных — экземпляров содержания документов, подготовленных по данной форме. Для передачи группы документов нужно один раз передать форму и экземпляры содержания.

На рис. 15 представлен возможный вариант решения проблем стыковки по данным разных систем. При наличии единого стандарта представления данных и передаче данных всегда в этом формате, каждой системе достаточно обеспечить только две функции преобразования форматов — из стандартного в свой внутренний и наоборот. Такой стандарт позволяет решить как внутренние проблемы передачи данных между подсистемами документооборота, так и взаимодействие с внешними системами. В частности, например, переход с одной платформы документооборота на другую может осуществляться экспортом данных из исходной системы документооборота в стандартный формат, а затем импортом в новый.

На практике подсистемы отличаются по многим характеристикам, и создать формат данных, с которым работали бы абсолютно все подсистемы и при этом использовались бы возможности их всех, практически невозможно. Поэтому подсистемы, работающие с данным представлением, можно разделить на три группы по уровню совместимости:

- Подсистемы, которые работают только со своей жесткой структурой. Им самим не нужна схема документа. Она де-факто «вшита» в программу. Для таких подсистем файл со схемой является постоянным. Они экспортируют данные в другие подсистемы, но импортируют только свои или сделанные другими специально для них.

- Подсистемы, которым все равно, с какими именно документами работать. Они обрабатывают документы со структурой, неизвестной в момент написания программы. Такие подсистемы удовлетворяют требованиям, указанным в предыдущем пункте, а также импортируют данные из любой другой подсистемы как первого, так и второго типа. Примерами таких подсистемы являются архивы.
- Подсистемы, которые способны изменять данные, созданные другими подсистемами, с сохранением схемы. Примером такой подсистемы является редактор форм.

Единый формат файлов [33] представляет собой подмножество XML-схем, что обеспечивает с одной стороны возможность работы с файлами в данном формате стандартным XML-инструментарием, а с другой стороны, упрощает разрабатываемые программы для импорта/экспорта структурированных данных в этом формате.

Что касается эффективности, то данный формат предназначен для организации передачи данных между различными информационными системами. Поэтому программы разбора будут иметь содержательную функциональность, и, как следствие, не самую высокую производительность. Следовательно, не удастся применять полновесный разборщик в приложениях реального времени, где от этого страдает общая производительность системы.

Файл в едином формате представляет собой пакет документов. С логической точки зрения пакет состоит из одного или нескольких структурированных документов. Полями документов могут быть, в том числе, и бинарные (неструктурированные) данные. Это позволяет хранить в пакете данные произвольной природы. С физической точки зрения пакет состоит из одного или нескольких файлов. Один из файлов является главным, остальные — присоединенными. Главный файл содержит основную информацию о пакете, присоединенные — документы или бинарные данные, если они не хранятся внутри главного файла.

Главный файл представляет собой сообщение в формате SOAP [34]. Содержимое SOAP-заголовка данный стандарт не регламентирует. Прикладные программы должны заполнять заголовок с учетом потребностей конкретной среды передачи данных. Тело пакета состоит из одного или нескольких документов. Каждый документ может содержаться непосредственно в сообщении, а может в присоединенном файле.

Экспортирующая программа может самостоятельно решать вопрос о распределении документов — во внешние файлы или в тело сообщения. Однако, размер сообщения и каждого из документов не должен превышать 10 Mb. Размер пакета в целом не ограничивается. Импортирующая программа может целиком загружать в память файл сообщения

или отдельный документ, например, средствами DOM (Document Object Model). Однако она не должна загружать целиком пакет, так как никаких гарантий о его размере импортирующая программа не имеет.

В пакет могут входить разные типы документов. Стандарт предусматривает несколько стандартных типов, прикладные программы могут вводить свои. Среди стандартных типов особо выделяется схема. Схема используется для проверки документов и для передачи информации об их структуре. Каждая схема имеет собственный глобально уникальный идентификатор. В качестве этого идентификатора используется URI.

Передача пакета осуществляется одним из двух способов:

- В виде файла с расширением .XFF (XML File Format). Этот файл представляет собой ZIP-архив, содержащий все файлы пакета.
- В виде директории.

Данный формат является ограниченно платформонезависимым. Это означает, что при отсутствии шифрования и электронно-цифровой подписи файл может быть записан на одной операционной системе, а прочитан на другой. Поддерживаются следующие операционные системы: WIN32, DOS, UNIX, OS/400, MacOS.

Формат основывается на следующих стандартах: XML, XML Schema, SOAP/XML Protocol, XML Encryption, XML Signing, XML Key Management.

6.5. Хранилище документов

Большая группа проблем возникает с организацией хранения введенных документов. Модель данных документов может иметь сложную структуру, которая естественным образом поддерживается моделями XML [23]. Главный вопрос — хранить ли документы как отдельные XML-файлы, или разносить значения полей документа в БД? Сейчас принято деление XML-баз данных на два класса: ориентированные на XML-документы в целом и ориентированные на данные, получаемые из XML-документов (см. [35, с. 27] и [36]). Каждое решение имеет достоинства и недостатки. XML-документы могут быть очень большими, и время их разборки по таблицам при записи в БД (и сборки при чтении) критичным. Ряд авторов сравнивает хранение документа в БД в разобранном виде с полной разборкой автомобиля при заезде в гараж и сборкой его при выезде из гаража. С другой стороны, чтение или корректировка отдельных данных, проведение аналитических исследований намного проще при хранении данных в БД. Отображению XML документов в реляционные и объектно-ориентированные БД посвящено много работ (см. монографию [35], в которой имеется обширный обзор литературы). Аналогично, при введении естественных ограничений на спецификации XML достигается взаимно однозначное отображение

данных в СУБД НИКА [37]. Возможно и смешанное решение, когда документ отображается как в БД, так и в файл (архив документов) (см. [27]). При этом в архиве XML документов хранятся все версии документов, а в БД только последнее актуальное значение данных. Архив документов является, по существу, протоколом всех изменений, вносимых в БД; по нему возможно восстановление БД на любое заданное время.

В Хранилище документов (независимо от метода его реализации) на логическом уровне выделяются *коллекции документов* и *документы*, которые имеют иерархическую структуру [27]. Для этого в схеме БД выделяются вершины (обычно массивы), соответствующие коллекциям, и вершины (обычно структуры, подчиненные массивам), соответствующие корневым элементам документов. Структура — корень документа выделяет в схеме БД поддерево, которое и отождествляется с моделью данных документа. Каждому типу документов ставится в соответствие набор форм для просмотра и редактирования данных.

Редактор позволяет создавать коллекции для хранения документов. Пользователь задает тип новых документов и определяет набор форм, соответствующих данному типу; при этом автоматически модифицируется схема хранилища. После чего редактор позволяет создавать, редактировать и осуществлять экспорт/импорт документов нового типа.

Редактор не зависит от конкретного типа БД, работая с данными документов через интерфейс источника структурированных данных и с хранилищем через интерфейсы объектной модели хранилища документов. Реализация такого интерфейса называется драйвером.

Реализации объектной модели хранилища документов делятся на локальные и серверные и обладают следующими свойствами:

- Все данные хранилища доступны через интерфейсы *коллекции* и *документа*.
- У хранилища есть схема.
- Схема хранилища и документа является реализацией абстрактной модели данных на разных источниках (РСУБД, СУБД НИКА, XML, директория файлов, Евфрат [28], Cognitive Forms [42]).
- Объект хранилища позволяет получать схему хранилища в формате XML Schema [38] или в формате модели содержания форм.
- Некоторые хранилища позволяют обновлять свою схему, используется формат описания модели содержания форм или формат XML Schema.
- Хранилище может организовывать документы в списки. Простейшим списком является коллекция.
- Списки могут быть получены через иерархию данных, индекс или поиск.

- Хранилище может обладать индексом. Индекс позволяет получить список документов (с возможностью фильтрации по типу документов) по указанным имени и значению индексного атрибута.
- Каждый документ хранилища имеет тип.
- Хранилище позволяет искать документы. Запросы к хранилищу должны быть оформлены на основе языка запросов XQuery [39].

В работе [27] описан драйвер над СУБД НИКА. В качестве формата обмена данными используется XML. Для использования других СУБД требуется реализовать интерфейсы объектной модели хранилища документов, что поддержано основными производителями СУБД реализацией экспорта/импорта XML-данных.

Следует отметить, что хранилище данных, реализованное в системе [27], можно отнести к разряду *XML-совместимых СУБД* (XML-Enabled Databases), в то же время оно имеет много общего с *естественными XML СУБД* (Native XML Databases) [36].

6.6. Индексация XML документов

Ряд авторов отмечает трудности индексации XML-документов (см. обзор в [35]). М. Грейвс пишет (см. [35]. С. 465): «Структура данных XML отличается большей гибкостью по сравнению со структурой данных в реляционных или объектно-ориентированных БД, поэтому для нее должна применяться иная стратегия индексации. Структура элемента может изменяться, поэтому подход, предусматривающий применение статической структуры индексации для элемента, является неосуществимым».

Возникают следующие проблемы. Базы данных XML будут хранить огромные объемы информации; какие методы использовать для эффективного поиска, которые позволяли бы исключить полный перебор? Существует ли возможность эффективного создания междокументных связей в одной БД и в нескольких БД? Как быть с быстроменяющимся источником данных, когда поддержание индекса в актуальном состоянии требует больших ресурсов? Когда применять полную и частичную индексацию? Отмечается, что полная индексация в наибольшей степени приемлема, если часто происходит доступ к относительно небольшой и неизменной БД, но где грань ее применения?

М. Грейвс далее пишет: «Хранение информации о ссылках, выходящих за границы документа, может стать более эффективным, если в индексах будет представлен только документ, на который указывает ссылка, а не фрагмент этого документа. В таком случае можно выполнять поиск соответствующего фрагмента в документе, на который указывает индекс. В результате появляется возможность свертывать и разворачивать

внутреннюю структуру документа, а также вносить в нее необходимые изменения без корректировки междокументных индексов» (см. [35, с. 466]). В статьях [40, 41] обсуждается построение индекса и система запроса, которые частично решают перечисленные проблемы.

6.7. Язык запросов

Для поиска нужных XML-документов консорциум W3C разрабатывает язык XML Path Language (XPath) [43]. Язык XPath 1.0 реализован рядом производителей: Microsoft XML Parser 4.0 [44], XSLT процессор Xalan компании Apache [45] и др. В настоящий момент на сайте консорциума опубликована версия XPath 2.0 от 15 ноября 2002, которая имеет статус рабочего варианта. В октябре 1999 года консорциум W3C образовал рабочую группу XML Query Working Group [46], куда вошли представители от известных крупных фирм программного обеспечения (Microsoft, IBM, Oracle и др.), с целью разработки языка запросов, получившего название XQuery [39]. В настоящий момент опубликована рабочая версия языка XQuery 1.0 от 4 апреля 2005 г. Язык XPath 2.0 стал подмножеством языка XQuery 1.0. Многие производители систем хранения данных уже создали свои трансляторы и процессоры для обработки запроса, написанного на языке XQuery (Microsoft, Oracle, SoftWare AG), несмотря на то, что язык имеет пока только рабочую версию. В [40, 41] рассматривается реализация запросной системы для хранилища сложно-структурированных данных на основе языка XPath и СУБД НИКА, описаны принципы построения индекса, функции работы с индексом и поиск в СУБД НИКА.

Следует отметить, что синтаксис языка XPath очень лаконичен, строго формализован и труден для понимания конечными пользователями. В работе [41] предложен способ реализации запросной системы над хранилищем данных СУБД НИКА с интерфейсом для конечного пользователя. При этом автоматически условие запроса записывается в нотации языка XPath для работы и сохранения запросов. Конечный пользователь может написать сложные условия запроса над хранилищем данных в нотации языка XPath без знания его синтаксиса. Схема хранилища отображается в виде наглядного дерева составных и простых элементов, условия на составные элементы представляются в табличном виде, при этом условия могут содержать сложные скобочные выражения и вложенные запросы и т. д. Диалог формирования запроса не имеет жесткой привязки к СУБД и может быть использован для работы над другими хранилищами данных, имеющими в своей основе XML базу данных.

6.8. Генерация отчетов

Процедуры ввода/вывода документов можно разбить на три основных этапа.

Ввод документа:

- расформатирование (выделение содержания документа по описанию его формы, проверка выполнения правил заполнения),
- отображения структуры содержания документа в структуру БД (при этом может потребоваться реструктуризация данных и выполнение расчетов),
- загрузка в БД.

1-й этап выполняет человек, вводящий документ, или программа распознавания,

2-й этап специальные функции системы ввода или функции СУБД,

3-й этап функция СУБД.

Вывод документа — генерация отчетов (задача обратная вводу):

- выборка по запросу нужных объектов из БД,
- отображение структуры объектов (БД) в структуру выходного документа,
- собственно вывод документа на экран или печать.

1-й этап функция СУБД, 2-й и 3-й — обычно относятся к программным средствам, называемым генераторами отчетов.

Таким образом, генерация отчетов тесно связана с языком запросов. Существующие генераторы (наиболее известный из них — Crystal Report [47]) используют язык запросов SQL и ODBC интерфейс с БД. При работе со сложно структурированными XML документами и XML базами данных громоздко и не удобно выписываются иерархические и ссылочные связи, нелеп также в SQL результат выполнения запроса — всегда двумерная таблица, когда хочется получить его в XML формате, как и исходные структурированные документы. Для XML баз данных сейчас активно развивается язык XQuery [39]. Мы отмечали создание первых трансляторов и процессоров обработки запросов, написанных на языке XQuery, что подчеркивает актуальность развития и стандартизации языка.

Другая проблема создания генератора — желание ядром его иметь Word, чтобы получать отчеты сразу в формате редактора, ставшего международным стандартом, а также поддерживающего, начиная с версии Word 11.0, XML документы [48].

6.9. Средства аналитических исследований

Простейшим средством аналитических исследований информации является генератор отчетов, если он предоставляет гибкие, ориентированные на конечного пользователя [49], средства запроса к информационному базису и проведения расчетов. Но обеспечение конечного пользователя аналитикой типа OLAP (или еще более функциональной [50])

над информационным базисом в виде XML баз данных выдвигает много новых проблем.

Система, построенная над СУБД НИКА [51], предоставляет возможность произвести некоторый срез БД, как результат запроса, выбрать объект для анализа (тип XML документа), выбрать оси для построения графиков и функции расчета значений. Эта система позволяет конечному пользователю производить большой класс аналитических исследований, но в случае более сложных запросов и расчетов, например, над несколькими взаимосвязанными типами документов, необходимо привлекать квалифицированного программиста.

7. Системы обработки документов. Основные компоненты

Термин «система обработки документов» не соответствует на сегодня никакому общепринятому содержанию. Одни авторы понимают под ними СУБД с минимальным набором средств ввода и выгрузки данных, другие — системы рассылки (передачи) документов и контроля их исполнения. Однако исключать из рассмотрения даже в этих простейших случаях, например, редакторы крайне рискованно хотя бы потому, что никакие драйверы и форматы (будь то DBF, ODBC, XML или что-нибудь другое) не решают важнейших проблем целостности и (или) безопасности данных. Эти проблемы могут быть решены только на уровне базы данных, но должны проверяться при редактировании. Таким образом, отсутствие глубоких связей между СУБД и редактором (ни за что, ни про что) перекладывает большую часть работы на прикладного программиста. Аналогично обстоит дело и с генерацией выходных документов.

Однако у вопроса содержания термина есть еще одна сторона. В одном из описаний самых крупных и распространенных систем обработки документов [52–55] читаем: «У нас в стране⁸⁾ можно встретить абсолютно не поддающиеся здравой логике ситуации. Заказчик использует созданные приложения, поддерживающие сложные процессы подготовки, работы с документами и контроля их прохождения. И вот, когда эти сложные процессы завершены, документ готов, всеми утвержден (так и хочется сказать, „почти мертв“), его регистрируют в отдельной системе, созданной часто на дорогостоящей РСУБД, которая состоит из более чем 30 связанных между собой таблиц»⁹⁾.

⁸⁾ Авторы имеют в виду Россию.

⁹⁾ См. Системы электронного управления документами: обзор, классификация и оценка возврата от внедрения (<http://www.lotus.ru> — на главной странице, без указания автора — как передовая).

Авторы, очевидно, за деревьями не видят леса. Каждый проверенный и утверждённый документ содержит бесценную порцию знаний о своей предметной области и/или объекте управления. Поэтому не удивительно, что в реальных прикладных системах обработки документов, как правило, необходимы средства архивации и извлечения этой находящейся в документах информации. Если же системы обработки документов таких средств не предоставляют, то разработчики вынуждены их создавать сами.

Мы будем исходить из целостного понимания процесса обработки документов. В статье рассматриваются основные средства сбора, хранения и гибкого использования не только самих документов но и извлекаемых из них знаний. Один из лозунгов подхода: «Создание хранилища знаний о тех информационных объектах, которые описаны в документах». Наряду с обычными компонентами систем управления документами, рассматриваются также системы массовой загрузки потоков документов, автоматической рубрикации, создания БД содержания документов, системы типа OLAP, классификаторы, сложные формы документов и другое, перечисленное ниже.

Ссылки на литературу в статье не претендуют на полноту. Они служат иллюстрацией предлагаемой классификации средств, поэтому в основном использованы работы, опубликованные в серии сборников трудов ИСА РАН «Интеллектуальные проблемы ввода и обработки информации» [56], «Развитие безбумажной технологии в организационных системах» [57], «Методы и средства обработки документов» [58] и настоящий сборник.

7.1. Классификация систем обработки документов

Некоторые авторы (в частности, см. [52]) предлагают такую классификацию: *Делопроизводство, Документооборот, Управление документами, Управление знаниями*. При этом не вводится ясных критериев деления на эти классы (особенно туманны 3-й и 4-й), только отмечается, что каждый следующий включает все предыдущие.

Мы попытаемся более конкретно определить эти классы:

* *Делопроизводство* — регистрация документов (бумажных и/или электронных как таковых, вообще говоря, без регистрации их содержания — только ввод краткой неструктурированной аннотации), ввод их в систему (сканирование бумажных), задание поручений по ним и контроль исполнения.

** *Документооборот* — средства первого класса и регистрация содержания документов, выполнение бизнес-процессов, связанных с этим документом, маршрутизация.

***** Управление документами** — средства первых двух классов и средства массовой загрузки документов, архивы, хранение на CD, DVD, MO, разнообразные средства вывода и т. п.

****** Управление знаниями** — средства первых трех классов и средства рубрикации, классификаторы, информационно — аналитические средства (OLAP)¹⁰.

Основные компоненты систем обработки документов

В этом разделе дается перечень основных компонент систем обработки документов в порядке прохождения их жизненного цикла. Эти компоненты помечены одной, двумя, тремя или четырьмя звездочками в соответствии с тем классом систем обработки документов, которому, на наш взгляд, они наиболее свойственны.

Ввод собственно документов

— бумажных и/или электронных (частей-страниц в терминологии системы Евфрат [55]):

- Средства сканирования бумажных документов (*).
- Средства ввода из Internet (on-line, off-line) (*).
- Средства ввода из Windows приложений (*).

Регистрация по определенным формам

содержания документов (вообще говоря, сколь угодно сложной структуры) [59]:

- Средства описания форм сложно структурированных документов (*) [60].
- Редактор документов на основе их форм (**).
- Средства генерации новой (расширение имеющейся) схемы БД информационных объектов по описаниям форм входных документов и, наоборот, построение формы ввода на основе фрагмента (или всей) схемы БД (**).
- Средства ввода данных, представленных по формам, в БД (набор драйверов для разных СУБД) (**).

Распознавание [61, 62]

- Средства распознавания всего документа (*).
- Распознавание выделенного фрагмента во время переноса атрибута — (Drag&Recog) (*).

¹⁰ Надо отметить, что некоторые компоненты класса (**) имеют смысл и без (*), (***) — без (**), (****) — без (*) и (**).

- Средство распознавания блока атрибутов, что позволяет в образе документа указать только место расположения группы атрибутов (индексация по имиджу, разметка имиджем) (**).

Массовая загрузка данных [63, 64]

- Средства проектирования форм для автоматического ввода и распознавания (***)
- Специальный редактор для корректировки ошибок распознавания (***)
- Система обеспечения конвейера сканирования, распознавания, корректировки, ввода в БД (***)
- Средства разбора и загрузки данных из новостных лент (***)
- Средства разбора (полностью или изменений) Internet сайтов (***)
- Средства экспорта данных в архив (***)

Индексация

- Система индексации по атрибутам документа (*)
- Средства полнотекстовой индексации, стоп словари (*)
- Система автоматической рубрикации (**)
- Система индексации по атрибутам объектов содержания документов (**) [65]
- Автоматическая индексация дат, географических наименований, имен и фамилий, e-mail и http — адресов и т. п. (***)

Лингвистический анализ [66, 67]

Эти компоненты, как правило, используются в средствах индексации и поиска.

- Морфологический и семантический анализ, средства нормализации (приведение к единственному числу, именительному падежу и т. п.) (*)
- Работа с другими языками, не только русским (*)
- Автоматическое выделение понятий (****)
- Элементарные средства перевода с одного языка на другой (например, английский—русский) (****)

Хранилище документов и описаний информационных объектов

- Хранилище документов (файл-сервер, документно-ориентированная или универсальная БД, см., например, [52, 68]) (*)
- Хранилище информационных объектов, описанных в документах, (содержание документов) — универсальная БД на промышленной СУБД, (а не документоориентированной) (**)

- Средства хранения документов на CD, DVD, MO (***)
- Средства сжатия данных (***)

Поиск

- Средства поиска по любому логическому выражению (с указанием скобок, фрагментов контекста и др.) (*)
- Средства полнотекстового поиска (в частности, с учетом близости слов в документе) (*)
- Средства комбинированного поиска (по всем словам и атрибутам) (*)
- Средства сохранения типовых запросов (*)
- Средства оценки релевантности и уточнения запроса (*)
- Средства автоматизации предложений к уточнению запроса (****)
- Система поиска объектов и подобъектов любой глубины вложенности по собственным и унаследованным атрибутам (****) [65].

Экспорт/импорт данных

- Экспорт/импорт в скрипт (специализированный, например, скрипт системы Евфрат [55]) или универсальный XML-документ (*) [69].
- Экспорт в объектно-ориентированную базу данных (ООБД). Импорт из ООБД (**)
- Экспорт в реляционную или объектно-реляционную базу данных (ОРБД). Импорт из реляционной или ОРБД (**)
- Экспорт/импорт в форматы HTML, PDF, PostScript и др. пригодные для публикации в Internet или на бумаге (**)

Web сервер над архивом документов

(для обеспечения работы с документами средствами стандартных браузеров) [70].

- Средства обеспечения доступа к архиву документов в Интернет (*)
- Средства включения файлов как в качестве приложений ко всему документу, так и компонент описания объектов документа (фотографии, чертежи, тексты, звук и т. п.) (**)
- Средства выделения подмножеств атрибутов (подсхем БД) доступных для разных типов пользователей (***)
- Средства гибкого оперативного управление режимом отображения информационных объектов (списки, таблицы, иерархии) (****)

Обмен сообщениями

(электронная почта, сокет, папки сообщений) (см., например, [52], теперь эти средства стали общими для систем обработки документов).

- Средства передачи сообщений (*).
- Хранилище сообщений (*).
- Адресная книга (*).

Репликации

- Слежение за версиями и копиями (***) [52, 71].
- Объединение версий (специальный редактор, «автоматическое» и др.) (***)
- Совместная разработка частей документа группой пользователей (***) [60].
- Репликация для не всегда подключенных, удаленных пользователей (***)
- Выборочная репликация (***)
- Фоновая репликация (репликация должна не прекращать другие работы) (***)
- Синхронизация с логикой приложений и дизайном (пересылаются не только данные но и изменения в логике и дизайне) (***)

Работа в сетях

- Локальных с малым числом пользователей (до 10 пользователей) (*)
- Локальных (глобальных) — с большим числом пользователей (> 100).
- Средства обмена между клиентами и сервером скриптами или XML документами (*)
- Монитор транзакций (*)
- Сервер приложений (*)

Маршрутизация исполнительской деятельности [72]

- Средства описания поручений (**)
- Средства задание маршрута (**)
- Средства определения текущего состояния по поручениям (**)
- Средства контроля исполнительской деятельности (**)
- Средства отображения истории обработки поручений и документов (**)

Защита от несанкционированного доступа [73]

- Система идентификаторов и паролей (*).
- Средства описания групп пользователей (*).
- Система разграничения доступа к объектам и функциям (*).
- Средства протоколирования всех операций (ввод, обработка, корректировка, чтения, распечатки документов и др. объектов) (*).
- Средства криптографии (*).
- Средства электронной подписи (*).
- Средства администратора безопасности (**).

Вывод документов

- Генератор отчетов (например, на основе Crystal Report) (*) [74].
- Вывод документов по формам ввода (***)
- Печать форм массового ввода, заполненных полностью, частично или пустых, в виде готовом для ввода и автоматического распознавания (***)
- Печать каталогов, справочников на основе издательских систем (***)

Публикация информации на CD в виде отчуждаемого продукта

- Обеспечение работы средствами стандартных браузеров (*).
- Защита от несанкционированного копирования дисков целиком (***)
- Защита от копирования отдельных объектов, образов документов (***)

Средства разработки приложений. API Интерфейсы

- API интерфейсы ко всем объектам системы (**).
- Интерфейсы ко всем перечисленным выше компонентам системы (**).
- Средства поддержки стандартных интерфейсов систем управления документами (ODMA, DMA, CORBA и др.) (***)

Средства управления знаниями

- Управление голосом (****) [75].
- Классификаторы, тезаурусы (****).
- Средства ввода, обработки, хранения, поиска и отображения сложно структурированных знаний (****) [77].
- Системы автоматического перевода (****).
- Системы типа OLAP (****).

7.2. Спектр решаемых задач

Спектр задач, решаемых обсуждаемыми в этой статье средствами весьма широк. Мы их так же при помощи «звездочек» относим к одному из четырех классов систем обработки документов.

- Делопроизводство (*) — регистрация документов, задания по ним, контроль исполнения.
- Справочники папок и файлов на внешних носителях (флорпи дисках, CD, DVD, MO и др.) (*).
- Электронные фотоальбомы (*).
- Документооборот (**) — поддержка процессов на основе данных из документов [76].
 - Обработка бухгалтерских документов (платежные поручения, счета, счета-фактуры, договора, акты, накладные и др.) (**) [77].
 - Электронная коммерция (**).
 - Учет и обработка документов материально-технического снабжения (**).
 - Управление кадрами (**).
 - Системы руководства проектами — планы, отчеты, задания (**).
- Библиографические системы для библиотек (**).
- Системы учет музейных экспонатов (**).
- Большие архивы документов (***) [78].
- Системы обработки документов налоговых инспекций и пенсионных фондов (***)).
- Информационно-аналитические системы (****).

Никогда все перечисленные выше компоненты не объединяются в один программный продукт (например, в одну коробку). Обычно в продуктах фирм есть ядро (может быть близкое, например, к LN [52] или Евфрату [55], и которое может быть коробкой) и набор средств окружения, которые имеют согласованный и, по возможности, единый интерфейс.

8. Заключение

Суть предлагаемого подхода к системам документооборота кратко можно сформулировать, как предоставление пользователю средств описания содержания документов, а не только факта их существования, как «глухих бумаг», которые идентифицируются Регистрационной Карточкой (от кого, кому, дата, адрес отправителя и т. д.).

Обращаясь к такому подходу, мы попадаем в область систем управления контентом (Content Management), которые строятся на универсаль-

ных СУБД. Если мы хотим предоставить пользователю (менеджеру, врачу, учителю) возможность самому создавать системы для описания содержания своих документов (объектов, людей, историй болезней и т. п.), то должны предоставить ему средства описания форм ввода и корректировки данных, генератора отчетов, языка запросов (типа XQuery, если данные сложно структурированные), средства аналитики (типа OLAP) и т. д.

То есть, предоставить пользователю все средства функционально полной СУБД, наряду со стандартными средствами систем делопроизводства. Добиться сбалансированного комплекса возможностей и доступности их использования пользователями непрограммистами — интересная, но не простая задача.

Мы не обсуждаем в этой статье круг проблем, связанных с целостностью и защитой данных, распределенным доступом, хотя понижение уровня квалификации администраторов систем документооборота и баз данных очень актуален. Этим проблемам посвящено много работ и они решаются средствами тех СУБД, на которых строится документооборот (универсальных промышленных или специализированных «своих»). Обзор возникающих здесь проблем требует отдельного рассмотрения.

Литература

1. Козлов М. Совершенствуем Делопроизводство // Cognitive Technologies. М., 1995. 124 с.
2. Арлазаров В. Л., Емельянов Н. Е. Документооборот или управление знаниями? // Организационное управление и искусственный интеллект / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2003. С. 6–29.
3. Документооборот. Прикладные аспекты / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2004. С. 183.
4. Евфрат—Документооборот // <http://www.cognitive.ru/products/euph-doc.htm>
5. Богданов А. С., Емельянов Н. Е., Ерохин В. И., Скорняков В. А., Романов Б. Л. НИКА — технология построения информационных систем // Организационное управление и искусственный интеллект / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2003. С. 52–67.
6. Акимова Г. П., Пашкин М. А., Славин О. А. Специфика документооборота электронной торговли // Документооборот. Прикладные аспекты / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2004. С. 12–19.
7. Боженов А. Ю. Концепция построения системы «Информационный фонд» // Документооборот. Прикладные аспекты / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2004. С. 12–19.

- ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2004. С. 20–49.
8. *Романов А. Н., Славин О. А.* Техническое проектирование общероссийской системы информационно-маркетинговых центров // Документооборот. Прикладные аспекты / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2004. С. 108–125.
 9. *Даниленко А. Ю., Минкин Ю. И.* Анализ основных принципов построения и особенностей защиты информации в системах электронного документооборота // Документооборот. Концепции и инструментарий / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2004. С. 31–41.
 10. *Емельянов Н. Е., Ерохин В. И.* Особенности систем, создаваемых на основе НИКА-Технологии // Документооборот. Концепции и инструментарий / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2004. С. 42–57.
 11. *Карпенко С. М., Кузнецов А. В., Куратов П. А. и др.* Система электронного документооборота и управления банковскими технологиями // Документооборот. Прикладные аспекты / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2004. С. 61–77.
 12. *Глинников М.* На защите интеллектуальной собственности // Директор ИС. 2004. № 5. С. 12–16.
 13. *Славин О. А., Федоров Г. О.* Об использовании штрих-кодирования и специализированных устройств в корпоративном электронном документообороте // Организационное управление и искусственный интеллект / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2003. С. 185–197.
 14. *Елагин В. В., Зимовцев В. Н.* Об оптимизации информационных потоков, содержащих конфиденциальную информацию, в информационно-управляющей системе и использовании интеллектуального потенциала организации при решении задач // Документооборот. Прикладные аспекты / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2004. С. 12–19.
 15. *Першиков В. И., Савинков В. М.* Толковый словарь по информатике // Ф и С. 1991. 543 с.
 16. Федеральный закон от 06.05.1999 № 305 «О конкурсах на размещение заказов на поставки товаров, выполнение работ, оказание услуг для государственных нужд».
 17. Указ Президента от 08.04.1997 № 305 «О первоочередных мерах по предотвращению коррупции и сокращению бюджетных расходов при организации закупки продукции для государственных нужд».
 18. *Арлазаров В., Безмозгий И., Емельянов Н.* Документы — основа автоматизированной банковской системы. В сборнике: Методы и средства работы с документами. М.: УРСС, 2000. С. 3–14.
 19. *Арлазаров В. Л., Емельянов Н. Е.* Системы обработки документов // Управление информационными потоками / Сборник трудов ИСА РАН. Под ред.

- д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2002. С. 3–20.
20. Колесов А. Наука управления документами // ВУТЕ. 2003. № 2. С. 30–33.
 21. Смирнов А., Криворученко И., Криворученко В. Электронный документооборот: еще раз о системном подходе к терминологии // PC Week/RE. 2004. № 18. С. 41–43.
 22. Кантор Н. MuseumPlus: от управления коллекцией до комплексной системы управления музеем // Материалы конференции «Музей в XXI веке». СПб., 18–21 мая 2004.
 23. XML (*Extensible Markup Language 1.0 (Second Edition)*). Рекомендация W3C. 6 октября 2000 г. <http://www.w3.org/TR/REC-xml>.
 24. Udell J. XML for the rest of us // http://www.infoworld.com/article/02/11/15/021118plmsxml_1.html
 25. Постникова М., Славин О. // Организационное управление и искусственный интеллект / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2003. С. 30–51.
 26. Колесов А. Наука управления документами // ВУТЕ № 2, 2003. С. 30–33.
 27. Богданов А., Емельянов Н., Ерохин В., Скорняков В., Романов Б. НИКА технология построения информационных систем // Организационное управление и искусственный интеллект / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2003. С. 52–67.
 28. Евфрат—документооборот. <http://www.cognitive.ru/products/euph-doc.htm>
 29. Романов Б. Л. Представление структурированных информационных объектов в виде электронных форм // Управление информационными потоками: Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н., проф. Емельянова Н. Е. М.: УРСС, 2002. С. 306–311.
 30. Богачева А. Н., Емельянов Н. Е. Семантическая модель документа // Системные исследования. Ежегодник 2001. М.: УРСС, 2003. С. 360–375.
 31. XForms 1.0 // W3C Working Draft. 18 January 2002.
 32. Ввод платежных документов. http://www.cognitive.ru/products/forms_vpp.htm
 33. Порай Д. С. Представление документов в формате XML // Управление информационными потоками: Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н., проф. Емельянова Н. Е. М.: УРСС, 2002. С. 278–289. См. также: <http://www.cognitive.ru/innovation/format1.htm>
 34. SOAP. <http://bodq.org.ua/w3c/www.w3.org/TR/soap12-part1/>
 35. Грейвс М. Проектирование баз данных на основе XML. М.: Издательский дом «Вильямс», 2002. 640 с.
 36. Bourret R. XML and Databases // <http://www.rpbourret.com/xml/XMLAndDatabases.htm>
 37. Годунов А., Емельянов Н., Космынин А., Солдатов В. Система НИКА // Системы управления базами данных и знаний / Под ред. Наумова А. Н. М.: Финансы и статистика, 1991. С. 209–248.

38. XML Schema — <http://www.w3.org/XML/Schema>
39. XQuery 1.0 and XPath 2.0 (XQuery 1.0 and XPath 2.0 Data Mode). Рабочий проект W3C. 4 апреля 2005 г. <http://www.w3.org/TR/query-datamodel/>
40. *Богданов А., Емельянов Н., Ерохин В.* Индексация и поиск объектов в СУБД НИКА // Управление информационными потоками / Сборник трудов ИСА РАН под ред. д. т. н. проф. Арлазарова В. Л., д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2002. С. 69–84.
41. *Богданов А., Емельянов Н., Ерохин В., Романов Б.* Реализация запросной системы на основе языка XPath для СУБД НИКА // Организационное управление и искусственный интеллект / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2003. С. 132–148.
42. *Арлазаров В., Постников В., Шоломов Д.* Cognitive Forms — система массового ввода структурированных документов // Управление информационными потоками. Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н., проф. Емельянова Н. Е. М.: УРСС, 2002. С. 35–46.
43. XPath (XML Path Language (XPath) 2.0). Рабочий проект W3C. 15 ноября 2002 г. <http://www.w3.org/TR/xpath.html>.
44. Microsoft XML Core Services (MSXML) 4.0. SDK. http://msdn.microsoft.com/library/default.asp?url=/library/en-us/xmlsdk/hm/sdk_intro_6g53.asp
45. Xalan: XSL stylesheet processors in Java & C++. <http://xml.apache.org/xerces>
46. World Wide Web Consortium. XML Query Working Group. <http://www.w3.org/XML/Query>.
47. *Пек Дж.* Seagate Crystal Reports 8. Полное справочное руководство // Seagate. 2002. 420 с.
48. *Соловьев А. В.* Генератор отчетов для систем электронного документооборота // Организационное управление и искусственный интеллект / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2003. С. 109–111.
49. *Емельянов Н., Соловьев А., Соловьев Д.* Средство конечного пользователя для генерации документов по базе данных // Методы и средства работы с документами: Сборник трудов ИСА РАН / Под ред. Д. т. н. проф. Арлазарова В. Л. и д. т. н., проф. Емельянова Н. Е. М.: УРСС, 2000. С. 183–203.
50. *Логинов А. С.* Некоторые разделы теории информации и их приложения к задачам прогнозирования и распознавания // Организационное управление и искусственный интеллект / Сборник трудов ИСА РАН. Под ред. члена-корр. РАН Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2003. С. 198–211.
51. *Емельянов Н. Е., Ерохин В. И.* Оперативный анализ данных СУБД НИКА // Управление информационными потоками. Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н., проф. Емельянова Н. Е. М.: УРСС, 2002. С. 163–173.
52. Lotus. См. <http://www.lotus.com/home.nsf/welcome/russia>
53. DOCS OPEN. См. <http://www.hummingbird.ru/product/index.asp>

54. Documentum 4i. См. <http://www.documentum.ru/docum4i.html>
55. Евфрат. См. <http://www.cognitive.ru/products/euphrates/index.shtml>
56. Интеллектуальные технологии ввода и обработки информации // Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 1998. 164 с.
57. Развитие безбумажной технологии в организационных системах // Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 1999. 384 с.
58. Методы и средства работы с документами // Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2000. 376 с.
59. Романов Б. Л. Представление структурированных информационных объектов в виде электронных форм // Управление информационными потоками. Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н., проф. Емельянова Н. Е. М.: УРСС, 2002. С. 306–311.
60. Романов Б. Л. Средства разработки форм документов // Методы и средства работы с документами // Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2000. С. 304–311.
61. Арлазаров В. Л., Астахов А. Д., Троянker В. В., Котович Н. В. Адаптивное распознавание символов // Интеллектуальные технологии ввода и обработки информации: Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 1998. С. 39–58.
62. Арлазаров В. Л., Крутов П. А., Славин О. А. Распознавание строк печатных символов // Методы и средства работы с документами // Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2000. С. 31–51.
63. Постников В. В. Разработка методов наложения форм на графическое изображение объектов // Интеллектуальные технологии ввода и обработки информации: Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 1998. С. 157–163.
64. Арлазаров В. В. Управление информационными потоками в системе автоматического ввода документов // Управление информационными потоками. Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н., проф. Емельянова Н. Е. М.: УРСС, 2002. С. 21–34.
65. Богданов А. С., Емельянов Н. Е., Ерохин В. И. Индексация и поиск объектов в СУБД НИКА // Управление информационными потоками. Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н., проф. Емельянова Н. Е. М.: УРСС, 2002. С. 69–84.
66. Бузикашвили Н. Е., Самойлов Д. В., Бродский Л. И., Усков А. В. Задача поиска в неструктурированном тексте и лингвистический анализ // Интеллектуальные технологии ввода и обработки информации // Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 1998. С. 96–104.
67. Бузикашвили Н. Е., Оберляйтнер М. С., Усков А. В. Выявление частеречевых N-грамм русского языка по неразмеченным текстам // Методы и средства

- работы с документами: Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2000. С. 154–166.
68. *Годунов А. Н., Емельянов Н. Е., Космынин А. Н., Солдатов В. А.* Система НИКА // Системы управления базами данных и знаний / М.: Финансы и статистика, 1991. С. 209–249.
 69. *Порай Д. С.* Представление документов в формате XML // Управление информационными потоками. Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н., проф. Емельянова Н. Е. М.: УРСС, 2002. С. 278–289.
 70. *Емельянов Н. Е., Тищенко В. А.* Методы отображения объектов для построения Web-сервера объектно-ориентированной базы данных // Развитие безбумажной технологии в организационных системах: Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 1999. С. 96–109.
 71. *Плискин Е. Л.* Управление версиями в системах коллективного создания документов // Развитие безбумажной технологии в организационных системах // Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 1999. С. 214–251.
 72. Стандарты Work Flow. См. <http://www.wfmc.org>
 73. *Бардин М. Д., Емельянов Н. Е., Соловьев Д. В.* Защита данных в системе AS/400 с сетью персональных машин // Интеллектуальные технологии ввода и обработки информации // Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 1998. С. 57–73.
 74. *Емельянов Н. Е., Соловьев А. В., Соловьев Д. В.* Средство конечного пользователя для генерации документов по базам данных // Методы и средства работы с документами // Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2000. С. 183–203.
 75. *Богданов Д. С., Брухтий А. В., Подрабинович А. Я., Усков А. В.* Язык описания сценария диалога для речевого управления // Развитие безбумажной технологии в организационных системах: Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 1999. С. 47–61.
 76. *Арлазаров В. Л., Безмозгий И. М., Емельянов Н. Е.* Документы — основа автоматизированной банковской системы // Методы и средства работы с документами: Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2000. С. 3–14.
 77. *Арлазаров Е. В., Долгоруков А. Ю., Емельянов Н. Е. и др.* Построение системы внутрибанковской бухгалтерии на основе базы данных финансовых документов // Методы и средства работы с документами: Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 2000. С. 15–31.
 78. *Акимова Г. П., Дега В. Н., Емельянов Н. Е., Порай Д. С.* Об одном подходе к построению согласованного ведения фактографической базы данных и базы изображений // Интеллектуальные технологии ввода и обработки информации: Сборник трудов ИСА РАН / Под ред. д. т. н. проф. Арлазарова В. Л. и д. т. н. проф. Емельянова Н. Е. М.: УРСС, 1998. С. 4–14.