

На правах рукописи

ЯФАЕВ Виль Эмарович

**XML-ТЕХНОЛОГИЯ СОЗДАНИЯ НА WEB-СЕРВЕРЕ
ПЕРСОНАЛИЗОВАННЫХ WORD-ДОКУМЕНТОВ
НА ОСНОВЕ XSL-ТРАНСФОРМАЦИИ**

**Специальность 05.13.11 – Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей**

**АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук**

Уфа - 2009

Работа выполнена в ГОУ ВПО
«Уфимский государственный авиационный технический университет»
на кафедре автоматизированных систем управления

Научный руководитель	д-р техн. наук, проф. МИРОНОВ Валерий Викторович
Официальные оппоненты	д-р техн. наук, проф. МАРТЫНОВ Виталий Владимирович зав. каф. экономической информатики УГАТУ канд. техн. наук, АЛИМБЕКОВА Софья Робертовна нач. отдела управления проектами НИИ техниче- ских систем «Пилот»
Ведущая организация	ФГУП уфимское научно-производственное пред- приятие «Молния»

Защита состоится «25» декабря 2009 г. в 15:00 часов
на заседании диссертационного совета Д-212.288.07
при Уфимском государственном авиационном техническом университете
по адресу: 450000, г. Уфа, ул. К. Маркса, 12

С диссертацией можно ознакомиться в библиотеке университета

Автореферат разослан 24 ноября 2009 г.

Ученый секретарь
диссертационного совета
д-р техн. наук, проф.

С. С. Валеев

ОБЩАЯ ХАРАКТЕРИСТИКА

Актуальность работы. Современный уровень развития информационных технологий характеризуется тенденцией доминирования электронных документов (ЭД) над традиционными бумажными носителями информации. Разнообразие и сложность задач, в которых задействованы электронные документы, способствуют развитию соответствующих научных исследований за рубежом и в нашей стране¹.

Как известно, одной из важных задач при обработке электронных документов является их персонализация. Распространенный подход использования обобщенных электронных форм документов, в которые пользователь вводит совокупность требуемых реквизитов, становится все менее актуальным. Так как важно не просто предоставить такую форму, а максимально адаптировать ее под конкретного пользователя. Такие документы называются *персонализированными*.

В основе построения таких документов лежит идея персонализации, которая применительно к теории разработки электронных документов заключается в настройке документов под определенного пользователя или подстановке персональных данных (реквизитов) пользователя в заданный макет.

Известно несколько подходов к персонализации электронных документов. В простейшем случае формы документов «вручную» заполняются пользователями. В автоматизированном режиме персонализация документов может быть выполнена с помощью серверных генераторов отчетов или систем электронного документооборота. На программном уровне персонализация документов обычно выполняется на основе СОМ-технологий с использованием объектных моделей соответствующих форматов.

В рамках исследований, проводимых на кафедре АСУ УГАТУ², была предложена технология персонализации ЭД, основанная на использовании макросов для генерации документов в формате Word. Такой подход обеспечивает эффективную подстановку в документы повторяющихся значений реквизитов пользователя, но не пригоден для использования на сервере.

В ходе анализа различных способов персонализации документов был выявлен ряд существенных недостатков. Во-первых, при вводе большого числа персональных реквизитов пользователя (при персонализации «вручную») возникает значительное число ошибок. Во-вторых, сложность установки и сопровождения автоматизированных систем персонализации неоправданно велика в контексте задачи генерации электронных документов. В-третьих, генерируемые в автоматизированном режиме документы обычно представлены в формате, закрытом для возможной модификации конечными пользователями. В-четвертых,

¹ Электронные документы исследуются в научных работах российских (О. В. Барышева, Р. Гиляревского, А. В. Конявского, Г. Г. Куликова, Л. Р. Черняховской) и зарубежных (Д. Шнайда, М. Стивенса) учёных.

² Обеспечение целостности комплекса электронных документов на основе встраиваемых динамических моделей: дис. ... канд. техн. наук / Т. А. Гарифуллин; науч. рук. проф. В. В. Миронов. Уфа: УГАТУ, 2006. 149 с.

использование СОМ-технологий сопровождается созданием на сервере дополнительных процессов (для каждого документа), что может привести к перегрузке сервера и снижению производительности его работы. В-пятых, персонализация, например, на основе макросов VBA может быть использована только на стороне клиента.

Выявленные недостатки требуют разработки научно обоснованных технологий генерации персонализированных документов, позволяющих сократить число ошибок подстановки значений реквизитов, не требующих трудоемкого сопровождения и запуска дополнительных процессов на сервере, позволяющие создавать документы редактируемого формата в процессе подстановки пользовательских реквизитов в шаблоны документов и доступных для использования на сервере.

В рамках выполненных исследований предлагается реализовать персонализацию электронных документов на основе XML и сопутствующих технологий. Персонализация электронных документов на основе XML-технологий основана на представлении пользовательских реквизитов в XML-формате (XML-база реквизитов³), к которым применяется заданное XSL-преобразование⁴ (XSL-спецификация персонализации).

Вместе с тем персонализация электронных документов на основе XML не является очевидной. Необходимо строго формализовать структуру персонализированного документа в контексте его XML-представления. Неясно, как построить шаблон персонализации на основе XML, как определить в нем правила подстановки пользовательских данных, как выполнить саму персонализацию, как сопоставить шаблон персонализации с пользовательскими данными.

Цель работы. Разработка научно обоснованной технологии автоматизированного создания на Web-сервере модифицируемых персонализированных электронных документов на основе XML.

Задачи исследования. Для достижения поставленной цели в работе сформулированы и решены следующие задачи:

1. Разработка концепции информационной технологии персонализации электронных документов на основе XML.
2. Разработка модели класса персонализированных документов.
3. Разработка метода создания стилей XSL-трансформации пользовательских XML-реквизитов.
4. Реализация программного обеспечения для создания спецификаций XSL-трансформации XML-реквизитов.

Методика исследования. В работе используются принципы и методы построения электронных документов, методы системного анализа, объектно-ориентированного программирования, теории информации, теории множеств, моделирования сложных объектов, принципы и методы алгоритмизации.

³ Под XML-базой реквизитов понимается множество персональных данных пользователя, представленных в формате XML.

⁴ XSL-преобразование позволяет трансформировать XML-данные в документ заданной структуры с помощью XSL-спецификаций (таблиц стилей).

Результаты, выносимые на защиту

1. Концепция информационной технологии персонализации электронных документов, основанная на использовании серверных макетов электронных документов и базы персональных реквизитов.

2. Модель класса персонализированных документов, основанная на представлении документа в виде иерархии вложенных фрагментов.

3. Метод XSL-инъекций для создания стилей XSL-трансформации XML-реквизитов в экземпляр персонализированного документа по модели класса персонализированных документов.

4. Программное обеспечение для создания спецификаций XSL-трансформации XML-реквизитов, основанное на технологиях обработки XML-данных, принципах XSLT-программирования и XML-организации документов Word.

Научная новизна результатов в целом связана с новой идеей использования XSL-трансформации для построения персонализированных Word-документов.

1. Новизна концепции информационной технологии персонализации электронных документов состоит в том, что с целью создания документов, допускающих дальнейшую модификацию и доработку, впервые макеты документов строятся в форме XSL-спецификаций, база реквизитов – в формате XML, а документы генерируются в формате Word путем XSL-трансформации.

2. Новизна модели класса персонализированных документов заключается в том, что с целью персонализации на основе XML-базы реквизитов в соответствии с концепцией информационной технологии персонализации электронных документов, впервые узлы иерархии включают: форматированные целевые фрагменты документа, внутри которых предусмотрены точки подстановки значений реквизитов, ссылающиеся на базу реквизитов, и XSL-инструкции, обеспечивающие установку контекста обработки фрагментов, выбор одного из нескольких фрагментов и циклическое копирование фрагментов в зависимости от состояния базы реквизитов.

3. Новизна метода XSL-инъекций для создания стилей XSL-трансформации XML-реквизитов состоит в том, что с целью снижения трудоемкости, впервые программирование таблиц стилей ведется в визуальной среде текстового процессора Microsoft Word, причем создается макет целевого документа с XSL-инструкциями вставки значений реквизитов выбора фрагментов и циклической обработки, далее макет, сохраненный в форме XML-документа, преобразуется в XSL-таблицу стилей.

4. Новизна разработанного программного обеспечения для создания спецификаций XSL-трансформации XML-реквизитов заключается в том, что с целью реализации предложенной концепции информационной технологии персонализации электронных документов, модели класса персонализированных документов и метода XSL-инъекций для создания стилей XSL-трансформации XML-реквизитов, впервые разработан комплекс шаблонов трансформации, со-

держащих правила обработки XML-макета конечного документа и корректного преобразования XSL-инъекций в XSL-инструкции соответствующего формата.

Внедрение результатов. Практическая значимость результатов подтверждается их использованием в УГАТУ в учебном процессе и в научно-производственной фирме «РД-Технология» в качестве методики программирования персонализированных электронных документов.

Связь с плановыми исследованиями. Работа выполнена в рамках плановых исследований кафедры АСУ УГАТУ по разработке информационных технологий веб-портала, реализующих создание и ведение персонализированных электронных документов, сопровождающих учебный процесс.

Апробация и публикации. Основные положения, представленные в диссертации, были заявлены и обсуждены на 6 конференциях всероссийского и международного уровня.

Список публикаций автора по теме диссертации включает 9 публикаций, в том числе 2 статьи в рецензируемом научном журнале из списка ВАК (23 с.), 1 свидетельство о государственной регистрации программы для ЭВМ, 6 публикаций в трудах конференций всероссийского и международного уровней (36 с.). 2 публикации выполнены без соавторов (8 с.)

Структура диссертации. Диссертация состоит из введения, четырех глав, заключения, приложения, библиографического списка из 97 наименований, всего на 187 листах.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во введении обсуждается актуальность решаемой научной задачи, формулируются цель работы, задачи, решаемые для достижения поставленной цели, результаты, выносимые на защиту. Обсуждается методика исследования, раскрывается научная новизна полученных в ходе исследования результатов, их значение для теории и практики. Рассматривается связь работы с плановыми исследованиями.

В первой главе рассматриваются понятие электронного документа, задача персонализации, анализируются известные подходы к персонализации электронных документов.

Электронные документы являются предметом многих научных исследований как в нашей стране, так и за рубежом. Особое место в этих исследованиях занимает терминология. В большинстве источников под электронным документом понимается зафиксированная на материальном носителе информация с реквизитами, позволяющими ее идентифицировать. Разнообразие видов электронных документов способствовало появлению значительного числа задач с их использованием. Особую роль в классе таких задач играет персонализация.

Понятие «персонализация» изначально применялось в полиграфии, а со временем распространилось и на другие сферы. В контексте электронных документов персонализация предполагает автоматизированное формирование документов, настроенных под конкретного пользователя.

В качестве основных составляющих персонализации документа выделены: шаблон конечного документа (статическая часть, одинаковая для всех пользователей) и персональные пользовательские реквизиты (переменная часть, которая по определенным правилам подставляется в шаблон документа).

На сегодняшний день задача персонализации документов решается только частично. Наиболее распространенным является решение на основе объектной модели соответствующих форматов. Каждый новый запрос на генерацию документа приводит к запуску на сервере нового процесса (исполняемого файла), что является недостатком этого подхода, поскольку может привести к перегрузке сервера. Другой подход ориентирован на использование инструментальных средств, к примеру, серверных генераторов отчетов. Его недостатком является специфичный формат формируемых документов, не поддерживающий модификацию пользователями.

В дополнение в первой главе обсуждается задача получения данных пользователя с помощью автономных форм, проводится анализ известных автономных технологий получения данных – InfoPath и Adobe Forms.

На основе результатов проведенного анализа в главе обосновывается необходимость и актуальность использования XML-технологий для персонализации электронных документов. Формулируются цели и задачи исследования, включающие разработку концепции информационной технологии персонализации электронных документов на основе XML, модели класса персонализированных документов, метода создания стилей XSL-трансформации пользовательских XML-реквизитов, реализацию программного обеспечения для создания спецификаций XSL-трансформации XML-реквизитов.

Вторая глава посвящена разработке концепции информационной технологии персонализации электронных документов на основе XML. Формулируются основы построения и применения персонализированных документов, рассматриваются возможные классификации персонализированных документов. Выполняется формализация модели класса персонализированных электронных документов и метода ее интерпретации.

Персонализированный электронный документ представляет собой электронный документ, содержащий общую и индивидуальную информацию, формируемую в соответствии с переданными пользователем данными.

В зависимости от предоставляемых пользователю возможностей персонализированные документы можно разделить на две группы: готовые документы (конечные документы) и документы-заготовки (модифицируемые документы). Последние с точки зрения возможности дальнейшей модификации делятся на однократные и многократные. В общем виде процедура формирования персонализированного документа представлена на рис. 1.

Одним из важных аспектов персонализации электронных документов является формат их представления. Несмотря на обилие форматов электронных документов, доминирующим по-прежнему остается формат Word. В этой связи было предложено разработать и исследовать технологию персонализации на примере этого формата.

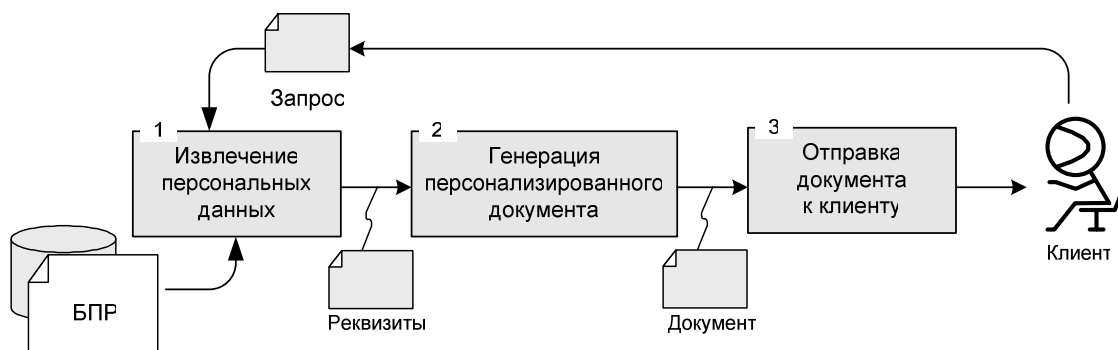


Рисунок 1 – Схема формирования персонализированного документа (БПР – база пользовательских реквизитов)

Основным условием персонализации документов является наличие на сервере персональной информации пользователя. Для этого на сервере размещается источник данных, содержащий персональные реквизиты пользователя. По запросу пользователя из источника данных извлекаются реквизиты (база пользовательских реквизитов), которые в работе предлагается представлять в XML-формате (XML-база реквизитов). Это позволит применить к ним любое преобразование и представить в удобном для пользователя виде.

Другим компонентом персонализации является макет конечного документа, в котором указывается, что должно и чего не должно быть в результирующем персонализированном документе. Макет представляет собой «константу» конечного документа, независимую от его переменной части – пользовательских реквизитов.

Макет содержит в себе определенную программную логику. Это инструкции, определяющие правила расстановки в макете пользовательских реквизитов. Именно эти правила активизируются при формировании конечного документа и по ним макет документа заполняется конечными пользовательскими данными. Правила расстановки пользовательских реквизитов можно разделить на инструкции подстановки значений реквизитов и инструкции формирования структуры конечного персонализированного документа.

Следующим компонентом механизма подстановки пользовательских данных является средство интерпретации правил подстановки реквизитов, которое должно обрабатывать все виды правил подстановки пользовательских реквизитов, анализировать архитектуру пользовательских данных, генерировать конечные персонализированные документы заданного формата.

Модель класса персонализированных документов в общем виде может быть представлена как совокупность фрагментов. Под фрагментом экземпляра документа понимается его часть, имеющая некоторое визуальное представление для пользователя. В модели класса фрагменту соответствует модель фрагмента – настраиваемая часть документа.

Графическое изображение модели ПЭД дано на рис. 2. Модель может быть представлена на макро- и микроуровнях.

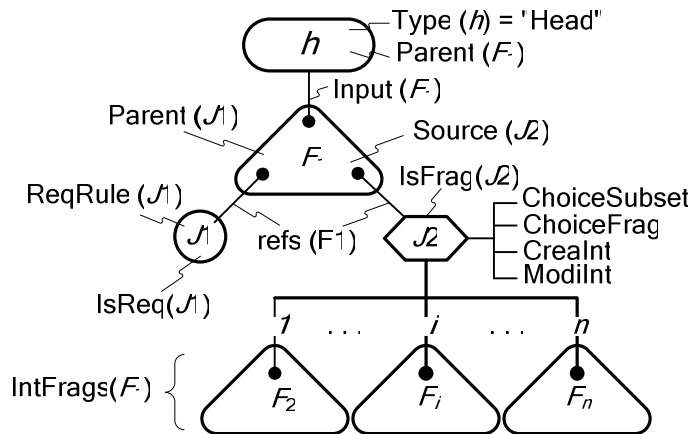


Рисунок 2 – Графическое изображение модели ПЭД

На макроуровне модель содержит конечное упорядоченное множество элементов $M = \{h\}$. Один из элементов $h \in M$ имеет тип «головной»: $Type(h) = \text{"head"}$, остальные элементы $f \in M, f \neq h$ имеют тип «фрагмент»: $type(f) = \text{"Frag"}$. $F_M = \{f\}$ – множество всех фрагментов модели. Множество M можно рассматривать как множество вершин ориентированного односвязного графа модели документа, дуги которого интерпретируются как отношения типа «родитель-ребенок».

На микроуровне каждый фрагмент $f \in F_M$ есть некоторое дерево, где $V(f)$ – множество вершин фрагмента f ; $root(f)$ – корневая вершина; $N(v)$ – нагрузка вершины $v, v \in V(f)$. Каждая вершина j модели фрагмента может быть нагружена значением (value) или ссылкой (ref): $vals(f)$ – множество вершин фрагмента f , нагруженных значениями; $refs(f)$ – множество вершин f , нагруженных ссылками. Ссылка j может быть одного из двух типов:

- реквизитной ($IsReq(j)$), которая интерпретируется как механизм, обеспечивающий подстановку значения реквизита. Для этого с вершиной ассоциируется правило $ReqRule$;

- фрагментной ($IsFrag(j)$), которая интерпретируется как механизм, обеспечивающий подстановку в фрагмент его дочерних (внутренних) фрагментов.

Из множества всех дочерних фрагментов, на которые ссылается ссылка j , при порождении конкретного экземпляра ПЭД выбирается только один. Для этого со ссылкой ассоциировано правило $ChoiceFrag$.

В экземпляре ПЭД выбранный фрагмент может копироваться в нескольких экземплярах. Для этого используется правило формирования множества внешних субконтекстов $ChoiceSubset$. Копии экземпляров выбранного фрагмента обрабатываются по-разному, для чего предусмотрены два правила управления внутренним контекстом: $CreaInt$ – правило создания внутреннего контекста, $ModiInt$ – правило модификации внутреннего контекста.

На уровне модели ПЭД база реквизитов представляется в виде иерархической модели данных – в виде дерева типов реквизитов. Тип реквизита подразумевает множество экземпляров (значений). База реквизитов представляет собой множество экземпляров деревьев, нагруженных значениями реквизитов.

Процесс порождения экземпляра рассматривается как процесс сборки ПЭД из отдельных фрагментов, модели которых содержатся в общей модели ПЭД. Процесс управляется правилами, ассоциированными с реквизитными и фрагментными ссылками в той же модели.

Обработка фрагмента происходит в буфере результата, куда процессор помещает обрабатываемые фрагменты, извлеченные из модели ПЭД. Здесь процессор настраивает фрагмент: реквизитные ссылки он заменяет значениями реквизитов, а фрагментные ссылки – соответствующими внутренними фрагментами. Чтобы обработать некоторый фрагмент, нужно выполнить подстановку в него внутренних фрагментов, которые предварительно нужно обработать аналогичным образом. У внутренних фрагментов могут быть свои внутренние фрагменты, которые требуют аналогичной обработки.

В третьей главе разрабатывается метод персонализации электронных документов на основе XML-технологий. Обсуждаются различные методики построения XSL-макетов персонализированных документов. Предлагается новый способ программирования XSL-макетов – на основе введения XSL-инъекций в визуальной среде Word.

В общем виде процедура персонализации документа может быть представлена как преобразование множества пользовательских реквизитов в персонализированный документ. Предложенная в работе XML-организация персональных данных сводит задачу персонализации к трансформации этих данных в заданную структуру.

Поскольку в работе персонализация рассматривается на примере документов формата Word, то необходима инструкция трансформации, позволяющая на программном уровне преобразовать XML-реквизиты пользователя в персонализированный документ Word. Этому способствует еще один аргумент в пользу XML – появление XML-формата для документов Word – WordML (Wordprocessing Markup Language). Документы WordML по своей функциональности являются XML-документами, поэтому к ним могут быть применены известные решения по программной обработке XML-данных – XPath-адресация, DOM-объекты, XSL-трансформация и т. д.

Задача персонализации представляется как преобразование XML-реквизитов в формат WordML конечного документа. Из компонентов персонализации правила подобной трансформации на программном уровне могут быть заданы в макетах документов.

Анализ способов преобразования XML-данных (конверторы, COM-технологии и др.) показал эффективность и гибкость подхода, основанного на механизмах XSL-трансформации. В этой связи в качестве макетов персонализированных документов предлагается использовать XSL-таблицы стилей, названные XSL-спецификациями персонализации.

В соответствии с правилами XSL-преобразования WordML-разметка представляет собой так называемые конечные литеральные элементы, предназначенные для копирования в результирующий документ. Поэтому можно сразу составить необходимую иерархию WordML-элементов, образующих струк-

туру формируемого документа. Чтобы дополнить полученный шаблон пользовательскими данными, следует применить инструкции, указывающие XSL-процессору, какие элементы и атрибуты исходного XML-документа должны быть размещены в той или иной позиции, как в зависимости от их значений может варьироваться структура конечного документа. XSL-спецификация персонализации формируется путем комбинирования WordML-разметки с XSL-инструкциями подстановки значений реквизитов и управления структурой конечного документа.

Библиотека WordML насчитывает несколько десятков XML-элементов и атрибутов, описывающих структуру документа Word (абзацы, слова, символы, объекты и т. д.). Корректно комбинируя эти элементы, можно построить простой документ Word даже «вручную», вводя программные инструкции в любом текстовом или XML-редакторе. Однако документы со сложной структурой и форматированием (каких, по сути, большинство и для которых вообще целесообразно выполнять персонализацию) построить таким способом весьма сложно, а зачастую и невозможно. В условиях распространения средств и методов визуального программирования большой интерес представляет идея разработки таблицы стилей «с экрана». При таком подходе текстовый процессор Microsoft Word можно использовать как среду визуального программирования документов формата WordML.

Визуальная среда Microsoft Word позволяет сохранять документ Word в классе XML-документов с сохранением текстового содержимого и правил его форматирования. В результате независимо от сложности документа конечные литеральные элементы WordML генерируются автоматически. Полученная разметка дополняется некоторыми XSL-инструкциями и доводится до уровня XSL-таблицы стилей (указываются корневой элемент, шаблон и т. д.).

Формирование XSL-спецификации персонализации в общем виде можно представить следующим образом. На начальном этапе разработки проектировщик в знакомой ему среде приложения Microsoft Word создает стандартный текстовый документ, определяет расположение отдельных текстовых фрагментов и задает правила их форматирования. Полученный документ сохраняется в XML-формате и на следующем шаге дополняется инструкциями подстановки реквизитов пользователя.

Наиболее сложным представляется этап расстановки XSL-инструкций, поскольку WordML-код может быть сгенерирован автоматически. В этой связи важно определить, каким образом идентифицировать позиции в документе, куда должны быть добавлены пользовательские данные из XML-базы реквизитов. Для этого могут быть использованы, к примеру, текстовые метки – слова в тексте документа, которые легко отличить от остальных текстовых фрагментов. Впоследствии текстовые метки, найденные в WordML-коде, могут быть заменены на соответствующие XSL-инструкции. Полученная XSL-таблица стилей является XSL-спецификацией персонализации.

Однако такой подход не лишен существенных недостатков. Это весьма трудоемкий процесс: нужно сначала определить в Word-макете расположение XSL-инструкций, а затем расставить их в WordML-коде. Независимо от того, в

какой среде выполняется подстановка – в простом текстовом редакторе или в XML-редакторе – разработчик имеет дело с неструктурированным текстом огромного объема. Чтобы правильно расставить XSL-инструкции, нужно просмотреть каждую строку такого текста, а, подставляя XSL-инструкции, важно не повредить корректность всего XML-документа.

В работе предлагается подход, позволяющий совместить этапы подготовки XSL-макета конечного документа и использовать Microsoft Word в качестве среды визуального программирования.

Вместо внесения XSL-инструкций в WordML-заготовку таблицы стилей, можно вносить их на более раннем этапе – в исходном Word-макете целевого документа, из которого формируется таблица стилей. Такие «вкрапления» XSL-инструкций в Word-документе названы *XSL-инъекциями*. Очевидным преимуществом этого подхода является то, что вставка инструкций выполняется в визуальном режиме – в среде Microsoft Word, а не в коде, отображенном текстовым редактором.

При таком подходе программирование XSL-существенно упрощается. Разработчик совмещает решение сразу двух задач – определяет внешний вид целевого документа и расставляет в нем XSL-инструкции подстановки реквизитов, организации циклов и модулей и т. д. В результате получается WordML-документ, который в текстовых элементах своей разметки содержит «вкрапления» разметки XSL.

В процессе исследования предложенного подхода было выявлено следующее затруднение применения идеи XSL-инъекций. В разметке WordML открывающая и закрывающая угловые скобки являются зарезервированными символами. Поэтому они автоматически заменяются на конструкции `<` и `>`. XML-процессор, который должен выполнить преобразование, воспринимает XSL-инструкции как текст, который без изменений должен быть скопирован в результат трансформации.

Для решения этой проблемы была разработана XSL-таблица стилей преобразования WordML → XSL, которая формирует XSL-спецификацию персонализации из документа WordML. Применение трансформации позволяет интерпретировать XSL-инъекции как программные XSL-инструкции.

Спецификация преобразования WordML → XSL представлена множеством XSL-шаблонов форсирующей трансформации. Одни из них ориентированы на обработку определенных WordML-элементов, а другие – на принудительный вызов (как функции в традиционных языках программирования).

В общей сложности в спецификации трансформации было разработано одиннадцать программных шаблонов. В шаблонах заданы программные инструкции, выполняющие блокирование зарезервированных символов (открывающей и закрывающей угловых скобок), «склеивку» расчлененных инструкций (XSL-инъекция может быть «разорвана» между несколькими WordML-элементами), баланс парных тегов, объявление модулей.

Кроме того, метод XSL-инъекций позволяет в визуальном режиме задавать переменные, вычисляемые выражения, проверку условий и др., обеспечи-

вая тем самым создание таблиц стилей персонализированных документов сложной структуры и наполнения.

В четвертой главе обсуждаются вопросы применения концепции информационной технологии персонализации электронных документов, модели класса персонализированных документов, метода XSL-инъекций и программного обеспечения для решения практических задач.

В качестве примера рассматривается актуальная задача – подготовка документов из диссертационного дела соискателя ученой степени. К ним относятся документы, с формированием которых соискатель сталкивается на трех последних этапах диссертационной работы – предзащите, утверждении оппонентов и непосредственно защите.

Анализ данных, которые должны быть использованы в документах диссертационного дела, показал, что их условно можно разделить на три группы:

- общие сведения о соискателе и его диссертационной работе. К ним относятся фамилия, имя и отчество диссертанта, название специальности, по которой планируется защита, тема работы, данные научного руководителя и т. д.;
- отзывы, полученные на диссертационную работу, от ведущей организации, официальных оппонентов, от сторонних организаций с замечаниями и ответами соискателя на них;
- результаты диссертационной работы: основные результаты, научная и практическая значимость, плановые исследования и т. д.

Для получения этих данных от соискателя были разработаны соответствующие интерфейсные документы, позволяющие сформировать XML-базу реквизитов каждого пользователя.

Анализ структуры диссертационного дела показал, что на разных стадиях защиты диссертационной работы соискателем должно быть подготовлено 18 различных документов. Важно отметить, что большинство реквизитов повторяются из документа в документ. В то же время структура этих документов может быть в определенной степени обобщена – можно выделить постоянную часть (которая присутствует в документе независимо от данных соискателя степени) и переменную (реквизиты отдельных соискателей).

Для осуществления персонализированной генерации документов были разработаны соответствующие XSL-спецификации персонализации. Для этого были разработаны документы Word с XSL-инъекциями, преобразованные в XSL-спецификации персонализации с помощью трансформации WordML → XSL.

На рис. 3 представлено поэтапное создание XSL-спецификации персонализации одного из документов из диссертационного дела соискателя. В визуальной среде Microsoft Word был разработан «макет» соответствующего документа. В нем были определены основные структурные компоненты, задано их информационное содержимое и правила форматирования (рис. 3, а). На следующем шаге макет документа был дополнен XSL-инъекциями, задающими XSL-инструкции подстановки пользовательских реквизитов (рис. 3, б). К этому документу была применена трансформация WordML → XSL, результатом которой стала XSL-спецификация персонализации документа (рис. 3, в).

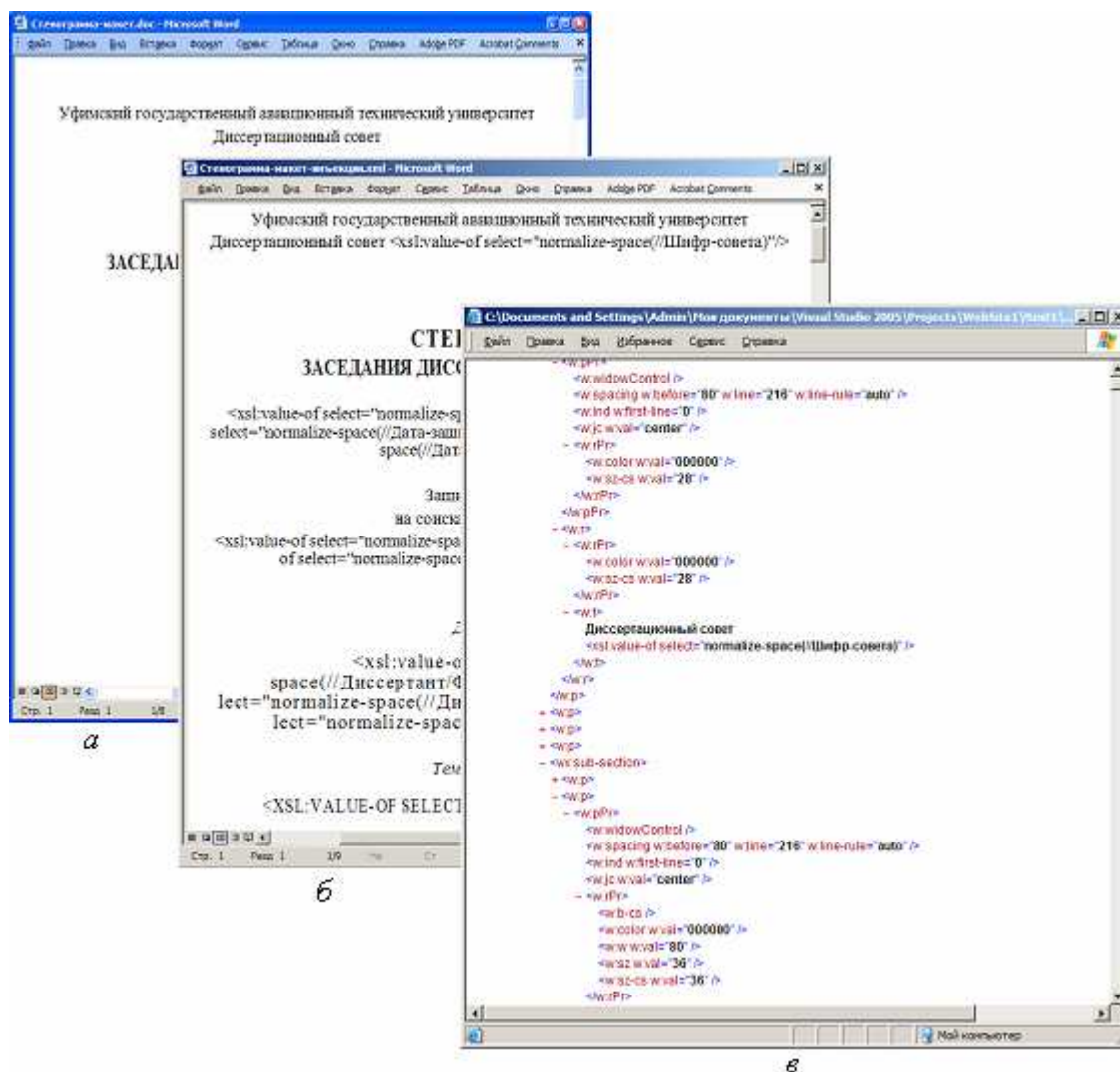


Рисунок 3 – Построение XSL-спецификации персонализации документа из диссертационного дела соискателя: *а* – макет документа Word; *б* – макет Word с XSL-инъекциями; *в* – отображение XSL-спецификации персонализации в браузере

Для практической иллюстрации применения автономных интерфейсных документов получения данных и XSL-спецификаций персонализации документов из диссертационного дела было разработано Web-приложение информационной поддержки соискателей ученой степени (рис. 4). В качестве основной программной технологии, использованной при разработке Web-приложения, была выбрана технология разработки серверных сценариев ASP.NET.

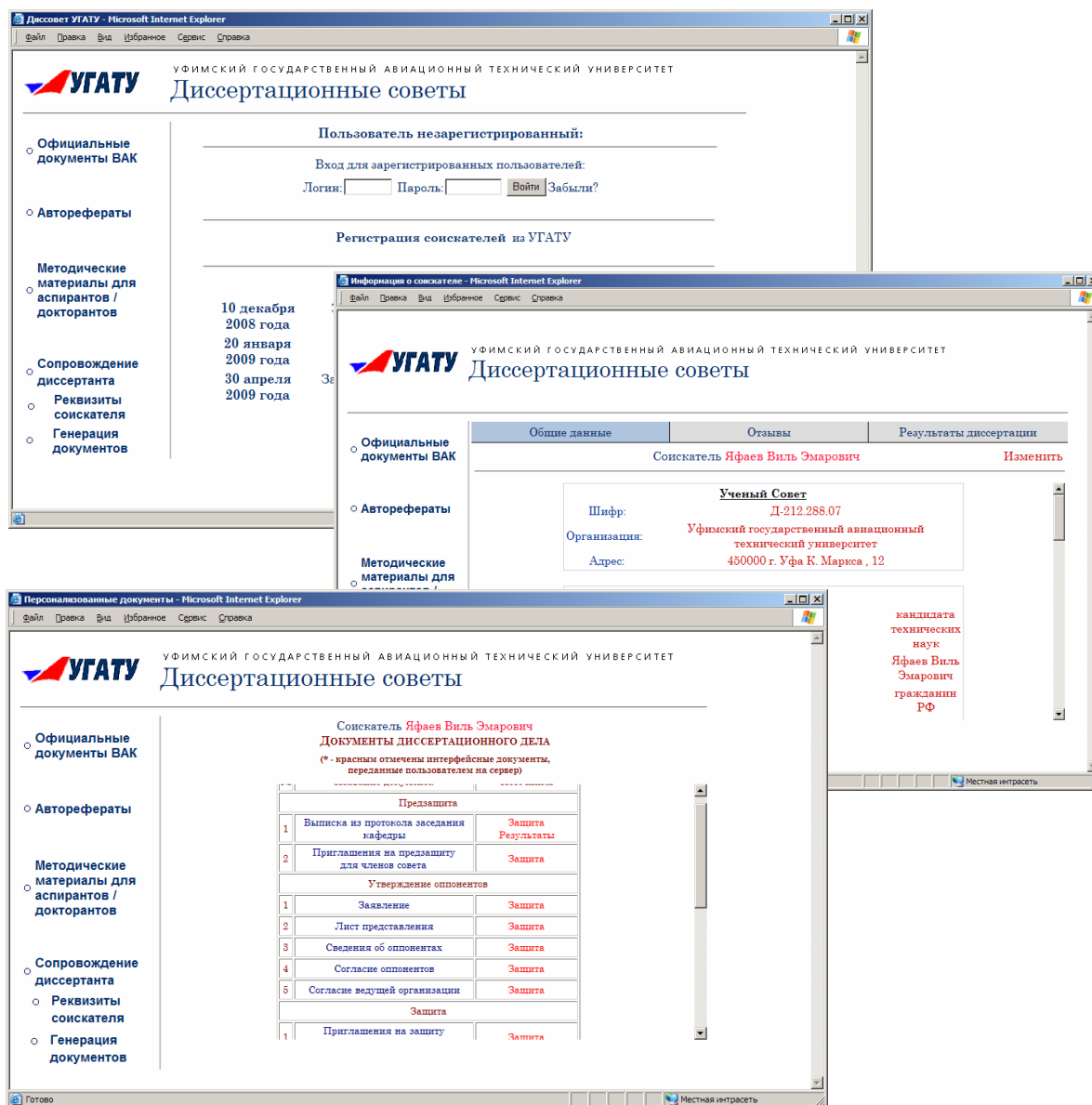


Рисунок 4 – Web-приложение поддержки соискателя ученой степени

Для оценки эффективности разработанного инструментально-программного средства был проведен сравнительный анализ трудоемкости и сокращения ошибок при программировании XSL-спецификаций персонализации с помощью различных инструментально-программных средств. С этой целью были рассмотрены четыре подхода:

- текстовый редактор «Блокнот» как самое простое и доступное программное средство создания и редактирования XML и XSL-кода;
- «связка» текстового процессора Microsoft Word, используемого для автоматической генерации WordML-конечных литеральных элементов, и текстового редактора «Блокнот» для редактирования полученного программного кода;
- «связка» текстового процессора Microsoft Word, используемого для автоматической генерации WordML-конечных литеральных элементов, и визуальной среды разработки Microsoft Visual Studio 2008 для редактирования по-

лученного программного кода, позволяющей работать с XML- и XSL-разметкой в частично автоматизированном режиме;

– спецификация трансформации Word2XSL, предложенная в диссертационной работе для решения задачи автоматической генерации XSL-спецификации персонализации на основе документа формата Microsoft Word, содержащего XSL-инъекции подстановки значений реквизитов и управления структурой документа.

При программировании XSL-спецификаций персонализации были выделены два ключевых момента. Во-первых, это трудоемкость собственно программирования. Эту характеристику можно условно разбить на две подгруппы. Первая из них предполагает оценку трудоемкости с точки зрения времени, которое было затрачено разработчиком на программирование XSL-спецификации персонализации. Вторая подгруппа представляет собой оценку трудоемкости вообще, с точки зрения трудоемкости каждого отдельного действия, выполняемого разработчиком при программировании XSL-спецификации персонализации.

Экспериментальные исследования показали, что применение предложенного метода XSL-инъекций и разработанной спецификации трансформации WordML → XSL позволяет существенно сократить трудоемкость программирования XSL-спецификаций персонализации по количеству выполняемых процедур в среднем до 60 раз; по затрачиваемому времени в среднем до 35 раз; по сложности выполняемых процедур в среднем до 25 раз. Такие результаты получены за счет того, что при использовании XSL-инъекций разработчик оперирует только XSL-инструкциями, а WordML-разметка генерируется автоматически, в отличие, скажем, от разработки XSL-спецификации персонализации с помощью текстового редактора.

Второй параметр, который был рассмотрен при сравнительной оценке технологии программирования XSL-спецификаций персонализации, связан с определением и анализом тех ошибок, которые могут возникнуть со стороны разработчика при программировании. Использование XSL-инъекций также показало существенное повышение качества XSL-спецификаций персонализации. Это связано с тем, что предложенная технология не предполагает непосредственной работы с WordML-кодом. Все необходимые корректировки вводятся разработчиком непосредственно в документ в визуальной среде Word. В результате WordML-код не повреждается, XSL-спецификация, а соответственно, и сгенерированные на ее основе персонализированные документы остаются корректными и могут быть использованы конечными пользователями таких документов.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

В работе решена задача разработки научно обоснованной технологии создания персонализированных электронных документов на основе XML. Получены следующие результаты.

1. Концепция информационной технологии персонализации электронных документов, основанная на использовании серверных макетов электронных

документов и базы персональных реквизитов, причём макеты документов строятся в форме XSL-спецификаций, база реквизитов – в формате XML, а сами документы генерируются в формате Word путем XSL-трансформации. Это позволит создавать документы, допускающие дальнейшую модификацию и доработку.

2. Модель класса персонализированных документов, основанная на представлении документа в виде иерархии вложенных фрагментов, причём с целью персонализации на основе XML-базы реквизитов в соответствии с концепцией информационной технологии персонализации электронных документов, узлы иерархии включают: форматированные целевые фрагменты документа, внутри которых предусмотрены точки подстановки значений реквизитов, ссылающиеся на базу реквизитов, и XSL-инструкции, обеспечивающие установку контекста обработки фрагментов, выбор одного из нескольких фрагментов и циклическое копирование фрагментов в зависимости от состояния базы реквизитов. Это позволит выполнять персонализацию на основе XML-базы реквизитов в соответствии с концепцией информационной технологии персонализации электронных документов.

3. Метод XSL-инъекций для создания стилей XSL-трансформации XML-реквизитов в экземпляр персонализированного документа по модели, причём с целью снижения трудоемкости программирование таблиц стилей ведется в визуальной среде текстового процессора Microsoft Word, при этом создается макет целевого документа с XSL-инструкциями вставки значений реквизитов, выбора фрагментов и циклической обработки, далее макет, сохраненный в форме XML-документа, преобразуется в XSL-таблицу стилей. Это позволит снизить трудоемкость программирования таблиц стилей.

4. Программное обеспечение для создания спецификаций XSL-трансформации XML-реквизитов, основанное на технологиях обработки XML-данных, принципах XSLT-программирования и XML-организации документов Word, причём с целью реализации предложенных концепции, модели и метода разработан комплекс шаблонов трансформации, содержащих правила обработки XML-макета конечного документа и корректного преобразования XSL-инъекций в XSL-инструкции соответствующего формата. Это позволит реализовать предложенные концепцию информационной технологии персонализации электронных документов, модель класса персонализированных документов и метод XSL-инъекций для создания стилей XSL-трансформации XML-реквизитов.

ОСНОВНЫЕ ПОЛОЖЕНИЯ ДИССЕРТАЦИИ ОПУБЛИКОВАНЫ В РАБОТАХ

В рецензируемых журналах из списка ВАК

1. Информационная технология персонализации электронных документов Microsoft Office в Web-среде на основе XML / В. В. Миронов, Г. Р. Шакирова, В. Э. Яфаев // Вестник УГАТУ: науч. журн. Уфимск. гос. авиац. техн. ун-та. Серия «Управление, вычислительная техника и информатика». 2008. Т. 10, № 2(27). С. 112–123.

2. Иерархическая модель персонализированных документов и ее XML-реализация / В. В. Миронов, Г. Р. Шакирова, В. Э. Яфаев // Там же. 2008. Т. 11, № 1(28). С. 164–174.

В других изданиях

3. XSL-трансформация для преобразования XML-базы реквизитов в документ Word: свид. об официальной рег. программы для ЭВМ № 2007613608. Зарег. 24 августа 2007 года / К. Э. Маликова, В. Э. Яфаев. М.: Роспатент, 2007.

4. Информационная интернет-технология ведения электронных документов с использованием Microsoft Office 2003 на основе языка XML / К. Э. Маликова, В. Э. Яфаев // XXXIII Гагаринские чтения: науч. тр. Междунар. молодежн. науч. конф. М.: МАТИ, 2007. Т. 4. С. 183–184.

5. Прямое и обратное преобразование Word и XML документов стандартными средствами Microsoft / В. Э. Яфаев // Актуальные проблемы в науке и технике: сб. ст. 2-й рег. зимн. шк.-сем. аспирантов и молодых ученых. Уфа: Технология, 2007. С. 146–148.

6. Автоматизированная информационная система ведения электронных документов на основе интернет-технологий / К. Э. Маликова, Г. Р. Шакирова, В. Э. Яфаев // Актуальные проблемы в науке и технике: сб. тр. 3-й Всерос. зимн. шк.-сем. аспирантов и молодых ученых. Уфа: Технология, 2008. Т. 1. С. 48–55.

7. XML-ориентированный подход к автоматизации создания и ведения электронных документов / К. Э. Маликова, Г. Р. Шакирова, В. Э. Яфаев // Там же. С. 65–75.

8. Концептуальный XML-ориентированный алгоритм персонализации электронных документов Microsoft Office / В. В. Миронов, Г. Р. Шакирова, В. Э. Яфаев // Тр. междунар. конф. компьют. наук и информ. техн. (CSIT'2008). Анталия, Турция, 2008. Т. 1. С. 15–20. (Ст. на англ. яз.).

9. Серверная технология создания персонализированных электронных документов на основе XML / В. Э. Яфаев // Актуальные проблемы в науке и технике: сб. ст. Всерос. зимн. шк.-сем. аспирантов и молодых ученых. Уфа: Диалог, 2009. Т. 1. С. 87–91.

ЯФАЕВ Виль Эмарович

XML-ТЕХНОЛОГИЯ СОЗДАНИЯ НА WEB-СЕРВЕРЕ
ПЕРСОНАЛИЗОВАННЫХ WORD-ДОКУМЕНТОВ
НА ОСНОВЕ XSL-ТРАНСФОРМАЦИИ

Специальность 05.13.11 - Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Подписано к печати __.__.__. Формат 60x84 1/16.
Бумага офсетная. Печать плоская. Гарнитура Таймс.
Усл. печ. л. 1,0. Усл. кр. – отт. 1,0. Уч.-изд. л. 1,0.
Тираж 100 экз. Заказ № 581.

ГОУ ВПО Уфимский государственный авиационный технический университет
Центр оперативной полиграфии
450000, Уфа-центр, ул. К.Маркса,12