

Содержание:

ВВЕДЕНИЕ

Совершенствование технических возможностей средств вычислительной техники, развитие коммуникационных средств и технологий управления информационными ресурсами в последние годы привели к появлению более крупных информационных систем. Речь идет о масштабах систем не только относительно объема поддерживаемых информационных ресурсов, но и числа их пользователей. Объем информационных ресурсов Web в настоящее время исчисляется многими миллионами страниц.

В связи с этим развитием информационных технологий, сетей, а также информационных систем получил широкое распространение язык гипертекстовой разметки HTML. Информационные системы при этом рассматриваются как инструмент моделирования реальности, реализующий различные подходы. В последние годы стали появляться инструментальные средства и крупные информационные системы, в которых совместно используются различные информационные технологии. Сейчас существует множество специализированных программ для разработки Web сайтов. Такие программы, облегчают работу разработчикам в создании Web страниц со сложным дизайном, позволяют динамически генерировать страницы Web.

Информационные системы сегодня применяются во всех областях общественной жизни и научной деятельности. Курсовая работа предназначена для обобщения накопленного отечественного и зарубежного опыта в разработке информационных систем связанная с Web-технологиями, выявление общих положений и принципов их построения и развития. Представленная работа показывает значимость и эффективность использования информационных систем в первую очередь для поддержки человеческой деятельности в различных областях науки, образования и культуры.

Цель курсовой работы - изучить теоретический материал по тематике курсовой работы

Для достижения поставленной цели были выделены следующие **задачи**:

- проанализировать литературу по теме курсовой работы;
- рассмотреть и изучить понятия: «информационная система», «технология Web»;
- охарактеризовать основные технологии Web, такие как: HTML, XML, XHTML;
- обозначить новые тенденции в развитии технологий Web;
- рассмотреть и проанализировать спецификации Web-языков;

Структура курсовой работы: работа состоит из введения, трех глав, заключения, списка литературы, включающего в себя 26 источников и двух приложений.

Первая глава посвящена изучению основных понятий, таких как «информационная система» и «технологии Web».

Вторая глава посвящена изучению технологии Webна основе языков разметки: HTML, XML, XHTML.

В третьей главе рассматриваются спецификаций Web-языков.

1. ОПРЕДЕЛЕНИЕ СОДЕРЖАНИЯ ОСНОВНЫХ ПОНЯТИЙ

Понятие «информационная система» в Web

Создание современных электронных вычислительных машин позволило автоматизировать обработку данных во многих сферах человеческой деятельности. Без современных систем обработки данных трудно представить сегодня передовые производственные технологии, управление экономикой на всех ее уровнях, научные исследования, образование, издательское дело, функционирование средств массовой информации, проведение крупных спортивных состязаний и т.д. Значительно расширило сферу применения систем обработки данных появление персональных компьютеров.

Одним из наиболее распространенных классов систем обработки данных являются информационные системы (ИС). Любой вид деятельности основывается на информации о свойствах состояния и поведения той части реального мира, с которой связана эта деятельность. Для получения такой информации во многих случаях необходимо регулярно через некоторые интервалы времен проводить

измерения или наблюдения, позволяющие определить характеристики состояния сущностей реального мира и протекающих процессов, соответствующие моментам времени, когда эти изменения производятся. Именно для этого существует специальный класс систем обработки данных – автоматизированные информационные системы (АИС).

Автоматизированной информационной системой называется комплекс, включающей вычислительное и коммуникационное оборудование, программное обеспечение, лингвистические средства информационные ресурсы, а так же системный персонал и обеспечивающий поддержку динамической информационной модели некоторые части реального мира для удовлетворения информационных потребностей пользователей [Когаловский, М.Р. Перспективные технологии информационных систем / М.Р. Когаловский. – М.: Компания АйТи, 2003.- С. 13].

Под динамической моделью в данном понятии изменяемость модели во времени. Это «живая», действующая модель, в которой отображаются изменения, происходящие в предметной области. Такая система должна обладать памятью, позволяющей ей сохранять не только сведения о текущем состоянии предметной области, но и в некоторых случаях предысторию. Поскольку эта модель, поддерживаемая ИС, материализуется в форме организованных необходимым образом информационных ресурсов, она называется информационной моделью.

АИС не всегда функционирует самостоятельно. Она может входить в качестве компонента (подсистемы) в более сложную систему, такую, например, как система управления торговой компанией, САПР или система управления производством, учреждением и т. д.

Информационные системы уже многие десятки и даже сотни лет существуют и используются на практике в фактографических системах, которые основаны на технологиях баз данных и оперируют структурированными данными, системы текстового поиска, оперирующие документами на естественных языках, глобальную гипермедийную информационную систему Web и др. По этой причине в определении используется обобщенный термин информационные ресурсы. Частными его случаями являются данные для систем баз данных, документы для систем текстового поиска, HTML-страницы или XML-документы для Web и т.д.

Важный факт состоит в том, что единого устоявшегося и общепринятого определения понятия «информационная система» в настоящее время не

существует, да и вряд ли оно может существовать. Дело в том, что в зависимости от необходимости в разных случаях используются разные точки зрения на такой сложный продукт высоких технологий, каким являются современные информационные системы.

Специалисты по системному проектированию трактуют понятие ИС более широко, чем комплекс, о котором идет речь в приведенном выше определении. При этом в состав ИС включаются, например, организационно-методические и технологические документы.

Информационные системы используют ресурсы нескольких категорий — средства вычислительной техники, системное и прикладное программное обеспечение, информационные, лингвистические и человеческие ресурсы. Кроме того, хотя об этом не говорится в известных определениях автоматизированных информационных систем, но подразумевается как само собой разумеющееся, для функционирования системы необходимы и другие ресурсы — помещения, их техническое оснащение, всевозможная оргтехника, электроснабжение и т.д.

Пользовательские информационные ресурсы в Web — это страницы Web-сайтов, ресурсы «скрытого» Web — базы данных, а также различные доступные пользователям Web-документы, представленные в форматах, отличных от HTML. В Web нового поколения к информационным ресурсам, кроме того, относятся не только представленные на Web-сайтах XML-документы, но и различные метаданные. Они описывают схемы XML-документов, их семантику, онтологии.

Во многих публикациях употребляется словосочетание специализированная информационная система. Из выше приведенного определения информационной системы следует, что универсальных информационных систем не бывает. Каждая из них существует в единственном числе, ее тиражирование бессмысленно, поскольку такая система моделирует конкретную предметную область, поддерживает характеризующие ее свойства информационные ресурсы, которые ассоциированы с конкретными моментами или периодами времени. Поэтому специализированной является каждая информационная система.

Усиливается тенденция глобализации ИС. Глобализация информационных систем имеет две стороны – обеспечение глобального доступа пользователей к системе и интеграция информационных ресурсов, распределенных в глобальной сети. Уникальной глобальной ИС является Web. В нем воплощаются обе указанные стороны глобализации ИС. Он обеспечивает глобальный доступ к явно

представленным на Web-сайтах информационным ресурсам, а также к ресурсам «скрытого» Web. Вместе с тем на платформе Web создаются разработанные приложения, обеспечивающие интеграцию распределенных в Web информационных ресурсов. Многочисленные глобальные системы создаются в настоящее время как приложения Web для электронного бизнеса, для поддержки научной кооперации различных коллективов ученых во многих областях знаний в международном и национальном масштабе, в библиотечном деле и в других сферах. Среда Web предоставляет для поддержки таких систем идеальные условия.

Понятие «технология Web»

Создание глобальной гипермедийной распределенной информационной системы *WorldWideWeb*, функционирующей в среде Internet и часто называемой в отечественной литературе *Всемирной паутиной*, является одним из крупнейших научно-технических достижений последнего десятилетия XX в., основой ряда новых информационных технологий, имеющих весьма значимые социально-экономические последствия.

Технологии, сформировавшиеся в процессе создания и развития этой системы, оказывают влияние на развитие других областей информационных технологий. На их основе развиваются подходы и методы интеграции неоднородных ресурсов — весьма актуальная тенденция в разработках информационных систем. Разрабатываются новые подходы к созданию распределенных систем. Возникли новые, чрезвычайно важные сферы применения — электронный бизнес, электронные библиотеки. Создаются крупные научные и образовательные системы.

Достигнутые за короткую историю существования Web масштабы этой глобальной распределенной информационной системы по количеству ее пользователей, по объему предоставляемых информационных ресурсов, по составу функционирующих в ее среде приложений привели к существенному росту функциональных требований к ней. Потребовались радикально новые подходы, которые могли бы обеспечить дальнейшее развитие Web. Их реализация составляет одну из главных задач консорциума W3C (*WorldWideWebConsortium*) ,начиная с середины 90-х годов.

Термину “**Web**” разные энциклопедии и словари дают следующее усредненное толкование: “глобальное информационное пространство, образованное связанными

ссылками гипертекстовыми документами, основанное на физической инфраструктуре Internet и протоколах передачи данных этой сети, непрерывно эволюционирующее”. Сразу можно отметить, что социальная подсистема в приведенном собирательном определении присутствует исключительно в неявной форме – подразумевается, что кто-то (или что-то) создает и эти самые гипертекстовые документы, и физическую инфраструктуру, и протоколы.

WWW (WorldWideWeb) или **Web** – поддерживаемая в Internet глобальная открытая бесконечно масштабируемая распределенная гипермедийная информационная система с архитектурой “клиент-сервер” распределение и неоднородность ресурсов которой прозрачны для пользователей. Система обладает огромным интенсивно наращиваемым информационным ресурсом, большинство из которых предоставляется для свободного доступа в любой момент времени. Среда WWW способна интегрировать ресурсы других информационных сервисов Internet – Gopher, FTP, Archie, WAIS, Telnet, электронной почты. Она обеспечивает также телекоммуникационный доступ к базам данных. Наиболее активно используемыми сервисами Internet являются WWW, электронная почта, сервисы передачи файлов, поддержки телеконференций, удаленного доступа к вычислительным ресурсам.

Все информационные сервисы Internet строятся на основе архитектуры «клиент-сервер». Некоторые из них, например WWW, поддерживают распределенные информационные ресурсы.

Для информационных сервисов Internet создано разнообразное свободно распространяемое и коммерческое программное обеспечение, функциональные возможности которого не зависят от специфических особенностей конкретных аппаратно-программных платформ, на которых оно используется. Это достигается благодаря стандартизации технологий, на которых эти сервисы базируются, и поддержке этих стандартов в указанном программном обеспечении. Благодаря тому, что сеть Internet построена на основе стандарта *эталонной модели сетевого взаимодействия открытых систем (OpenSystemInterconnection— OSI)*, это программное обеспечение не зависит так же от особенностей сетей, входящих в состав Internet. Его место в эталонной модели — *прикладной уровень*.

В глобальной коммуникационно – вычислительной сети Internet функционирует ряд информационных и других сервисов, услуги которых свободно доступны для любого пользователя или предоставляются при условии, если пользователь обладает необходимыми полномочиями. В последнем случае обычно используется меха

низм, предусматривающий предъявление пользователем своего идентификатора (имени) и пароля для подтверждения его полномочий доступа.

Термин “технологии Web” или “Web – технологии” объединяет в себе два выше рассмотренных понятия “Web” и “Информационная технология”. В большинстве энциклопедий и словарей под *Web - технологией* понимается “технология построения Всемирной паутины, представление различного рода документов, находящихся в Интернете в виде связанных между собой системой ссылок” [Ваулина, Ч.Ю. Информатика: толковый словарь / Ч.Ю. Ваулина. – М.: Изд-во Эксмо, 2005. – С. 474].

Глава 2. ТЕХНОЛОГИИ WEB

Современные информационные Web технологии быстро изменяют наш мир и непосредственно влияют на развитие Web-технологий. Эта технологическая революция сильно повлияла не только на бизнес, но также на частную и профессиональную жизнь. Новейшие Web-технологии проникают во все сферы жизни общества, изменяют способы общения и принципы ведения Web-проектов современных компаний, определяя судьбу предприятий. Внутренняя сложность и предельная простота применения современных информационных Web технологий делает их доступными каждому, кто ежедневно сталкивается с применением их в своей профессиональной деятельности.

И в быту, и в бизнесе, в переписке и торговле люди и организации используют Web, создают собственные Web-узлы, где предлагают информацию, товары и услуги. Средства создания Web-ресурсов развиваются стремительно и без остатков, позволяют создавать сложные Web-документы, не требуя специальных знаний об их структуре и внешнем виде, освобождая время для продуктивной творческой деятельности. Главное преимущество Web-технологий в современных условиях заключается в их простоте и как следствие в повышении эффективности их применения.

Язык гипертекстовой разметки HTML

Популярность Internet во многом вызвана появлением WorldWideWeb (WWW), так как это первая сетевая технология, которая предоставила пользователю простой современный интерфейс для доступа к разнообразным сетевым ресурсам.

Простота и удобство применения привели к росту числа пользователей WWW и привлекли внимание коммерческих структур. Далее процесс роста числа пользователей стал лавинообразным, и так продолжается до сих пор. На основе необходимости объединить все множество информационных ресурсов начала развиваться технология при помощи, которой определяется гипертекстовая навигационная система. Этой технологией стал язык HTML. Технология HTML на начальном этапе была чрезвычайно проста, и практически все пользователи сети одновременно получили возможность попробовать себя в качестве создателей и читателей информационных материалов, опубликованных во Всемирной паутине. Дело в том, что при разработке различных компонентов технологии предполагалось, что квалификация авторов информационных ресурсов и их оснащенность средствами вычислительной техники будут минимальными.

Язык HTML (HyperTextMarkupLanguage, язык разметки гипертекста) относится к числу так называемых языков разметки текста (markuplanguages). Под термином "разметка" понимается общая служебная информация, которая не выводится вместе с документом, но определяет; как должны выглядеть те или иные фрагменты документа. Например, вы можете потребовать, чтобы какое-либо слово выводилось жирным или курсивным шрифтом, вывести отдельный абзац особым шрифтом или оформлять заголовки увеличенным шрифтом.

В наши дни существует множество разных языков разметки. Например, в коммуникационных программах особая форма разметки определяет смысл каждого пакета из нулей и единиц, пересылаемого в Internet. Впрочем, любой язык разметки должен решать две важные задачи:

- 1) язык определяет синтаксис разметки;
- 2) язык определяет смысл разметки.

Наиболее распространенным из языков разметки Web-страниц является HTML. Этот язык разметки был создан и рекламировался как одна из конкретизаций SGML. Впервые предложенный в 1974 году Чарльзом Голдфарбом в дальнейшем после значительной доработки принятый в качестве официального стандарта ISO, SGML (StandardGeneralizedMarkupLanguage, Стандартный обобщенный язык разметки) представляет собой метаязык – систему для описания других языков.

Появление стандарта SGML было обусловлено необходимостью совместного использования данных разными приложениями и операционными системами. Даже в далеких 60-х годах у пользователей компьютеров возникло немало проб

лем с совместимостью. Проанализировав недостатки многих нестандартных языков разметки, трое ученых из IBM — Чарльз Гольдфарб (Charles Goldfarb), Эд Мошер (Ed Mosher) и Рэй Лори (Ray Lorie) — сформулировали три общих принципа, обеспечивающих возможность совместной работы с документами в разных операционных системах.

1) Использование единых принципов форматирования во всех программах, выполняющих обработку документов. Вполне логичное требование — всем нам хорошо известно, как трудно договориться между собой людям, говорящим на разных языках. Наличие единого набора синтаксических конструкций и общей семантики заметно упрощает взаимодействие между программами.

2) Специализация языков форматирования. Благодаря возможности построения специализированного языка на базе набора стандартных правил программист перестает зависеть от внешних реалий и их представлений о потребностях конечного пользователя

3) Четкое определение формата документа. Правила, определяющие формат документа, задают количество и маркировку языковых конструкций, используемых в документе. Применение стандартного формата гарантирует, что пользователь будет точно знать структуру содержимого документа. Обратите внимание: речь идет не о формате отображения документа, а о его структурном формате. Набор правил, описывающих этот формат, называется "определением типа документа" (document type definition, DTD).

Эти три правила были заложены в основу предшественника SGML — GML (Generalized Markup Language). Исследования и разработка GML продолжались около десяти лет, пока в результате соглашения, заключенного международной группой разработчиков, не появился стандарт SGML.

HTML (HypertextMarkupLanguage, Язык разметки гипертекста) — это компьютерный язык, лежащий в основе WorldWideWeb. HTML основан на стандарте SGML гипертекстовый язык разметки документов для их представления в Web. Стандарты языка HTML, одного из ключевых стандартов Web, разрабатываются и поддерживаются консорциумом W3C. Основателем этого международного консорциума является Тим Бернерс-Ли (TimBerners-Lee). Консорциум помимо создания стандартов форматирования, является центром разработки SemanticWeb (семантическая сеть). Средствами языка HTML обеспечивается форматная разметка документов, определяются гиперсвязи между документами и/или их фрагментами.

В качестве основы написания кода HTML был выбран обычный текстовый файл. Таким образом, гипертекстовая база данных в концепции WWW — это набор текстовых файлов, размеченных на языке HTML, который определяет форму представления информации (разметка) и структуру связей между этими файлами и другими информационными ресурсами (гипертекстовые ссылки).

Разработчики HTML смогли решить две задачи:

- предоставить дизайнерам гипертекстовых баз данных простое средство создания документов;
- сделать это средство достаточно мощным, чтобы отразить имевшиеся на тот момент представления об интерфейсе пользователя гипертекстовых баз данных.

Первая задача была решена за счет выбора теговой модели описания документа. Язык HTML позволяет разметить электронный документ, который отображается на экране с полиграфическим уровнем оформления; результирующий документ может содержать самые разнообразные метки, иллюстрации, аудио- и видеофрагменты и так далее. В состав языка вошли развитые средства для создания различных уровней заголовков, шрифтовых выделений, различные списки, таблицы и многое другое.

Вторым важным моментом, повлиявшим на судьбу HTML, стало то, что в качестве основы был выбран обычный текстовый файл. Среда редактирования HTML является нейтральной полосой между простейшим текстовым файлом и приложением WYSIWYG (whatyouseeiswhatyouget – что вы видите, то и получаете). Выбор среды редактирования дает все преимущества текстового редактирования.

Гипертекстовые ссылки, устанавливающие связи между текстовыми документами, постепенно стали объединять самые различные информационные ресурсы, в том числе звук и видео. Система гиперссылок HTML позволяет построить систему взаимосвязанных документов по различным критериям. Язык HTML содержит команды (тэги), позволяющие управлять формой и размером шрифтов, размером и расположением иллюстраций, позволяет осуществлять переход от фрагмента текста или иллюстрации к другим html - документам – так называемую гипертекстовую ссылку. Документ в html - формате представляет собой текстовый файл, содержащий все необходимые сведения о выводимой на экран информации. Для управления сценариями просмотра страниц Website (гипертекстовой базы данных, выполненной в технологии World Wide Web) можно использовать языки

программирования этих сценариев, например, JavaScript, Java и VBScript. Формы для ввода пользователем данных, которые позднее подвергаются обработке и другую информацию можно обрабатывать с помощью специальных серверных программ (например, на языках PHP или Perl). Язык HTML позволяет помещать на страницы гипертекстовые ссылки и интерактивные кнопки, которые соединяют ваши Web-страницы с другими страницами того же Web-сайта, равно как и с другими Web-сайтами по всему миру.

HTML является языком разметки текста, а не языком программирования, который всего лишь один из инструментов (точнее, язык описания страниц), используемый при создании Web-страниц. В HTML ограничены возможности форматирования текста по сравнению с возможностями издательских программ, особенно при издании текста, насыщенного сложными элементами.

Анализируя особенности языка HTML и оценивая уровень его проработки, можно прийти к выводу, что уже в ближайшие годы следует ожидать появления более совершенных его модификаций, новых языков и прикладных пакетов для работы с web-страницами.

Динамический и статический HTML-документы

Различают два вида html-документов – статические и динамические. Статические документы хранятся в файлах той файловой системы, которая используется web-сервером или браузером при просмотре локальных файлов. При размещении информации на web-сервере можно использовать динамические документы – такие, которые не существуют постоянно в виде файлов, а генерируются в момент запроса клиента.

Для генерирования динамического документа HTML требуется специально написанная программа по правилам, определяемым web-сервером. Динамический способ определяет хранение данных в формализованном виде, например в базе данных. Для генерирования статических документов можно использовать любые средства отчетов, имеющиеся в той системе управления баз данных (СУБД), которой обработаны и формализованы данные.

Перспективы HTML

Новых версий языка HTML не будет, однако существует дальнейшее развитие HTML под названием XHTML (англ. Extensible Hypertext Markup Language — расширяемый язык разметки гипертекста). Пока XHTML по своим возможностям сопос

таким с HTML, однако предъявляет более строгие требования к синтаксису. Как и HTML, XHTML является подмножеством языка SGML, однако XHTML, в отличие от предшественника, соответствует спецификации XML.

2.2 Расширяемый язык разметки XML

Вторая половина 90-х годов прошедшего века ознаменовалась радикальными переменами в технологиях Web. Менее чем за пятилетнюю историю своего существования Web приобрел многие сотни миллионов пользователей на всех континентах, в его среде сформированы и поддерживаются огромные информационные ресурсы. Эта глобальная информационная система интенсивно вторгается в другие области информационных технологий, стала одним из важных звеньев инфраструктуры информационного общества.

Вместе с тем ряд ограничений, свойственных действующим технологиям Web (*Web первого поколения* или *Web-1*), стал сдерживающим фактором дальнейшего его развития. Новые подходы в области технологий Web, которые начали конструктивно воплощаться в жизнь на пороге XXI века, направлены, прежде всего на преодоление этих ограничений и создание технологической платформы[1], которая бы обеспечила потенциал для появления нового поколения Web (*Web второго поколения* или *Web-2*) и возможностей его развития на длительную перспективу. основополагающую роль в технологическом переоснащении Web стал играть разработанный консорциумом W3C новый язык разметки XML. Язык XML (Extensible Markup Language, расширяемый язык разметки) – это метаязык, являющийся подязыком SGML и определяющий процедуру порождения языков разметки для специфических целей.

Консорциум W3C, созданный для проведения единой технической политики в рамках Web и развития его технологий, ведет в настоящее время разработку и поддерживает более полутора сотен стандартов. Конечно же, невозможно представить их здесь в достаточно полном виде и приходится ограничиться лишь обсуждением концептуальных аспектов наиболее важных из них. Для основательного изучения стандартов платформы XML нужно обратиться к их оригинальным спецификациям и другим материалам консорциума W3C.

Следует заметить, что аббревиатуру XML довольно часто используют для обозначения не только самого языка XML, но и некоторых других связанных с ним понятий — определяющего языка стандарта W3C, информационных ресурсов XML,

комплекса основанных на языке XML стандартов консорциума W3C, составляющих платформу XML.

В то время как язык XML все чаще используют в среде Web по прямому своему назначению — как выразительное средство для представления информационных ресурсов в этой среде, он вместе с тем энергично внедряется в другие технологии. Развитые выразительные возможности языка, а главное, его поддержка механизмами среды Web позволяют использовать XML в качестве языка-посредника для определения форматов обмена данными между различными системами, которые используют Internet в качестве коммуникационной среды.

Главная сфера применения стандартов платформы XML — это представление слабоструктурированных данных[2] Web-сайтов в форме XML-документов. Собственно, для этой цели и создавался язык XML. Применение XML в этой области позволяет не только представлять в среде Web гипермедийные страницы в форме XML-документов, но и поддерживать связанные с ними метаданные[3]. Благодаря этому можно создать такие поисковые машины Web, которые будут обеспечивать в результате обработки пользовательского поискового запроса гораздо более низкий уровень информационного шума по сравнению с нынешними HTML-технологиями.

Одной из важнейших целей создания платформы XML является привнесение в среду Web метаданных, описывающих свойства поддерживаемых в ней информационных ресурсов[4]. Речь идет прежде всего об описании структуры XML-документов и их смыслового содержания (семантики). Необходимость решения этой задачи аргументируется стремлением к получению возможностей автоматической проверки правильности структуры XML-документов и снижения уровня информационного шума при отыскании нужных данных в Web с помощью различных поисковых машин. Имеется в виду, что при наличии явного описания структуры документов проверку их правильности может осуществлять браузер.

Однако чаще всего не учитывается еще одно важное назначение метаданных, описывающих информационные ресурсы Web. Метаданные необходимы для создания принципиально новых высокоуровневых приложений Web, в частности основанных на интеграции информационных технологий и обеспечивающих интеграцию неоднородных информационных ресурсов. Приведенный в приложении рис. 1. иллюстрирует упрощенную архитектуру системы, в которой метаданные используются для обеспечения интеграции неоднородных информационных ресурсов.

Пространства имен XML

Простейшая возможность задания семантики — использование пространства имен. В отличие от языка HTML, обеспечивающего форматную разметку текста, которая определяет его представление на экране, XML служит для структурной разметки.

Разметка в XML позволяет выделять в тексте содержательные структурные единицы, называемые элементами XML-документа. Для выделения каждого типа элементов используется свой тег, указывающий имя типа элемента. Поэтому с каждым таким тегом можно ассоциировать семантику соответствующих элементов XML-документа (адрес организации, номер телефона и т. д.).

Если некоторое сообщество разработчиков и пользователей XML-документов договорится о единой интерпретации имен, принадлежащих некоторому набору, то этот унифицированный набор, возможно, с каким-либо описанием их смысла (например, в виде обычного текста на естественном языке или представленный как им-либо иным образом), может использоваться как пространство имен. Адрес документа, представляющего в Web этот набор имен, будет рассматриваться как уникальный идентификатор пространства имен, и на него можно ссылаться в XML-документе, где используются принадлежащие этому пространству имена. И тем самым им придается некоторый смысл.

Примером достижения консенсуса о составе пространства имен является набор элементов метаданных для описания семантики представленных в Web документов, названный Дублинским ядром (Dublin Core, DC).

Дублинское ядро с принятой в нем семантикой элементов метаданных может использоваться в рамках платформы XML различными способами. Например, можно применять DC в качестве пространства имен для некоторого типа XML-документов или в RDF-спецификации (ResourceDefinitionFramework, стандарт схемы описания источников).

RDF-спецификации представляют собой более высокий уровень семантического описания информационных ресурсов. Информационные ресурсы в RDF — это ресурсы Web, идентифицируемые уникальным образом с помощью их URI (Uniform Resource Identifier, обобщение концепции URL в WWW). Они могут также представлять собой коллекции других информационных ресурсов или литералов, называемые контейнерами. Допускаются контейнеры типа мультимножества, последовательности и альтернативы.

Для того чтобы RDF-спецификация семантики информационных ресурсов была полной, необходимо ассоциировать с ней описание семантики используемых в этой спецификации свойств, которое в терминологии стандарта RDF называется схемой.

Метаданные, представленные средствами RDF, могут использоваться для более эффективного поиска ресурсов поисковыми машинами Web, в электронных библиотеках, в описаниях коллекций страниц Web, составляющих некоторый виртуальный документ, для представления содержания информационных ресурсов в конкретных предметных областях, а также для поддержки различных Web-приложений, нуждающихся в семантической информации о ресурсах.

В задачу RDF не входит стандартизация как их-либо наборов семантических свойств, и они могут быть различными в разных случаях.

В последнее время начали создаваться сервисы регистрации и поддержки пространств имен в интересах различных сообществ разработчиков и пользователей. Зарегистрированное пространство имен является своего рода стандартом для сообществ клиентов сервиса регистрации.

Перспективы XML

XML — отнюдь не модное направление, а естественный результат развития Web-технологий, следствие стремления к более эффективному использованию уникальных возможностей открытой глобальной информационной среды, которую они поддерживают. Создание платформы XML — это новая эпоха в развитии Всемирной паутины, это — начало нового, более наукоемкого и технологически более совершенного этапа в ее истории. Сегодня XML, несомненно, стал стандартом де-факто. Все ведущие поставщики программного обеспечения не только Web, но и систем баз данных, включают в свои программные продукты поддержку языка XML или даже создают специализированные системы.

Большую работу по продвижению стандартов платформы XML в практику ведет крупный Международный, некоммерческий консорциум OASIS в составе, которого более 600 корпоративных и индивидуальных членов из различных стран мира. Эта деятельность является основной задачей консорциума. OASIS разрабатывает, координирует разработки и распространяет информацию о методологиях применения, технологиях и реализациях этих стандартов. В его задачу входит также создание приложений для «вертикальной» индустрии (например, разработки описания типов документов (DocumentTypeDefinition, DTD), схем XML и

пространств имен XML), спецификаций интероперабельности (в частности, создание спецификаций профилей, включающих стандарты рассматриваемой категории), тестов на соответствие рассматриваемым стандартам.

Распространению стандартов XML-платформы существенным образом способствует политика W3C, направленная на обеспечение доступности их спецификаций, создание ряда свободно распространяемых синтаксических анализаторов для языка, то большое внимание, которое создатели стандартов XML уделяют обеспечению преемственности для существующей HTML-платформы и накопленных на ее основе ресурсов.

Хотя язык XML и базирующиеся на нем стандарты получают все более широкое распространение, имеются вместе с тем факторы, которые сдерживают массовое распространение XML в среде Web.

Во-первых, существует связанная с экономическими и иными причинами естественная инерционность столь масштабной среды, как ой является сегодняшний Web. Эта инерция может преодолеться только постепенно.

Во-вторых, пока еще не завершена работа над двумя важнейшими стандартами платформы XML, которые позволяют строить из отдельных XML-документов и их компонентов гипермедийную среду. Речь идет о стандартах XPointer (XMLPointerLanguage, язык указателей XML) и XLink (XMLLinkingLanguage, язык ссылок XML). Эти стандарты решают задачу определения гиперссылок в языке XML. Возможности стандартов XPointer и XLink предусматривают существенно более богатые возможности работы с гиперссылками, чем у имеющихся в HTML.

Технологии XML начинают распространяться и в нашей стране. В этой связи приобретает важное значение русскоязычная терминология в этой области.

Платформа XML имеет благоприятные перспективы для широкого практического применения. В пользу этого свидетельствуют не только богатые функциональные возможности рассмотренного семейства стандартов, но и высокая активность в области разработки и развития стандартов, а также производства программного обеспечения, на них основанного.

Расширяемый язык разметки гипертекста XHTML

Активное распространение технологий XML порождает весьма острую проблему обеспечения преемственности в развитии среды Web, создания возможностей, позволяющих использовать огромные информационные ресурсы, накопленные в рамках технологий HTML.

Один из подходов к решению этой проблемы реализован в стандарте XHTML 1.0 (The Extensible Hypertext Markup Language, расширяемый язык разметки гипертекста), одобренном W3C в январе 2000 г. Как и HTML, XHTML является подмножеством языка SGML, однако XHTML, в отличие от предшественника, соответствует спецификации XML.

Идея предлагаемого подхода заключается в создании на основе XML языка разметки, эквивалентного по его функциональности языку HTML. Аналогичным образом в настоящее время разработаны многие другие языки разметки конкретизации XML. Создание таких языков сводится, по существу, к разработке соответствующего определения типа документов (DTD).

Авторы стандарта XHTML 1.0 трактуют функцию определяемого в нем языка как реформировку HTML в XML (более точно, речь идет о принятом W3C в декабре 1999 г. стандарте HTML 4.01 и об XML 1.0).

XHTML является преемником HTML. Потребность в более строгой версии HTML возникла из-за того, что веб-контент сегодня всё больше становится ориентированным на нетрадиционные виды устройств (например, мобильные телефоны), в которых зачастую ограничены ресурсы, в том числе и для обработки гибкого, нетребовательного HTML (чем свободнее синтаксис языка, тем сложнее его разбирать).

Практически все современные браузеры поддерживают XHTML. Он также совместим и со старыми браузерами, т. к. в основе XHTML лежит HTML. Такая совместимость, к сожалению, в числе прочего, замедляет процесс перехода от HTML к XHTML.

Настоящая сила XHTML проявляется в его сочетании с каскадными таблицами стилей. Это позволяет отделить оформление документа от его содержимого.

Отличия переходного (англ. transitional) XHTML от HTML незначительны и предназначены лишь для приведения его в соответствие с XML. Самое главное требование заключается в том, чтобы все теги были правильно вложены и семантически развиты. Кроме того, в XHTML все теги должны записываться строчными буквами,

все атрибуты (включая численные) должны быть заключены в кавычки (что не является обязательным в SGML и, следовательно, в HTML, где кавычки не требуются для чисел и некоторых символов, включая все буквы). Также все элементы должны быть закрыты, включая те, которые не имеют закрывающего тега (закрываются добавлением слэша ('/') в конец тега). Минимизация атрибутов (к примеру `<option selected>` или `<td nowrap>`) также воспрещена. Детальнее об отличиях можно узнать из спецификации XHTML <http://www.w3.org/>.

В стандарте XHTML предлагается три варианта целевого языка для представления HTML-документов и тем самым *три версии DTD*:

- *XHTML Strict* (строгий XHTML);
- *XHTML Transitional* (переходный XHTML);
- *XHTML Frameset* (XHTML с фреймами).

Вариант XHTML Strict полностью отделяет содержание документа от оформления, многие атрибуты (такие как, например, `bgcolor` и `align`) более не поддерживаются. Предназначен для чисто структурной разметки без применения элементов форматирования. Для целей форматирования можно при этом дополнительно использовать язык каскадных таблиц стилей (CSS).

Вариант XHTML Transitional будет, вероятно, наиболее популярным. Он допускает использование таблиц стилей, но имеется в виду, что будет внесена некоторая небольшая коррекция в разметку с тем, чтобы документ мог восприниматься и старыми браузерами, которые не поддерживают таблиц стилей. Предназначен для лёгкой миграции из HTML и для тех, кто использует инлайн-фреймы.

Вариант XHTML Frameset обеспечивает поддержку фреймов. Это позволяет разбить окно браузера на несколько разделов (фреймов), с которыми ассоциируется некоторый набор функций управления.

Но существуют и специализированные дополнительные версии XHTML:

XHTML 1.1 Модульный (Module-based): авторы могут импортировать дополнительные свойства в их разметку. Эта версия также поддерживает руби-разметку, необходимую для дальневосточных языков.

XHTML Основной (Basic): специальная облегчённая версия XHTML для устройств, которые не могут использовать полный набор элементов XHTML — в основном используется в миниатюрных устройствах, таких как мобильные телефоны.

Подразумевается, что он заменит WML и С-HTML.

XHTML мобильного профиля (Mobile Profile): основанный на XHTML Basic, добавляет специфические элементы для мобильных телефонов. Он является еще одним шагом вперед на пути к мобильным сервисам 3G. XHTML дает пользователям доступ к полноценному контенту, который отлично выглядит и имеет удобную навигацию. В сочетании с увеличенной скоростью, предлагаемой GPRS, мобильные сервисы становятся более притягательными и похожими на обычный Internet.

Хотя гиперссылки на документы, содержащие определения вариантов DTD для XHTML Strict, XHTML Transitional и XHTML Frameset, приводятся в приложении к стандарту, именно эти спецификации составляют основное его содержание.

В стандарте XHTML значительное внимание уделено вопросам поддержки развития языка HTML. Предполагается, что XHTML 1.0 определяет начальную версию развивающегося семейства типов документов, которые позволяют воспроизводить HTML, выделять его подмножества и расширять этот язык.

Обеспечение модульности языка воплощается в проекте новой версии стандарта — XHTML 1.1.

Прародителями XHTML 2.0 являются HTML 4, XHTML 1.0 и XHTML 1.1, но он не рассчитан на обратную совместимость с ними. Кроме того, первый Рабочий проект пока не включает реализации XHTML 2.0 ни в форме DTD, ни в виде XML-схемы. Эти реализации будут включены на более поздних этапах, как только будут урегулированы принципиальные вопросы.

В спецификации XHTML вводится специальное пространство имен XHTML. Однако для именования элементов и атрибутов в документах допускается использование наряду с ним также и других пространств имен, например пространства имен MathML (математический язык разметки) или RDF (Resource Definition Framework).

Пространство имен XHTML

XHTML более доступен, так как он использует пространство имен практически идентичное с HTML 4.01 и, таким образом, большая часть DTD уже "зашита" в браузер. DTD (Document Type Definition) критически важен для XML-документов. Другой вопрос связанный с XML: пространство имен этого языка очень велико и определяется именно DTD написанным специально для конкретной его разновидности. Для браузера во время разбора соответствующего XML-кода эти

документы должны быть доступны. Преимущество XHTML module в том что пространство имен может быть сокращено до того списка который вы используете на сайте.

Валидация XHTML документов

Валидным (т.е. отвечающим всем правилам) XHTML-документом считается документ, удовлетворяющий технической спецификации. В идеале, все браузеры должны следовать веб-стандартам и, в соответствии с ними, валидные документы должны отображаться во всех браузерах на всех платформах. Валидация XHTML-документа рекомендована даже несмотря на то, что она не гарантирует кросс-браузерной совместимости. Документ может быть проверен на соответствие спецификации с помощью онлайн-Службы валидации разметки W3C. Валидация обнаружит и разъяснит ошибки в XHTML-разметке.

Валидный документ должен содержать определение типа документа (DTD). DTD должен быть расположен до всех других элементов документа.

Валидный XHTML-документ, по правилам W3C, может быть снабжён специальным баннером, подтверждающим правильность XHTML-разметки.

Авторы стандарта рассматривают миграцию к XHTML как следующий шаг в эволюции Web-технологий. В настоящее время W3C продолжает работу по развитию XHTML. Одной из важных задач при этом признается обеспечение модульности языка.

Глава 3. СПЕЦИФИКАЦИИ ТЕХНОЛОГИЙ WEB

О спецификации HTML

Спецификация состоит из следующих разделов.

Во введении описывается место языка HTML в схеме World Wide Web, приводится краткая история развития языка HTML, описывается, что можно сделать с использованием HTML 4.0 и содержатся некоторые подсказки относительно создания документов в формате HTML.

Краткое руководство по SGML дает читателям понимание отношения языка HTML к языку SGML и предоставляет информацию о типах Определений типов документов HTML (Document Type Definition - DTD).

Главным содержанием руководства является справочник по языку HTML, в котором определены все элементы и атрибуты языка.

Этот документ упорядочен по разделам, а не по грамматике языка HTML. Разделы сгруппированы в три категории: структура, представление и интерактивность. Хотя конструкции языка HTML трудно разделить на эти три категории, такая модель отражает опыт Рабочей группы HTML, говорящий о том, что разделение структуры документа и его представления обеспечивает большую эффективность документов и лучшие возможности поддержки.

Информация о языке включает следующую:

- Какие символы могут отображаться в документе HTML.
- Основные типы данных документа HTML.
- Элементы, управляющие структурой документа HTML, включая текст, списки, таблицы, ссылки и объекты, изображения и апплеты.
- Элементы, управляющие представлением документа в формате HTML, включая таблицы стилей, шрифты, цвета, горизонтальные разделители и другое визуальное представление, а также фреймы (кадры) для многооконного представления.
- Элементы, управляющие интерактивностью документа HTML, включая формы для ввода данных пользователя и скрипты для активных документов.
- Формальное SGML-определение HTML: SGML-определение HTML;

три DTD: строгое, переходное и с кадрами; список ссылок на символы.

В первом приложении содержится информация об изменениях по отношению к HTML 3.2 с целью помочь авторам при переносе файлов в формат HTML 4.0. Во втором приложении содержатся замечания о производительности и применении, целью которых является помощь разработчикам в создании средств для использования HTML 4.0.

Список нормативных и информативных документов.

Три указателя предоставляют читателям быстрый доступ к определению: понятия, элементы и атрибуты.

Этот документ написан читателями с двумя типами мышления: авторами и разработчиками. Мы надеемся, что спецификация предоставит авторам средства, необходимые им для создания эффективных, привлекательных и доступных документов и не обременяющие их подробностями применения HTML. Разработчики, однако, должны найти здесь всю необходимую для разработки соответствующих средств информацию.

Эту спецификацию можно использовать несколькими способами:

Прочитать от начала до конца. Эта спецификация начинается с общего представления языка HTML, а количество технических подробностей постепенно повышается.

Обращаться к необходимой информации. Для обеспечения максимальной скорости получения информации о синтаксисе и семантике в оперативную версию спецификации включены следующие возможности:

Каждая ссылка на элемент или атрибут связана с его определением в спецификации. Каждый элемент или атрибут определяется только в одном месте.

На каждой странице имеются ссылки на указатели, поэтому Вы всегда сможете найти определение элемента или атрибута, используя не больше двух ссылок.

На первых страницах трех разделов руководства к исходному оглавлению добавляется более подробная информация о каждом разделе.

Названия элементов представляются символами в верхнем регистре (например, BODY). Названия атрибутов представляются символами в нижнем регистре (например, lang, onsubmit). Помните, что в HTML имена элементов и атрибутов не учитывают регистр; это используется для более легкого чтения.

В названиях элементов и атрибутов в этом документе используется разметка, поэтому агентами пользователей они могут генерироваться особым образом.

В каждом определении атрибута устанавливается тип его значения. Если имеется несколько возможных значений, приводится список значений, разделенных вертикальной чертой (|).

После информации о типе в каждом определении атрибута в квадратных скобках ("[]") указывается, учитываются ли в значениях регистр. Подробнее см. раздел информации о регистре.

Информативные замечания выделены, чтобы отличаться от остального текста и могут генерироваться агентами пользователей особым образом.

Все примеры, иллюстрирующие нежелательное использование, помечены как "ПРИМЕР НЕЖЕЛАТЕЛЬНОГО ИСПОЛЬЗОВАНИЯ". В примеры нежелательного использования входят также рекомендуемые альтернативные решения. Все примеры, иллюстрирующие недопустимое использование, помечены как "ПРИМЕР НЕДОПУСТИМОГО ИСПОЛЬЗОВАНИЯ".

В примерах и замечаниях используется разметка, поэтому некоторыми агентами пользователей они могут генерироваться особым образом.

О спецификации XML

Расширяемый Язык Разметки (XML) является поднабором SGML и полностью описан в спецификации. Он создан с целью обеспечения обслуживания, передачи и обработки в WEB исходного SGML теми же способами, которые в данный момент имеются в HTML. XML был разработан для облегчения создания конкретных реализаций и для взаимодействия с SGML и HTML.

Роль W3C в составлении Рекомендаций заключается в том, чтобы привлечь внимание к данной спецификации и способствовать её широкому распространению. Это расширит функциональность и возможности Web.

Этот документ специфицирует синтаксис, создаваемый путём подразделения существующих широко распространённых международных стандартов обработки текста для использования в World Wide Web.

Extensible Markup Language, сокращённо XML, описывает класс объектов данных, называемых XML-документы, и частично описывает поведение обрабатывающих их компьютерных программ. XML является профилем приложения или ограниченным вариантом SGML - The Standard Generalized Markup Language. По структуре документы XML являются "соответствующими" документами SGML.

Документы XML состоят из единиц хранения, называемых экземпляры, которые содержат разбираемые или неразбираемые данные.

Разбираемые данные состоят из символов, некоторые из которых образуют символьные данные, а другие - разметку. Разметка кодирует описание схемы и

логической структуры единиц хранения документа. XML предоставляет механизм наложения ограничений на схему и логическую структуру единиц хранения.

XML был разработан XML Working Group (ранее известной как SGML Editorial Review Board), сформированной под руководством World Wide Web Consortium (W3C) в 1996 году.

Её возглавил JonBosak из SunMicrosystems при активном участии XMLSpecialInterestGroup (ранее известной как SGMLWorkingGroup), также организованной W3C. Члены XML Working Group указаны в Приложении. Dan Connolly является контактным лицом рабочей Группы с W3C.

Цели создания XML:

1. XML будет широко распространён в Internet.
2. XML будет поддерживать большой диапазон приложений.
3. XML будет совместим с SGML.
4. Он будет лёгким для написания программ, обрабатывающих документы XML.
5. Количество свойств по выбору (optional) в XML будет сведено к абсолютному минимуму, в идеале - к нулю.
6. Документы XML должны быть разборчивыми и ясными по смыслу.
7. Дизайн XML должен выполняться быстро.
8. Дизайн XML должен быть формальным и кратким.
9. Документы XML должны легко создаваться.
10. Краткость в разметке XML имеет минимальное значение.

Эта спецификация, вместе с ассоциированными стандартами, предоставляет всю информацию, необходимую для понимания XML и создания компьютерных программ его обработки.

Символы – это разбираемый экземпляр содержит текст, последовательность символов, которая может представлять символичные данные или разметку. Текст состоит из смеси символических данных и разметки.

Комментарии могут появляться в любом месте документа вне прочей разметки; кроме того, они и могут появляться внутри объявления типа документа в тех местах, которые допускаются грамматикой. Они не являются частью символьных данных документа: процессор XML может, но не должен, давать приложению возможность запрашивать текст комментариев.

Инструкции процесса (ИП) позволяют вводить в текст документа инструкции для приложений. Разделы CDATA могут появляться там же, где и символьные данные; они используются для escape-блоков текста, содержащего символы, которые иначе будут распознаваться как разметка.

Документы XML должны начинаться объявлением XML, которое специфицирует используемую версию XML.

Поскольку будущие версии ещё не сформированы, эта конструкция даётся как средство предоставления возможности автоматического распознавания версии и должна, следовательно, быть включена обязательно. Процессоры могут сигнализировать об ошибке, если получат документ, помеченный неподдерживаемой версией.

Функцией разметки в документе XML является обязанность описывать структуру хранения данных и логическую структуру и ассоциировать пары атрибут-значение с их логическими структурами. XML предоставляет механизм объявления типа документа для определения ограничений в логической структуре и для поддержки использования предопределённых единиц хранения.

Документ XML является правильным/valid, если он имеет ассоциированное объявление типа документа и если документ выполняет ограничения, выраженные в нём.

Объявление типа документа XML содержит или указывает на объявления разметки, предоставляющие грамматику для класса документов. Эта грамматика известна как определение типа документа или DTD. Объявление типа документа может указывать на внешний поднабор (особый вид внешнего экземпляра), содержащий объявления разметки, или может непосредственно содержать объявления разметки во внутреннем поднаборе, или может иметь и то, и другое. DTD документа состоит из обоих соединённых поднаборов. Объявление разметки это объявление типа элемента, объявление списка атрибутов и объявление экземпляра, или объявление нотации. Эти объявления могут полностью или частично содержаться внутри экземпляров параметров.

Каждый документ XML содержит один или более элементов, ограниченных либо начальными и конечными тэгами, либо -для пустых элементов - тэгами пустых элементов. Каждый элемент имеет тип, идентифицируется по имени, которое иногда называется "generic identifier" (GI) - родовой идентификатор, и может иметь набор спецификаций атрибутов. Каждая спецификация атрибутов имеет имя и значение.

Начало каждого непустого элемента XML обозначается начальным тэгом. Окончание каждого элемента, начатого начальным тэгом, обязано быть отмечено конечным тэгом, содержащим имя, отражающее тип элемента, как это было дано в начальном тэге. Текст между начальным и конечным тэгами называется содержимым элемента.

Элемент без содержимого называется пустым. Пустой элемент представлен либо начальным тэгом, после которого непосредственно следует конечный тэг, либо тэгом пустого элемента. Тэг пустого элемента имеет особую форму.

Структура элемента документа XML может, для целей проверки, быть ограничена путём использования объявлений типа элемента и списка атрибутов. Объявление типа элемента ограничивает содержимое элемента.

Объявление типа элемента часто ограничивают типы элементов, которые могут появляться в качестве потомков элемента.

Тип элемента имеет содержимое элемента, если элементы данного типа обязаны содержать только дочерние элементы (а не символные данные), которые могут быть, по усмотрению, разделены пробелами.

В этом случае ограничение включает модель содержимого, простую грамматику, управляющую разрешёнными типами дочерних элементов и порядком, в котором они могут появляться.

Тип элемента имеет смешанное содержимое, если элементы этого типа могут содержать символные данные, перемежаемые дочерними (необязательными) элементами.

Атрибуты используются для ассоциирования пар имя-значение с элементами. Спецификации атрибутов могут появляться только в начальных тэгах и тэгах пустых элементов; поэтому продукты, используемые для их распознавания, появляются в разделе.

Прежде чем значение атрибута передается приложению и проверяется на правильность, процессор XML обязан нормализовать значение атрибута путём применения к нему нижеприведённого алгоритма или путём использования некоторых других методов так, чтобы значение, передаваемое приложению, было тем же, что и произведённое алгоритмом.

Документ XML может состоять из одной или более единиц хранения. Они называются экземплярами; они имеют содержимое и все (исключая экземпляр документа и внешний поднабор ОТД) идентифицируются по имени экземпляра. Содержимое разбираемого экземпляра называется его замещающим текстом; этот текст считается неотъемлемой частью документа.

Неразбираемый экземпляр это ресурс, чьё содержимое может, или может не быть, текстом, и, если это текст, может не быть XML. Каждый неразбираемый экземпляр имеет ассоциированную нотацию, идентифицируемую по имени. Помимо требования к процессору XML сделать идентификаторы экземпляра и нотации доступными приложению, XML не накладывает никаких ограничений на содержимое неразбираемых экземпляров.

Общие экземпляры это экземпляры для использования внутри содержимого документа. В этой спецификации ОЭ иногда называются неквалифицированным термином экземпляр, если это не приводит к неоднозначности.

Экземпляры параметров это разбираемые экземпляры для использования внутри ОТД. Эти два типа экземпляров используют разные формы ссылок и распознаются в различных контекстах. Следовательно, они занимают разные пространства имён; экземпляр параметра и общий экземпляр с одним именем - это два разных экземпляра.

Ссылка символа ссылается на специфический символ в наборе символов ISO/IEC 10646, например, ссылка на символ, не доступный напрямую из устройства ввода. Ссылка экземпляра ссылается на содержимое именованного экземпляра.

Если процессор XML обнаруживает ссылку на разбираемый экземпляр, то, для того чтобы проверить документ, процессор обязан включить его (экземпляра) замещающий текст. Если экземпляр является внешним, а процессор не пытается проверить документ XML, то процессор может, но это не является необходимым, включить замещающий текст экземпляра. Если непроверяющий процессор не включает замещающий текст, он обязан информировать приложение, что он обнаружил, но не прочитал, экземпляр.

Это правило базируется на том, что автоматическое распознавание, предоставляемое механизмом экземпляров SGML и XML, первоначально созданным для поддержки модульности в авторизации, не обязательно подходит для других приложений, особенно для просмотра документов. Браузеры, например, при обнаружении ссылки на внешний разбираемый экземпляр, могут избрать визуальное предупреждение о том, что экземпляр существует, и запрашивать его для показа только по требованию.

Литеральное значение экземпляра это завышенная строка, реально представленная в объявлении экземпляра, соответствующая нетерминальному EntityValue. Определение: Замещающий текст это содержимое экземпляра после замещения мнемоник символов и ссылок экземпляров параметров.

Нотации идентифицируют по имени формат неразбираемых экземпляров, формат элементов, которые породили атрибут нотации, или приложение, которому адресуется инструкция процесса. Объявления нотации предоставляют имя нотации для использования в объявлениях экземпляра и списка атрибутов и в спецификациях атрибутов, а также внешний идентификатор для нотации, который может позволить процессору XML или его клиентскому приложению локализовать вспомогательное приложение, способное обработать данные в данной нотации.

Соответствующие процессоры XML делятся на два класса: проверяющие и не проверяющие. Проверяющие и не проверяющие процессоры оба обязаны выводить сообщения о нарушениях ограничений правильности сформированности данной спецификации в содержимом экземпляра документа и любых других разбираемых экземплярах, которые они читают.

Проверяющие процессоры обязаны, по выбору пользователя, сообщать о нарушениях ограничений, выраженных объявлениями в ОТД, и невозможности выполнения ограничений правильности, данных в этой спецификации. Чтобы выполнить это, проверяющие процессоры XML обязаны читать и обрабатывать все ОТД и все внешние разбираемые экземпляры, на которые имеются ссылки в документе.

От не проверяющих процессоров требуется лишь проверить экземпляр документа, включая весь внутренний поднабор ОТД, на правильное формирование.

Поскольку не требуется проверять документ на правильность/верность, необходимо обработать все объявления, прочитанные во внутреннем поднаборе ОТД и во всех экземплярах параметров, которые прочитаны, до первой ссылки на

экземпляр параметра, который не прочитан; то есть информация в этих объявлениях обязана использоваться для нормализации значений атрибутов, включения замещающего текста внутренних экземпляров поддержки значений по умолчанию в атрибутах.

Формальная грамматика XML даётся в данной спецификации с использованием нотации Extended Backus-Naur Form (EBNF). Каждое правило грамматики определяет один символ.

3.3 О спецификации XHTML

В настоящей спецификации определяется XHTML 1.0, переформулировка HTML 4 в виде приложения XML 1.0, и три DTD, соответствующих типам, определяемым HTML 4. Семантика элементов и их атрибутов определена в рекомендации W3C HTML 4. Данная семантика представляет собой основу для будущего расширения языка XHTML.

XHTML представляет собой семейство имеющихся на данный момент и могущих появиться в будущем типов документов и модулей, являющихся копиями, подмножествами или расширениями языка HTML 4 [HTML]. Семейство типов документов XHTML базируется на XML и предназначено для работы с пользовательскими агентами на базе. Более подробную информацию об этом семействе и его эволюции можно найти в разделе "Направления развития".

Семейство XHTML является следующим шагом в эволюции Интернет. Переходя сегодня на XHTML, разработчики содержимого (контента) могут вступить в мир XML со всеми его преимуществами, сохраняя при этом совместимость содержимого с более старыми и более новыми версиями.

Преимущества перехода на XHTML 1.0 описаны выше. Вот несколько основных преимуществ:

Разработчики документов и создатели пользовательских агентов постоянно открывают новые способы выражения своих идей в новой разметке. В XML ввод новых элементов или атрибутов достаточно прост. Семейство XHTML разработано так, чтобы принимать расширения путем модулей и технологий XHTML для разработки новых соответствующих XHTML модулей (описанных в готовящейся спецификации Модуляризации XHTML). Модули позволят комбинировать существующие и новые

наборы функций при разработке содержимого и создании новых пользовательских агентов.

Постоянно вводятся альтернативные методы доступа в Интернет. По некоторым оценкам, в 2002 году 75% обращений к документам в Интернет будет выполняться с альтернативных платформ. Семейство XHTML создавалось с учетом общей совместимости пользовательских агентов. С помощью нового механизма профилирования пользовательских агентов и документов серверы, прокси и пользовательские агенты смогут преобразовывать содержимое наилучшим образом. В конечном счете станет возможной разработка соответствующего XHTML содержимого, пригодного для любого соответствующего XHTML пользовательского агента.

В настоящей спецификации используются следующие термины, которые расширяют определения, данные в [RFC2119] аналогично определениям ISO/IEC 9945-1:1990 [POSIX.1]:

Описываются общие термины XHTML: атрибут, DTD, возможности, документ, пользовательский агент, правильно построенный, представление (генерация), проверка корректности, реализация, синтаксический разбор, элемент.

В настоящей версии XHTML предоставляется определение строго конформных документов XHTML.

Строго конформный документ XHTML - это документ, которому необходимы только возможности, описанные в настоящей спецификации как обязательные.

Пространство имен XHTML может использоваться с другими пространствами XML в соответствии с [XMLNAMES], хотя такие документы не являются строго конформными XHTML 1.0 в соответствии с приведенным выше определением. В будущих работах W3C будут определены способы указания конформности документов, в которых используется несколько пространств имен.

Конформный пользовательский агент должен соответствовать всем определенным в спецификации критериям.

Говорится о различиях которые присутствуют в языке XHTML. Поскольку XHTML является приложением XML, некоторые приемы, допустимые в языке HTML, основанном на SGML, должны быть изменены.

К документам XHTML 1.0 не предъявляется требование совместимости с существующими пользовательскими агентами, но на практике оно достаточно легко реализуемо.

Спецификация XHTML 1.0 закладывает основу семейства типов документов, которые будут расширениями и подмножествами XHTML, для поддержания широкого диапазона новых устройств и приложений путем определения модулей и механизма объединения этих модулей. Такой механизм позволит унифицировать способы расширения XHTML 1.0 и использования его подмножеств путем определения новых модулей.

По мере перемещения XHTML с традиционных пользовательских агентов на рабочем столе на другие платформы становится ясно, что не все элементы XHTML будут необходимы на всех платформах.

Процесс модуляризации разбивает XHTML на ряд более мелких подмножеств элементов. Модуляризация дает определенные преимущества.

В профиле документа определяется синтаксис и семантика набора документов. Соответствие профилю документа обеспечивает основу гарантии совместимости. В профиле документа определяются возможности, необходимые для обработки документа этого типа.

Для авторов профилей устраняют необходимость написания нескольких различных версий документов для различных клиентов.

ЗАКЛЮЧЕНИЕ

Создание Web поправу можно считать одним из крупнейших научно-технических достижений после десятилетия XX века. Благодаря реализации этого проекта рождается целый ряд новых информационных технологий, имеющих весьма значимые социально-экономические последствия.

Одним из наиболее распространенных классов систем обработки данных являются информационные системы. В настоящее время усиливается тенденция глобализации ИС.

Современные информационные Web-технологии быстро изменяют наш мир и непосредственно влияют на развитие Web-технологий. Эта технологическая революция сильно повлияла на все сферы человеческой деятельности. Внутренняя сложность и предельная простота применения современных информационных Web-технологий делает их доступными каждому, кто ежедневно сталкивается с применением их в своей профессиональной деятельности.

Главное преимущество Web-технологий в современных условиях заключается в их простоте и как следствие в повышении эффективности их применения.

В этой курсовой работе мы попытались показать обширную проблематику технологий современных информационных систем, технологий Web, появившихся тенденций в развитии технологий Web. В данной работе проанализированы необходимые для данной курсовой работы спецификации и разработан фрагмент информационной системы «Учебно – методический ресурс».

БИБЛИОГРАФИЯ

1. Баранов, Д.В. Современные информационные технологии. / Д.В. Баранов. – Томск: ИДО (ТУСУР), 2005. – 130 с.
2. Ваулина, Ч.Ю. Информатика: толковый словарь / Ч.Ю. Ваулина. – М.: Изд-во Эксмо, 2005. – 480 с.
3. Когаловский, М.Р. Перспективные технологии информационных систем / М.Р. Когаловский. – М.: Компания АйТи, 2003. – 288 с.
4. Когаловский, М.Р. Энциклопедия технологий баз данных / М.Р. Когаловский. – М.: Финансы и статистика, 2005. – 800 с.
5. Крис, Д. Креативный Web-дизайн. HTML, XHTML, CSS, JavaScript, PHP, ASP, ActiveX. Текст, графика, звук и анимация. Учебник Пер с англ. / Д. Крис, К. Кинг, Э. Андерсон. – М.: ООО «ДиаСофтЮП», 2005. 672 с.
6. Мишенин, А.И. Теория экономических информационных систем / А.И. Мишенин. – М.: Финансы и статистика, 2002. – 240 с.
7. Непейвода, Н.Н. Основания программирования / Н.Н. Непейвода, Скопин И.Н. – Москва-Ижевск: Институт компьютерных исследований, 2003. – 868 с.
8. Основы Web – технологий : учеб. пособие / П.Б. Храмцов [и др.]. – М. : Изд-во Инт уит.ру “Интернет-Университет Информационных Технологий”, 2003. – 512 с.

9. Пауэл Томас, А. Справочник программиста / Томас А Пауэл, Д. Уитворт. – М.: АСТ, Мн.: Харвест, 2005. – 384 с.
10. Петров, В.Н. Информационные системы: учеб. пособие / В.Н. Петров. – СПб.: Питер, 2002. – 588 с.
11. Экономическая информатика: Введение в экономический анализ информационных систем: учебник. – М.: ИНФРА-М, 2005. – 958 с. – (Учебники экономического факультета МГУ и м. М.В. Ломоносова).
12. Когаловский М.Р. XML: возможности и перспективы [Электронный ресурс] / сост. и ред. М.Р. Когаловский – OSP.RU:Издательство "Открытые системы", [2001]. <<http://www.osp.ru/cio/2001/02/016.htm>> (15.02.2001).
13. Когаловский М.Р. XML: сферы применений [Электронный ресурс] / сост. и ред. М.Р. Когаловский – OSP.RU:Издательство "Открытые системы", [2001]. <<http://www.osp.ru/cio/2001/04/010.htm>> (17.04.2001).
14. Когаловский М.Р. Развитие стандартов XML: новые возможности и применения [Электронный ресурс] : (Материалы Второй всероссийской конференции «Стандарты в проектах современных информационных систем») [тез. докл.] / М.Р. Когаловский и Институт проблем рынка РАН <<http://www.cemi.rssi.ru/mei/articles/conf02st.htm>> (27-28.03.2002).
15. Международная организация OASIS: сообщество рабочих групп Журнал: Intersoft Lab [Электронный ресурс] Новая деловая газета: CitCity [Web-сайт]. 24.02.1997// <<http://www.citforum.ru/internet/xml/oasis/>> (09.04.2003).
16. Международный консорциум W3C: от Рабочего проекта до Рекомендации Журнал: Intersoft Lab [Электронный ресурс] Новая деловая газета: CitCity [Web-сайт]. 24.02.1997// <<http://www.citforum.ru/internet/xml/w3c/>> (08.04.2003).
17. Российские Электронные Библиотеки Extensible Hypertext Markup Language (XHTML) [Электронный ресурс] / Аннотированный указатель стандартов платформы XML <http://www.elbib.ru/index.phtml?page=elbib/rus/methodology/xmlbase/standarts_2002/XHTML> (13.01.2006).
18. Российские Электронные Библиотеки Hypertext Markup Language (HTML)[Электронный ресурс] / Аннотированный указатель стандартов платформы XML) / <http://www.elbib.ru/index.phtml?page=elbib/rus/methodology/xmlbase/standarts_2002/HTML> (23.12.2003).
19. Extensible Markup Language (XML) [Электронный ресурс] / Спецификация <<http://www.w3.org/XML/>>.
20. OASIS Advancing E-Business Standard Since [Электронный ресурс] /<<http://www.oasis-open.org/>> (1993).

21. SimpleSolutions – разработка и сопровождение Web-сайтов. Введение в учебник HTML, немного истории [Электронный ресурс] / Web-студия Ретюхина Александра <<http://www.sasha.by/doc2.php?page=html&theme=0>> (2000-2006).
22. SimpleSolutions – разработка и сопровождение Web-сайтов. PHP и XML. [Электронный ресурс] / Web-студия Ретюхина Александра <<http://www.sasha.by/doc2.php?page=html&theme=1>> (2000-2006).
23. SimpleSolutions – разработка и сопровождение Web-сайтов. Основные понятия [Электронный ресурс] / Web-студия Ретюхина Александра <<http://www.sasha.by/doc2.php?page=html&theme=1>> (2000-2006).
24. World Wide Web Consortium W3C, World Wide Web, Web, WWW, Consortium, computer, access, accessibility, semantic, worldwide, W3, HTML, XML, standard, language, technology, link, CSS, RDF, XSL, Berners-Lee, Berners, Lee, style sheet, cascading, schema, XHTML, mobile, SVG, PNG, PICS, DOM, SMIL, MathML, markup, Amaya, Jigsaw, free, open source, software [Электронный ресурс] / <<http://www.w3.org/>> (03.1989).
25. XHTML 1.0: The Extensible HyperText Markup Language (Second Edition) [Электронный ресурс] / Спецификация <<http://www.w3.org/TR/xhtml1>> (26.01.2000, 01.08.2002).
26. HTML 4.01 Specification [Электронный ресурс] / Спецификация <<http://www.w3.org/TR/html4> > (24.12.1999).

ПРИЛОЖЕНИЕ

Приложение WEB

Реляционное приложение

Механизм доступа XML

Сервер баз данных с расширением для XML

XML - документы

БАЗА данных SQL

Метаданные XML

SQL - схема базы данных

Преобразователь: метаданные XML⇒SQL

Преобразователь: метаданные SQL⇒XML

Объединенные метаданные XML

Объединенная схема SQL

Рис. 1. Упрощенная архитектура системы, обеспечивающей интеграцию информационных ресурсов XML и SQL

1. **Платформа** — целенаправленно разработанная для решения некоторых задач совокупность технологий и поддерживающих их стандартов. [↑](#)
2. Термин «слабоструктурированные данные» означает такие данные, которые в отличие от данных в БД не имеют регулярной структуры, определяемой с помощью предписывающей схемы. [↑](#)
3. **Метаданные** — свойства данных, определяющие их структуру, допустимые значения и способы их представления, взаимосвязи с другими данными, размещение и другие характеристики данных, которые помогают правильно их интерпретировать и использовать. Иначе говоря, это данные о данных. [↑](#)
4. **Информационный ресурс** — используемые в приложениях данные, которые представлены в базах данных, базах знаний, на Web-сайтах, в отдельных файлах различной природы или в процедурной форме с помощью продуцирующих их программных средств. [↑](#)