

## **Содержание:**

# **ВВЕДЕНИЕ**

World Wide Web(WWW) - глобальная компьютерная сеть на сегодняшний день содержит более 500 миллионов сайтов, на которых размещена всевозможная информация. Люди получают доступ к этой информации посредством использования технологии Internet. Для навигации в WWW используются специальные программы - Web-браузеры, которые существенно облегчают путешествие по бескрайним просторам WWW. Вся информация в Web-браузере отображается в виде Web-страниц.

Web-страницы, поддерживая технологию мультимедиа, объединяют в себе различные виды информации: текст, графику, звук, анимацию и видео. От того, насколько качественно и красиво сделана та или иная Web-страница, зависит во многом ее успех в сети.

Пользователю приятно посещать те Web-страницы, которые имеют стильное оформление, не отягощены чрезмерно графикой и анимацией, быстро загружаются и правильно отображаются в окне Web-браузера. Все большее и большее количество сайтов переходит под отображение операционной системы. Для использования «оконной среды»(windows).

Создать Web-страницу непросто, но наверно каждый человек хотел бы попробовать себя в роли дизайнера. И я, в данном случае, не являюсь исключением, потому и выбрал такую тему для курсовой работы.

В своей работе я сделал попытку разобраться в том, что необходимо знать и уметь для создания Web-страницы, какое программное обеспечение является инструментарием создания Web-страниц и как его эффективно использовать.

Это даст нам возможность ознакомиться со структура Web-страницы и приемами ее правильного оформления.

## **1. НАЗНАЧЕНИЕ ЯЗЫКОВ РАЗМЕТКИ**

Язык разметки текста представляет собой определённый набор символов, необходимых для его корректного вывода на страницу и придания нужной структуры. В итоге, документ содержит не только текст, но и данные о его составляющих – названиях разделов, подразделов, маркированных списках, использовании того или иного выделения ... В некоторых случаях использование языка разметки позволяет вводить в текст интерактивные элементы или части других документов.

Разметки бывают логические и визуальные. Логическая: речь идет только о том, какую роль играет данный участок документа в его общей структуре (к примеру, «данная строка является заголовком»). Визуальная — определяется, как именно будет отображаться этот элемент (к примеру, «данную строку следует отображать жирным шрифтом»).

Смысл языков разметки в том, что визуальное отображение документа должно автоматически получаться из логической разметки и не должно зависеть от его содержания. Это упрощает автоматическую обработку документа и его отображение в разных условиях (к примеру, один и тот же файл может по-разному отображаться на мониторе компьютера, на экране мобильного телефона и на печати, так как свойства этих устройств вывода существенно отличаются). Однако это правило часто нарушается: допустим, создавая текстовый документ в редакторе, подобном Microsoft Word, пользователь может выделить заголовки жирным шрифтом, но не указывать, что эта строка является заголовком.

Языки разметки используются везде, где требуется получение форматированного текста на основании только текста: в типографии (SGML, TeX, PostScript, RTF), в пользовательских интерфейсах компьютеров (MS WORD, OpenOffice, troff), Всемирной Сети интернет (HTML, XHTML, XML, WML, VML, PGML, SVG, XBRL).

## **1.1 ОСНОВНЫЕ ЯЗЫКИ РАЗМЕТКИ**

Существует достаточно много языков разметки, однако самыми на сегодняшний день популярными из них являются: XML;Wiki;BBC0de;Textile;HTML.

XML (в переводе с английского EXTENSIBLE Markup Language – расширяемый язык разметки).Его название связано с тем, что он не имеет зафиксированного формата. Это значит, что любой пользователь может создавать собственные теги, которые позволяют глубоко обрабатывать текстовые документы. Обычно XML используется

с целью осуществления описания грамматики других языков, а также для обеспечения контроля над правильностью составления документа. С виду документы HTML и XML очень схожи. Принципиальное отличие – более высокие требования к тегам в документах XML(к примеру, при простановке тега нужно следить за регистром). Язык XML дает проводить обмен данными даже между такими системами, которые даже не имеют совместимости. Это обеспечивает удобную работу с разными типами программ.

Wiki-разметка используется для таких веб-страниц, которые могут изменяться (дополняться, редактироваться) пользователями. Это позволяет создавать текстовые документы даже тем, кто не разбирается в особенностях HTML-разметки. Конечно, в Wiki существуют свои правила, но они очень простые, все изменения можно просмотреть и при необходимости вернуться к первоначальной версии. Благодаря этому исправлять ошибки намного проще. Администраторы могут ограничить право редактирования размещённых текстов, разрешить проведение манипуляций только определённым пользователям и т. д. Отличительная особенность Wiki разметки – гипертекстовость (связь документов посредством проставления контекстных гиперссылок).

BBCode. Этот язык применяется для форматирования сообщений, которые располагают на электронных досках с объявлениями, в блогах, на форумах и т. д. Теги похожи на аналогичные из HTML, однако они заключаются не в угловые, а в квадратные скобки. Появление такого языка связано с массовыми ошибками в отображении информации в чатах, на форумах и др. из-за относительной сложности стандартного языка. Особенность разметки заключается в том, что браузеры не воспринимают его самого по себе. Нужно устанавливать спецпрограмму, которая разберёт текстовый документ и преобразует его в понятный и универсальный HTML-код.

Язык Textile является одним из самых простых языков разметки, также он похож на BBCode, однако имеет немного большую функциональность. Его основная задача – автоматическое преобразование текста в разметку, сохраняя разбивки на абзацы. Язык используется в CMS Textpattern.

Больше всех в мире распространён язык HTML (в переводе с английского Hypertext Markup Language – «язык разметки гипертекста»). Он является стандартным для интернет-документов. С помощью HTML создаются все веб-страницы. Документы, содержащие в себе спецкод, обрабатываются браузерами и представляются пользователям в удобном интерфейсе, а значит, вы получите возможность

просматривать страницы, распечатывать их и использовать для передачи данных на серверы.

Язык разметки HTML открывает такие возможности работы с текстом:

1. выделение его составляющих полужирным или подчёркнутым шрифтом, курсивом
2. использование спецсимволов, которых нет в предустановленной пунктуации
3. изменение цвета буквенных и цифровых символов
4. проведение выравнивания текста, создание отступов
5. вставка гиперссылки, таблиц.

Начало и конец каждого элемента обозначается дескрипторами разметки (тегами(tag), заключёнными в угловые, треугольные скобки < >.

Теги языка HTML определяют:

- внешний вид док-та (стиль, формат шрифта, цвет фона и т. д.)
- структуру док-та: взаимное расположение текстовой, графической и другой мультимедийной информации
- ссылки на другие интернет-ресурсы, якоря и т.д. [2]

## **2. ГИПЕРТЕКСТ**

Тед Нельсон в 1965 году ввел термин гипертекст для обозначения «текста ветвящегося или выполняющего действия по запросу». Обычно гипертекст является набором текстов, которые содержат в себе узлы перехода от одного текста к любому другому, позволяющие выбрать читаемые сведения или последовательность чтения. Общеизвестным и достаточно выраженным примером гипертекста являются веб-страницы — документы HTML, которые размещаются в сети. В более широком понимании термина, гипертекстом является любая повесть, словарь или энциклопедия, где встречаются отсылки к другим частям данного текста.

В компьютерных терминах, гипертекст — это текст, который формируется при помощи языка разметки и потенциально содержащих в себе ссылки.

Всемирная паутина World Wide Web (WWW) является информационной гипертекстовой системой. Перемещаясь по этой паутине можно получать не только

текстовую информацию, а также и графику, таблицы, видео- и звуковые программы и другие виды спецпрограмм, хранящихся во многих сотнях тысяч управляющих компьютерных систем в более чем 150 странах мира. Популярность Internet во многом вызвана появлением всемирной паутины, так как это первая сетевая технология, которая предоставила пользователю простой современный интерфейс для доступа к различным сетевым ресурсам. Простота и удобство применения привлекли внимание коммерческих структур. Процесс роста числа пользователей стал лавинообразным, и рост продолжается до сих пор. На основе необходимости объединить все информационные ресурсы стала развиваться технология, на основе которой определяется гипертекстовая навигационная система. Этой технологией стал язык HTML. Технология HTML на начальном этапе была невероятно проста, и почти все пользователи сети разом получили возможность попробовать себя в качестве создателей и читателей информационных материалов, опубликованных в WWW. При разработке различных компонентов технологии предполагалось, что квалификация авторов информационных ресурсов и их оснащенность средствами вычислительной техники будут минимальными.

Вся технология WWW основана на таких основных элементах, как:

- язык HTML ;
- универсальный способ адресации ресурсов в сети URL (Universal Resource Locator);
- протокол обмена гипертекстовой информацией HTTP (Hyper Text Transfer PROTOCOL).
- универсальный интерфейс шлюзов CGI (COMMON Gateway Interface).

HTML относится к числу так называемых языков разметки текста (markup languages). Под термином "разметка" понимается общая служебная информация, которая не выводится вместе с документом, но определяет, как должны выглядеть фрагменты документа. Допустим, вы можете потребовать, чтобы какое-то слово выводилось жирным или курсивным шрифтом, можно вывести отдельный абзац особым шрифтом или оформить заголовки увеличенным шрифтом.

В настоящее время существует много языков разметки. К примеру, в коммуникационных программах особая форма разметки определяет смысл каждого пакета из нулей и единиц, пересылаемого в Internet. Любой язык разметки должен решать две важные задачи:

1) язык определяет синтаксис разметки;

2) язык определяет смысл разметки.

## 2.1 ЯЗЫК ГИПЕРТЕКСТОВОЙ РАЗМЕТКИ HTML

Самым распространенным языком разметки Web-страниц является HTML. Он был создан и рекламировался как одна из конкретизаций SGML. HTML впервые был предложен в 1974 году Чарльзом Голдфарбом и в дальнейшем после значительной доработки был принят в качестве официального стандарта ISO, SGML (Standard Generalized Markup Language - Стандартный обобщенный язык разметки). SGML представляет собой метаязык – систему для описания других языков.[1]

Появление стандарта SGML было обусловлено необходимостью совместного использования данных различными приложениями и ОС. В 60-х годах у пользователей компьютеров возникало немало проблем с совместимостью.

Проанализировав недостатки многих нестандартных языков разметки, трое ученых из IBM — Чарльз Гольдфарб, Эд Мо-шер и Рэй Лори — сформулировали три общих принципа, которые обеспечивали возможность совместной работы с документами в разных ОС:

1) Использование единых принципов форматирования во всех программах, которые выполняют обработку документов. Очень логичное требование — ведь всем нам хорошо известно, как сложно договориться между собой людям, которые говорят 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 реализаций 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, 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.

Стандарты языка HTML, одного из ключевых стандартов Web, разрабатываются и поддерживаются консорциумом W3C. Основателем этого международного консорциума является Тим Бернес-Ли. Консорциум помимо создания стандартов форматирования, является центром разработки Semantic Web (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 всемирной паутины — 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:

1. предоставить дизайнерам гипертекстовых баз данных простое средство создания документов
2. сделать это средство настолько 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 размечать электронный документ, который отображается на экране с полиграфическим уровнем оформления; результирующий документ может содержать самые разнообразные метки, иллюстрации, аудио- и видеофрагменты и прочее. В состав языка вошли развитые средства для создания различных уровней заголовков, шрифтовых выделений, разнообразные списки, таблицы и многое другое.

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

редактирования.

Гипертекстовые ссылки, которые устанавливают связь между текстовыми документами, постепенно начали объединять самые различные информационные ресурсы, в том числе звук и видео. Система гиперссылок HTML позволяет построить систему взаимосвязанных документов по разным критериям. Язык HTML содержит команды (тэги), которые позволяют управлять формой и размером шрифтов, размером и расположением иллюстраций, осуществлять переход от фрагмента текста или иллюстрации к другим html - документам – так называемую гипертекстовую ссылку. Документ в html - формате является текстовым файлом, содержащим все необходимые сведения об информации, которая выводится на экран . Для управления сценариями просмотра страниц Website (гипертекстовой базы данных, выполненной в технологии WWW) можно использовать языки программирования этих сценариев, к примеру, JavaScript, Java и VBScript. Формы для введения пользователем данных, которые позднее подвергаются обработке и другую информацию можно обрабатывать с помощью специальных серверных программ (к примеру, на языках PHP или Perl). Язык HTML позволяет помещать на страницы гипертекстовые ссылки и интерактивные кнопки, соединяющие ваши Web-страницы с другими страницами того же Web-сайта, а так же с другими Web-сайтами по всему миру.

HTML является языком разметки текста, а не языком программирования(в отличие от допустим PHP), который всего лишь один из инструментов (точнее, язык описания страниц), который используют при создании Web-страниц. В HTML Ограничены возможности форматирования текста по сравнению с возможностями издательских программ, особенно при издании текста, который перенасыщен сложными элементами.

Для создания HTML док-та необходимо:

1. Создать персональную папку, используя в имени папки исключительно английские буквы
2. создать текстовый документ, типа блокнот,
3. необходимо сохранить документ под новым именем, используя расширение .html либо .htm
4. закрыть док-т
5. открыть html документ и начать редактирование...

В настоящее время нет достаточно удобных редакторов HTML, чтобы можно было бы обойтись без текстового редактора и ручной расстановки тэгов (тег-управляющий символ, определяющий, в каком виде мы видим WEB-страницу в браузере интернета). Это делает работу с языком сложной, а также делает необходимым овладение ими совершенно несвойственных им функций.

HTML-документы бывают двух видов – статические и динамические:

статические страницы (рис.1) несут в себе некую жестко заданную информацию, для редактирования которой обязательно нужно внести изменения в гипертекстовый документ .

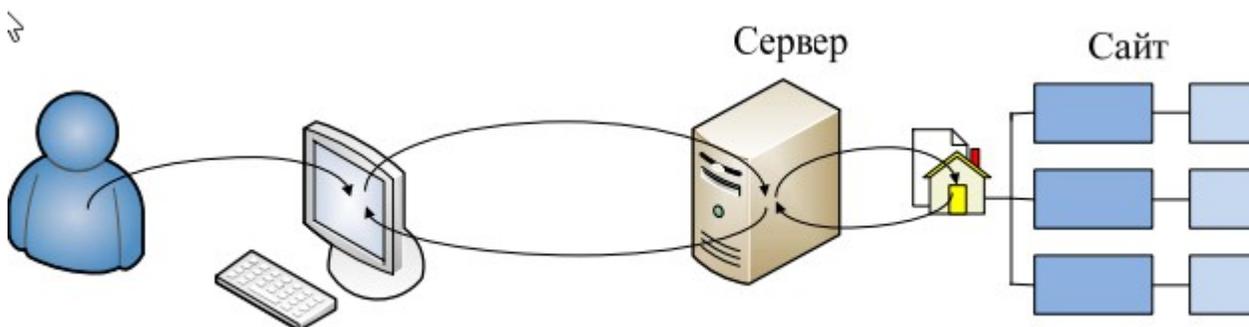


Рисунок 1

динамические страницы (рис.2) разрешают показ данных, которые могут меняться и без изменения самого HTML-документа (к примеру, информацию, которую можно извлечь из СУБД). При создании динамических HTML- документов чаще всего пользуются специальными серверными расширениями, которые называются сценариями (бывает для их обозначений используется термин «web-приложения» или сленговое выражение «скрипт»). Главная задача, которую выполняет сценарий, — это получение информации из некоего внешнего источника (к примеру, из СУБД), которая затем показывается в виде HTML- документов, передается серверу, а он уже, отправляет страницу клиенту. В свою очередь, сценарий позволяет обеспечивать интерактивное взаимодействие с клиентом, когда обрабатывает данные, которые передаются от клиента к серверу. Таким образом реализуется, к примеру, вариант поиска либо выборка из СУБД именно информации, запрашиваемой пользователем.[4]



Рисунок 2

Динамические сайты бывают разными, в зависимости от технологий, которые используются. Процесс их получения тоже может отличаться:

1. Генерация происходит на стороне сервера (осуществляется серверными скриптами на языках PHP, Perl, ASP.NET, Java, Python и др., а информационное наполнение хранится в СУБД).

2. Генерация происходит на стороне клиента (JavaScript).

3. Комбинированная генерация (на практике встречается именно комбинация первых двух методов, так как она более удобна). [2]

Статические документы хранятся в файлах файловой системы, используемой web-сервером или браузером при просмотре локальных файлов. Размещая информацию на web-сервере запросто можно использовать динамические документы - те, которые не существуют постоянно в виде файлов, а генерируются в момент запроса клиента. Причем для конечного пользователя нет разницы динамический или статический способ представления документов.

Для генерирования динамического документа HTML необходима специально написанная программа по правилам, определяемым web-сервером. При планировании размещения информации на web-сервере, для правильного определения использования, какого либо вида документов, нужно учитывать степень обновляемости данных, их объем и частоту обращения.

Динамический способ определяет хранение данных в формализованном виде, к примеру в СУБД. А если вдруг данные хранятся в формализованном виде, то, при использовании шаблонов документов, в которых были внесены изменения, генерируются статические документы. Для того, чтобы генерировать статические

документы допустимо использовать любые средства отчетов, имеющихся в этой системе управления баз данных (СУБД), которой обработаны и формализованы данные.

Новых версий языка HTML не должно быть, но существует дальнейшее развитие HTML под названием XHTML (англ. Extensible Hypertext Markup Language — расширяемый язык разметки гипертекста). Пока XHTML по своим возможностям сравним с HTML, однако он предъявляет более строгие требования к синтаксису. Как и HTML, XHTML является подмножеством языка SGML, однако XHTML, в отличие от HTML, соответствует спецификации XML. Вариант XHTML 1.0 был одобрен в качестве Рекомендации Консорциума Всемирной паутины (W3C) 26 января 2000 года. Но нужно учесть одну серьезную деталь — в этом формате создано огромное количество информационных ресурсов, и они долго еще будут "пониматься" web-браузерами и использоваться в своем первоначальном виде. Кроме того, все новые форматы будут разрабатываться (и уже разрабатываются — к примеру XML) с поддержкой технологий HTML.[3]

Стиль работы меняется, изменяются и средства доступа к содержимому. Язык HTML изначально создавался как платформа — независимый язык. в настоящее время применение HTML довольно далеки от его первоначальной задумки. И используются в основном при необходимости мультимедийного и графического оформления WEB-страниц.

К примеру, чтобы добавить на web-страницу элемент фонового звука применяется элемент `bgsound` (Background Sound), адрес аудиофайла прописывается с помощью, всем известного, атрибута `src`. Аудиофайл имеет некоторое короткое время звучания, поэтому количество его повторений записывается атрибутом `loop`. Если указать значение `loop="-1"`, аудиофайл будет повторяться и повторяться бесконечно. Зададим десятикратное повторение воспроизведения аудиофайла в этом примере HTML-кода:

```
< bgsound src="audio/file.wav" loop="10" >
```

Громкость воспроизведения аудиофайла регулируется атрибутом `volume`, значения которого записываются в пределах от -10 000 до 0. Максимальная громкость это нулевое значение.

Стереобаланс задается атрибутом `balance`, в пределах от -10 000 до 10 000. Нуль определяет несмещенный баланс (`balance="0"`). Код можно записать в таком виде:

```
< bgsound src="audio/file.wav" loop="10" balance="0" volume="-2000" >
```

Когда в состав HTML-документа включают графические изображения, то их хранят в отдельных файлах, а в тексте HTML-документа есть ссылка на соответствующий файл. Для хранения изображений используют чаще всего файлы формата JPEG, GIF или PNG.[4]

Новые технологии применяются повсюду и достаточно скоро пространство WWW перестанет быть достоянием только пользователей настольных персональных компьютеров, уже сейчас некоторые пользователи активно пользуются голосовыми браузерами для незрячих или браузерами, использующими азбуку Брайля; очень распространено, когда содержимое выводится не на монитор компьютера, а на телевизор, с применением различных приставок с выходом в сеть или на телетайп, или с использованием телевизоров с функцией SMART, или на монохромные дисплеи различных Организаторов-пейджеров и прочие.

В последнее время для того, чтобы оформить web-страницу кроме языка HTML достаточно широко применяются фрагменты на таких языках, как Java, JavaScript, VBScript. В отличие от HTML они предоставляют намного большие возможности для передачи информации, чем язык HTML.[4]

## **2.2 РАСШИРЯЕМЫЙ ЯЗЫК РАЗМЕТКИ XML**

Вторая половина 90-х годов прошедшего века ознаменовалась огромными переменами в технологиях Web. Меньше чем за пять лет в истории своего существования Web приобрел огромное количество пользователей на всех континентах, в его среде сформированы и поддерживаются невероятные информационные ресурсы. Эта глобальная информационная система интенсивно вторгается в другие области IT технологий, стала одним из важных звеньев инфраструктуры информационного общества.

Однако ряд ограничений, которые свойственны действующим технологиям Web (Web первого поколения или Web-1), стал сдерживающим фактором для дальнейшего его развития. Новые подходы в области технологий Web, которые конструктивно воплощались в жизнь на пороге XXI века, направлены, в первую очередь на преодоление этих ограничений и создание технологической платформы<sup>1</sup>, которая бы обеспечила потенциал для появления нового поколения Web (Web второго поколения или Web-2) и возможностей его развития на

дальнейшую перспективу. Главную роль в технологическом переоснащении Web стал играть разработанный консорциумом W3C новый язык разметки XML.[7]

XML (англ. Extensible Markup Language, расширяемый язык разметки)-это язык разметки, который был создан для описания данных. Можно сказать, что он сам является метаязыком, ибо он не ограничивается набором каких-либо определенных тегов и используются как средство для описания грамматики многих других языков и в качестве контроля за правильностью составления документов. XML-документ чаще всего состоит из процессинговых инструкций, элементов, атрибутов, сущностей и комментариев. Пример синтаксиса XML:

```
<?xml version = "1.0" encoding = "windows -1251"?>
```

```
<photos>
```

```
<photo
```

```
url = " .. /images/ettlingen01.jpg"
```

```
description="Ettlingen,Germany" />
```

```
</photos>
```

Обработка XML-документа чаще всего проходит при использовании технологий таких ,как DTD/XDR/XSD схем, XPath, XSLT и так далее. [6]

Консорциум W3C, который был создан для проведения единой технической политики в рамках Web и развития его технологий, в данное время разрабатывает и поддерживает более 150 стандартов. Естественно, нельзя себе представить их здесь в достаточно полном виде и остается ограничиться только обсуждением концептуальных аспектов наиболее важных из них. Для полного изучения стандартов платформы XML необходимо обратиться к их спецификациям и другим материалам консорциума W3C.[7]

Необходимо заметить, что аббревиатуру XML очень часто используют для того, чтобы обозначать не только сам язык XML, но и некоторые другие связанные с ним понятия — определяющего язык стандарта W3C, информационных ресурсов XML, комплекса основанных на этом языке стандартов консорциума W3C, составляющих платформу XML.

Так как язык XML все чаще используют в среде Web по прямому своему назначению — как выразительное средство для представления IT ресурсов в этой среде, XML однако энергично внедряется в другие технологии. Развитые выразительные возможности языка, а также, что немаловажно, его поддержка механизмами среды Web дают использовать XML в качестве языка-посредника для определения форматов обмена данными между разными системами, используемыми Internet в качестве коммуникационной среды.

Важная сфера применения стандартов платформы XML — это представление слабоструктурированных данных для Web-сайтов в форме XML-документов. Но в принципе, для этой ведь цели и создавался язык XML. Применение XML в этой области дает во первых представлять в среде Web гипермедийные страницы в форме XML-документов, а во вторых поддерживать связанные с ними метаданные. Именно благодаря этому создаются такие поисковые машины Web, которые обеспечивают в результате обработки пользовательского поискового запроса намного более низкий уровень информационного шума ,если сравнивать с настоящими HTML-технологиями.

Одной из наиглавнейших целей создания платформы XML является привнесение в среду Web таких метаданных, которые описывают свойства информационных ресурсов, которые в ней поддерживаются. В данном случае речь прежде всего об описании структуры XML-документов и их смыслового содержания (семантики). Смысл решения этой задачи аргументируется стремлением к получению возможностей автоматической проверки правильности структуры XML-документов и снижения уровня информационного шума при отыскании необходимых данных в Web с помощью различных поисковых машин. Это значит, что если есть явное описание структуры документов проверку их правильности может осуществить браузер. Описание семантики документов может быть полезным подспорьем для новых или модернизированных существующих поисковых машин, а также для разнообразных нуждающихся в нем Web-приложений.

Однако скорее всего не учитывается еще одно очень важное назначение метаданных, которые описывают информационные ресурсы Web. Метаданные очень нужны для создания принципиально новых высокоуровневых приложений Web, в частности тех, что основаны на интеграции IT технологий и обеспечивающих интеграцию неоднородных IT ресурсов.

Самый простой способ задания семантики —это использовать пространства имен. В отличие от HTML,который обеспечивает форматную разметку текста,

определяющую его представление на экране, язык XML нужен для структурной разметки.

Разметка в XML дает выделить в тексте содержательные структурные единицы, которые называются элементами XML-документа. Для того, чтобы выделить каждый тип элемента используют свой тег, указывающий имя типов элементов. Именно поэтому с каждым тегом можно ассоциировать семантику соответствующих элементов XML-документа (адрес организации, номер телефона ,сферу деятельности...)

Когда некоторое сообщество разработчиков и пользователей XML-документов договорится о единой интерпретации имен, которые будут принадлежать некоторому набору, то этот унифицированный набор, скорее всего, с каким-либо описанием их смысла (к примеру, в виде обычного текста на естественном языке или представленный каким-то другим способом), может использоваться как пространство имен. Адрес документа, который будет представлять в Web этот набор имен, будет рассматриваться как уникальный идентификатор пространства имен(VIN), и на него можно будет ссылаться в XML-документе, где используются имена, которые принадлежат этому пространству. И таким образом им придается некоторый смысл.

Ресурс Web, адрес которого символизирует некоторое пространство имен, может не содержать никакого явного описания смысла принадлежащих ему имен или вообще не существовать. В этом случае мы имеем дело с определением семантики имен этого пространства по умолчанию.

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

DC с принятой в нем семантикой элементов метаданных может использоваться в рамках платформы XML разными способами. К примеру, можно применять ядро в качестве пространства имен для некоторого типа XML-документов или в RDF-спецификации (Resource Definition Framework, стандарт схемы описания источников).

RDF-спецификации представляют собой высочайший уровень семантического описания IT ресурсов. IT 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 литеров , которые в данное время называют контейнерами. Допускаются контейнеры типа мультимножества, последовательности и альтернативы.

Для того чтобы RDF-спецификация семантики IT ресурсов была полной, необходимо ассоциировать с ней описания семантики используемые в этой спецификации свойства, которые в терминологии стандарта RDF называются схемой.

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

В задачу RDF не входит стандартизация любых наборов семантических свойств, и они могут быть разными в разнообразных случаях.

В настоящее время стали создаваться сервисы регистрации и поддержки пространств имен в интересах различных сообществ разработчиков и пользователей. Зарегистрированное пространство имен является так сказать стандартом для сообществ клиентов сервиса регистрации.

XML — далеко не модное направление, а естественный результат развития Web-технологий, следствие стремлений к более эффективному использованию уникальных возможностей открытой глобальной IT среды, которую они поддерживают. Создание платформы XML — это новая эпоха в развитии WWW, это — начало нового, более наукоемкого и технологически более совершенного этапа в ее истории. На сегодняшний день XML, конечно, стал стандартом де-факто , все ведущие поставщики ПО не только Web, но и СУБД, включают в свои программные продукты поддержку языка XML или даже создают специализированные системы.

Огромную работу по продвижению стандартов платформы XML в практику ведет крупный Международный, некоммерческий консорциум OASIS (Organization for Structured Information Standards, Организация по стандартизации структурированной информации) 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 (к примеру, разработки описания типов документов (Document Type Definition, DTD), схем XML и пространств имен XML), спецификаций интероперабельности (в частности, создание спецификаций профилей, включающих стандарты рассматриваемой категории), тестов на соответствие рассматриваемым стандартам.[6]

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

Несмотря на то, что язык XML и базирующиеся на нем стандарты получают все большее широкое распространение, все же имеются факторы, сдерживающие пока массовое распространение XML в среде Web.

В первую очередь, существует естественная инерционность столь масштабной среды, какой является сегодняшний Web, которая связана с экономическими и не только причинами. Эта инерция может преодолеваться только очень постепенно.

К тому же, пока еще не завершена работа над двумя самыми главными стандартами платформы XML, позволяющими строить из отдельных XML-документов и их компонентов гипермедийную среду. Речь в данном случае о стандартах XPointer (XML Pointer Language, язык указателей XML) и XLink (XML Linking Language, язык ссылок XML). Эти стандарты решают задачу определения гиперссылок в языке XML. Возможности стандартов XPointer и XLink обладают существенно более богатыми возможностями работы с гиперссылками, чем у имеющихся в HTML.[6]

Технологии XML стали распространяться и в нашей стране. В связи с этим начинает приобретать особое значение русскоязычная терминология в этой области.

У платформы XML очень благоприятные перспективы для широкого практического применения. Об этом свидетельствуют не только большие функциональные возможности рассмотренного семейства стандартов, но и высокая активность в области разработки и развития стандартов, а также производства программного обеспечения, которое основано на этом.

На основе этого языка создаются языки разметки информационных ресурсов в таких областях науки, как математика, химия, астрономия, геоинформатика, в области электронного бизнеса, и они достаточно хорошо применяются на практике. Предлагаются разнообразные способы структурированного представления библиографической информации при помощи XML и к созданию на этой основе электронных библиотек с поисковыми машинами, обладающими необходимой информацией. В ходе разработок используют средства платформы для представления метаданных, определенных стандартом RDF, RDFS, OWL.[5]

## **2.3. РАСШИРЯЕМЫЙ ЯЗЫК РАЗМЕТКИ ГИПЕРТЕКСТА XHTML**

В отличие от XML, язык HTML намного строже, чем данный язык разметки с неким ограниченным набором тегов. Во всяком случае, общий характер XML дает нам рассматривать HTML-документы как XML-документы с неким набором тегов для отражения в web-браузерах. Но, несмотря на это, первые стандарты HTML не до конца совместимы с XML. К примеру, в HTML совсем не обязательно закрывать тег `p`, а именно закрывающую часть тега `</p>` можно пропускать. Web-браузер может правильно интерпретировать эту конструкцию, так как он так запрограммирован, однако XML-парсер будет показывать ошибку, что HTML-документ неправильно построен (well-formed). Для того, чтобы это исправить, и был разработан XHTML. В принципе, это обычный язык HTML, который был дополнен синтаксическими правилами XML для построения well-formed документов.[2]

XHTML – представляет собой основанный на XML язык разметки гипертекста, который очень сильно приближен к текущим стандартам HTML, что не удивительно, ведь он является преемником языка гипертекстовой разметки. XHTML отличается от HTML исключительно строгостью написания кода. К примеру, если HTML позволял написание разнообразных конструкций, а интернет браузер их правильно распознавал, то с языком XHTML, это стало абсолютно невозможным. Потому что преемник требует самого строгого соблюдения абсолютно всех правил, которые предъявляет W3C. Причем такие строгие правила оформления XHTML-кода дают нам избежать очень многих ошибок даже в самом начале написания, а также на стадии отладки.[8]

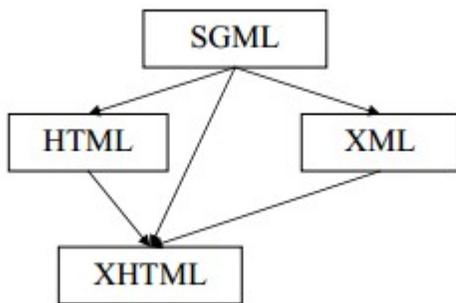


Рисунок 3 Эволюция языков разметки

Для приверженцев HTML будет полезно узнать, что XHTML - это совершенно новый язык, пришедший на смену HTML. Новых версий HTML больше не будет. Получается, все браузеры, как предполагают пользователи, перейдут на XHTML (и скорее всего, при переходе сохранится совместимость с языком HTML, на этом все). Здесь хочется привести цитату из русского перевода спецификации XHTML от W3C: "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".

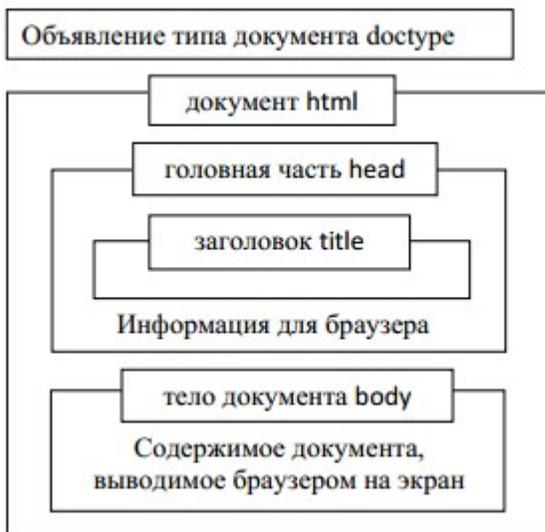


Рисунок 4 Структура документа XHTML

XHTML совместим с HTML при соблюдении некоторых правил, описание которых можно найти в стандарте. Это обозначает, что даже самые старые браузеры(к

примеру, internet explorer), которые понимают HTML, также будут корректно работать и с XHTML.

Чтоб проверить правильность написания XHTML-кода существуют разнообразные программы – валидаторы.

Примеры:

1.

```
<a href="index.php?pid=1&id=2">...</a>
```

Вероятное сообщение валидатора: Unknown entity...

Что делать: Всегда использовать &amp; вместо &.

Правильно:

```
<a href="index.php?pid=1&amp;id=2">...</a>
```

2.

```
<td rowspan=3>
```

Вероятное сообщение валидатора: Missing " "

Что делать: Все значения атрибутов должны быть в кавычках.

Правильно::

```
<td rowspan="3">
```

Плюс к этому, пользовательские агенты, поддерживающие XHTML, сами будут вам сообщать об ошибках в синтаксисе, если вдруг такое случится.

Существует несколько требований, которые обязательно должны выполняться (рис.4) :

1. в начале документа обязательно должен быть указан один из возможных DTD (Document Type Definition):

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
```

```
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
```

```
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Frameset//EN"
```

```
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-frameset.dtd">
```

2. в теле XHTML-документа всегда должны присутствовать такие тэги ,как : html, head, title и body;

3. обязательно наличие атрибута xmlns в элементе html;

4. названия тегов и атрибутов всегда должны быть записаны в нижнем регистре;

5. все значения атрибутов должны быть заключены в "кавычки";

6. все тэги должны быть закрыты, если у элемента нет закрывающего тэга, тогда необходимо добавить в конец слеш (<br/> или <br /> — слеш пробел для совместимости со старыми браузерами);

7. обязательно нужно соблюдать корректную вложенность тэгов ,к примеру (<b><i>текст</b></i> — неверно, необходимо писать <b><i>текст</i></b>);

8. ни в коем случае нельзя использовать минимизированные атрибуты (nowrap следует заменить на nowrap="nowrap"). Полный список таких атрибутов: checked, compact, declare, defer, disabled, ismap, noresize, noshade, nowrap, multiple, readonly, selected.

9. на некоторые элементы налагаются некие ограничения по включению в них каких-то еще элементов:

a не может включать в себя другие элементы a;

form не может включать в себя другие элементы form;

label не может включать в себя другие элементы label;

pre не может включать в себя img, object, big, small, sub или sup;

button не может включать в себя элементы input, select, textarea, label, button, form, fieldset, iframe или isindex;

10.специальные символы в реальном значении должны быть изменены на свои эквиваленты:

& на &amp;;

< на &lt;;

> на &gt;.

Более того, существует очень много необязательных рекомендаций, которые разработчик может не выполнять в версии XHTML 1.0, но в следующих версиях возможно эти рекомендации перерастут в обязательные:

1.декларация XML-документа в самом начале кода перед DTD (<?xml version="1.0" encoding="windows-1251"?>);

2.следование элемента title сразу после открытия тега head;

3.использование атрибута id вместо name (name уже считается устаревшим атрибутом);

4.наличие атрибута type в элементах подключаемых файлов (таблиц стилей и скриптов);

5.отказ от использования атрибута target.

Ниже представлен минимальный код правильной XHTML-страницы:

```
<?xml version="1.0" encoding="windows-1251"?>
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
```

```
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
```

```
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="ru" lang="ru">
```

```
<head>
```

```
<title>Заголовок</title>
```

```
</head>
```

```
<body>
```

Содержимое документа

```
</body>
```

```
</html>
```

Если вставить этот код в файл, сохранить его как "file.html" и открыть через веб-сервер, то вся информация будет получена клиентом как `>text/html<`. То есть, как и обычная HTML-страница. В принципе, для браузера это будет не XHTML, а HTML-документ. У XHTML есть свой собственный MIME-тип: `<application/xhtml+xml>`.

MIME - это специальный набор расширений, указывающие программам, как обработать входящую информацию. С самого начала MIME-типы были разработаны для почтовых программ, где и заимели своё название.

Итог, XHTML-данные правильно отдавать клиенту именно в формате `<application/xhtml+xml>`, так как все преимущества, помимо кросс-браузерности (увеличение скорости анализа кода процессором XML, сообщение об ошибках самим браузером и пр.), могут достигаться только в том случае, если пользовательский агент поддерживает XHTML и ему сообщается о том, что входящие данные - XHTML-код. Единственное, что надо помнить при отправке XHTML-кода: когда браузер понимает язык XHTML, то тогда и только тогда можно отправлять информацию как `<application/xhtml+xml>`; если же нет, то только как `<text/html>`. Список современных браузеров, которые поддерживают XHTML: MZ, Opera, Safari. Internet Explorer пока не понимает `<application/xhtml+xml>`. Проверить, поддерживает пользовательский агент нужный MIME-тип или нет, можно по исходящему от браузера заголовку `<Accept>`, который содержит все MIME-типы, известные клиенту. [8]

## **ЗАКЛЮЧЕНИЕ**

В связи с быстрым развитием Web-программирования и создания множественных CMS, была облегчена работа Web-дизайнера. Теперь для создания многофункционального мощного динамического web портала достаточно освоить несколько языков программирования (если будем дописывать расширения и скрипты), умело владеть графическими и растровыми программами, немного вдохновения, инноваций и мастерства.

# СПИСОК ИСТОЧНИКОВ

1. Основы Web – технологий : учеб. пособие / П.Б. Храмцов [и др.]. – М. : Изд-во Интуит.ру “Интернет-Университет Информационных Технологий”, 2003. – 512 с.
2. Web – технологии : учеб. пособие / И.Н. Васильева, Д.Ю. Федоров –СПб. : Изд-во СПб-Петербургского Государственного Экономического Университета, 2014. - 67 с.
3. Когаловский, М.Р. Перспективные технологии информационных систем / М.Р. Когаловский. – М.: Компания АйТи, 2003. – 288 с
4. Петров, В.Н. Информационные системы: учеб. пособие / В.Н. Петров. – СПб.: Питер, 2002. – 688 с.
5. Когаловский М.Р. Стандарты платформы XML и базы данных. Обзорный доклад. Сб. трудов Третьей Всероссийской конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции", Петрозаводск, 11-13 сентября 2001 г. – Петрозаводск: Карельский научный центр РАН, 2001.-188 с.
6. 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 Институт проблем рынка РАН <<http://www.cemi.rssi.ru/mei/articles/conf02st.htm>> (27-28.03.2002).
7. Extensible Markup Language (XML) [Электронный ресурс] / Спецификация <<http://www.w3.org/XML/>> .
8. 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).