

УДК 004.91:004.55

А.С. Власовских (2 курс, каф. АиВТ), Е.В. Пышкин, к.т.н., доц.

МЕХАНИЗМЫ ОТДЕЛЕНИЯ ЛОГИЧЕСКОЙ СТРУКТУРЫ ИНФОРМАЦИИ ОТ ЕЁ ПРЕДСТАВЛЕНИЯ В ЭЛЕКТРОННОМ ДОКУМЕНТООБОРОТЕ

В бумажном документообороте информация создаётся и распространяется в виде *цельного бумажного носителя*, поэтому выделять логические блоки, способные к относительно независимому существованию, не имеет смысла. В связи с этим возможности традиционного манипулирования информацией на основе бумажных носителей довольно ограничены. В электронном документообороте возможность выделения логической структуры информации позволяет облегчить и систематизировать не только создание и представление электронных документов, но и их распространение. Несмотря на все достоинства обработки потоков информации в формате *plain text* (который, вероятно, до последнего времени являлся единственным вполне независимым от платформы), многие пользователи предпочитают иметь дело с форматами, поддерживающими полноценную структурную разметку документа.

Для описания организации текста были созданы *языки разметки текста*. Они предполагают решение двух задач – задачи *структурной* разметки, то есть встраивания в документ управляющих кодов, указывающих на логические отношения между частями текста, и задачи *визуального форматирования*, то есть обеспечения желаемого оформления отображаемого документа. Примером языка разметки является HTML – основной инструмент описания содержимого Web-страниц в Internet. Абсолютное большинство тегов HTML версии 1.2, созданной в 1993 году, предназначались для структурной разметки. При этом для представления документа на экране HTML-браузеры документа пользовались некими правилами умолчания. Консорциум W3C, образованный в 1994 году для стандартизации языка гипертекстовой разметки, принял версию HTML 2.0 и начал работу над третьей версией.

Для решения задачи визуального форматирования документов была создана спецификация *иерархических стилевых таблиц (CSS)*. Возможность применения таблиц CSS была впервые предусмотрена в спецификации HTML 3. Система CSS создавалась как независимая от HTML и должна была решить очевидное к тому времени противоречие между идеологией структурной разметки и потребностями пользователей в гибких возможностях визуального форматирования. Растянутый во времени процесс принятия нового стандарта привёл к тому, что в различных браузерах эти возможности были реализованы различными HTML-элементами. Всё это усложняло сопровождение Web-сайтов, поскольку Web-дизайнеры вынуждены были обеспечивать межбраузерную совместимость. Отделение логической структуры информации от её представления в полной мере было реализовано в спецификации HTML 4.01, принятой в конце 1999 года. Этот стандарт предписывает использование HTML преимущественно для структурной разметки. Представление документа описывается независимой спецификацией CSS 2, при этом элементы HTML, управляющие представлением документа, объявлены *нерекомендуемыми (deprecated)*.

В настоящее время интенсивно развивается язык XML – расширяемый язык разметки, следующий идее исключительно структурной разметки документа. XML позволяет хранить информацию, структурируя ее логически при помощи элементов, определяемых пользователем. XML ориентирован на представление иерархически организованных данных и является открытым стандартом. XML хорошо адаптирован для машинного разбора специальными программными модулями – XML-парсерами. Развиваясь в качестве универсального кросс-платформенного формата, XML позиционируется рядом специалистов как *lingua franca* информационных технологий настоящего времени.

Складывающаяся практика применения XML заключается в его использовании не только для описания отдельных документов, но и для создания стандартизированных наборов элементов XML и структур документов, применяемых для особых классов документов. Подобные наборы элементов вместе с описаниями структур документов называются XML-приложениями. Данная концепция позволяет использовать XML-документы совместно с другими пользователями документов данного класса (т.е. совместно со всеми пользователями данного XML-приложения), а значит – использовать программное обеспечение, которое создано для данного XML-приложения.

Собственно, разработчики стандартов HTML из W3C так и поступили, переопределив элементы HTML 4.01 в виде набора XML-тегов. Получившийся стандарт стал называться XHTML 1.0. Таким образом, новые стандарты языка с аббревиатурой HTML больше разрабатываться не будут. На смену приходит спецификация XHTML (текущая версия 1.1), которая и является XML-приложением. Аспекты представления информации (визуального или любого другого, к примеру, аудио) теперь должны задаваться исключительно с помощью каскадных таблиц стилей CSS. В результате такой трансформации языка у разработчиков браузеров (и не только у них) появились в распоряжении все средства XML.

Актуальность исследований, являющихся темой настоящего доклада, обусловлена необходимостью систематизации материала, используемого при подготовке и проведении учебных занятий по курсу «Информационные технологии документооборота», который включён в учебный план кафедры автоматизации и вычислительной техники в 2002 году. Положения, изложенные в докладе, частично использованы при организации лабораторного практикума, посвящённого изучению принципов создания, обработки и распространения электронных документов.

Результатом изучения и грамотного использования изложенных концепций стала разработка Web-сайта учебной группы 2081/4, созданного в соответствии со стандартами XHTML 1.1 и CSS 2. Строгое соответствие спецификации подтверждено службой проверки W3C (W3C Validation Service). Помимо достижения актуальных для учебной группы целей, разработка сайта подтверждает, что, используя только структурную разметку и отделяя визуальное представление от логической структуры при помощи CSS-таблиц, можно создавать страницы с отличным дизайном.

В настоящее время интенсивные исследования ведутся в области управления потоками информации, хранящейся в структурно размеченных документах. В связи с этим необходимо отметить, что структурная разметка сама по себе не решает проблем управления документами, но, безусловно, является одной из необходимых предпосылок.