

**СОВРЕМЕННЫЙ ВЗГЛЯД
НА ТЕКСТОВУЮ РАЗМЕТКУ**
**С.Д. Иванов, В.В. Грязнова,
Е.Г. Андрианова**

Текстовая разметка состоит в явном указании пользователем директив предписывающих использование того или иного форматирования при отображения определенных участков текста или включение в документ дополнительных нетекстовых элементов, например изображений. На протяжении многих лет шел переход от тех или иных форм текстового форматирования к WYSIWYG технологиям (сокращение от What You See Is What You Get, англ. что видишь, то и получишь). Все современные текстовые процессоры являются WYSIWYG приложениями. Преимущество этой технологии на первый взгляд кажутся очевидными, однако у визуального редактирования есть и свои недостатки. В этой статье мы рассмотрим те из них которые связаны с подготовкой текстовых документов. Согласно современным реалиям любой текстовый документ состоит из 4 основных составляющих: содержимое (content); форматирование (markup); внедренные нетекстовые объекты; семантические элементы. Например HTML страница состоит из собственно HTML документа, таблицы CSS стилей определяющих форматирование и изображений/анимации в качестве внедренных объектов и ключевых слов, как семантических элементов. При WYSIWYG разработке естественно основной упор делается на содержание документа и его форматирование. Оно и понятно — внедренные объекты имеют не текстовую природу и подготавливаются

отдельно, а о семантических элементах автор как правило забывает или пренебрегает.

Рассмотрим более детально виды форматирования, применяемые в документах:

Физическое: форматирование зависит исключительно от желания пользователя выделить тем или иным образом части текста, согласно его вкусам и предпочтениями.

Логическое: пользователь определяет роль той или иной части текста в документе, а правила форматирования определяются соглашениями принятыми в данной информационной системе (например, посредством таблицы стилей).

Здесь мы сталкиваемся с первой проблемой WYSIWYG разработки — подготовка материала в нескольких форматах одновременно. Как правило, пользователю проще жестко указать форматирование, чем возиться со стилями (более того большинство пользователей даже не знают о такой возможности). Таким образом, становится практически невозможно изменить формат документа без нарушения его типографических качеств. Современные текстовые процессоры, как правило, предназначены либо для предпечатной подготовки текстовых документов, либо для подготовки их к публикации в Интернете. Однако, при всей своей мощности они не способны решить эти две задачи одновременно или делают это неэффективно.

Второй серьезной проблемой является так называемая проблема сложного форматирования. У нее есть две основные причины: автор видит только внешний вид документа, а не его реальную структуру, поэтому он не может оценить, как интерпретировал его указания WYSIWYG редак-

тор и различные пользователи используют разные приемы форматирования. Современные текстовые процессоры настолько сложны, что наиболее правильный способ изменения форматирования далеко не всегда очевиден. Поэтому создание конвертеров форматов требует очень сложной логики учитывающей все многочисленный способы форматирования, которые могут быть применены в исходном документе. Текстовое форматирование лишено этих проблем. Но все его достоинства становятся видны именно в контексте логического форматирования. Пользователь больше не должен заботиться о внешнем виде документа. За него это сделает система. Основная задача пользователя системы текстовой разметки правильно указать роли тех или иных фрагментов текста по средством расстановки соответствующих директив. Естественно это совершенно не обязательно делать в простом текстовом редакторе, вроде блокнота. Специализированный же текстовый редактор позволит пользователю быть уверенным в том, что компьютер правильно воспринял указанные им правила разметки. Это особенно важно, если над документом работает сразу группа людей, так как становится возможным контроль изменений и управления версиями документов в максимально удобной форме. Благодаря тому, что компьютер и пользователь гораздо лучше «понимают» друг друга появляется возможность использования так называемого семантического форматирования. При нем указывается даже не роль участка текста в документе, а его фактический смысл. Это позволяет автоматизировать процесс создания индексов, глоссариев, списков ссылок и другие трудо-

емкие процедуры. Кроме того существенно повышается качество обработки текста поисковыми роботами. От такого механизма текстового форматирования отказались на заре развития персональной вычислительной техники, однако сейчас эта технология становится все более популярной (на ее основе построены большинство Wiki систем, например Википедия, технологии разметки Textile, BBCode и другие).

Заменим, что все эти технологии так или иначе связаны с развитием средств коммуникаций, случившимся за последние годы. Еще пятнадцать лет назад основной целью редактирования текстовых документов была их предпечатная подготовка (хотя и там использовалось текстовое форматирование, но это касалось только профессиональных систем таких, как TeX являющейся по сути языком программирования). Десять лет назад с развитием Интернета основной задачей становится публикация материалов в сети. На сегодняшний день из-за колоссального роста объема информации в Интернете, на первое место встает не просто публикация информации, но и ее адаптация для быстрого и эффективного поиска. Наиболее перспективный путь развития Интернета в этом направлении — семантический веб. Благодаря этому текстовое форматирование получает второй шанс, как наиболее простой способ работы с семантической разметкой.