

2. Конспект лекций

1. О спецификации HTML 4.0

1.1 Организация спецификации

Спецификация состоит из следующих разделов:

Разделы 2 и 3: Введение в HTML 4.0

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 чтении Определений типов документов HTML (Document Type Definition - DTD).

Разделы 4 - 18: Справочное руководство по HTML 4.0

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 следующее:

- 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](#).

Приложения

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.

Ссылки

Список нормативных и информативных документов.

Указатели

Три указателя предоставляют читателям быстрый доступ к определению: [понятия](#), [элементы](#) и [атрибуты](#).

1.2 Соглашения

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 необ-

ходимую для разработки соответствующих средств информации.

Эту спецификацию можно использовать несколькими способами:

- **Прочсть 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:
 1. Каждая ссылка на элемент или атрибут связана с его определением в спецификации. Каждый элемент или атрибут определяется только в одном месте.
 2. На каждой странице имеются ссылки на указатели, поэтому Вы всегда сможете найти определение [элемента](#) или [атрибута](#), использовав не больше двух ссылок.
 3. На первых страницах трех разделов руководства к исходному оглавлению добавляется более подробная информация о каждом разделе.

1.2.1 Элементы и атрибуты

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, учитывается ли в значениях регистр. Подробнее см. раздел [информации о регистре](#).

1.2.2 Замечания и примеры

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.

2. Введение в HTML 4.0

2.1 Что такое World Wide Web?

World Wide Web (Web) - это сеть информационных ресурсов. Для того чтобы сделать эти ресурсы доступными наиболее широкой аудитории, в Web используются три механизма:

1. Единая схема наименования для поиска ресурсов в Web (например, URI).
2. Протоколы для доступа к именованным ресурсам через Web (например, HTTP).
3. Гипертекст для простого перемещения по ресурсам (например, HTML).

Связи между этими тремя механизмами становятся очевидными по прочтении этой спецификации.

2.1.1 Введение в URI

Каждый ресурс в Web - документ HTML, изображение, видеоклип, программа и т.д. - имеет адрес, который может быть закодирован с помощью *универсального идентификатора ресурсов (Universal Resource Identifier)*, или URI.

URI обычно состоят из трех частей:

1. Схема наименования механизма, используемого для доступа к ресурсу.
2. Имя машины, на которой располагается ресурс.
3. Имя собственно ресурса, заданное в виде пути.

Рассмотрите URI этой спецификации HTML на сервере W3C:

```
http://www.w3.org/TR/PR-html4/cover.html
```

Этот URI может читаться следующим образом: этот документ можно получить по протоколу HTTP (см. [RFC2068](#)), он располагается на машине www.w3.org, путь к этому документу - "/TR/PR-html4/cover.html". Кроме того, в документах в формате HTML Вы можете увидеть схемы "mailto" для электронной почты и "ftp" для протокола FTP.

Вот еще один пример URI. Он относится к почтовому ящику пользователя:

```
...текст...  
Комментарии отправляйте <A href="mailto:joe@someplace.com">Джо Кулу</A>.
```

Примечание. Большинство читателей уже знакомо с термином "URL", но не знает термина "URI". URL образуют подмножество более общей схемы наименования URI.

2.1.2 Идентификаторы фрагментов

Некоторые URI указывают на местоположение внутри ресурса. Этот тип URI заканчивается символом "#", за которым следует указатель (*идентификатор фрагмента*). Например, следующий URI указывает на фрагмент с именем section_2:

```
http://somesite.com/html/top.html#section_2
```

2.1.3 Относительные URI

Относительный URI не содержит информации о схеме наименования. Путь в нем указывает на ресурс на машине, на которой находится текущий документ. Относительные URI могут содержать компоненты относительного пути (например, ".." означает один уровень выше в иерархии) и [идентификаторы фрагментов](#).

Относительные URI [приводятся к полным URI](#) с помощью базового URI. В качестве примера приведения относительного URI предположим, что у нас имеется базовый URI "http://www.acme.com/support/intro.html". Относительный URI в следующей ссылке:

```
<A href="suppliers.html">Suppliers</A>
```

будет преобразован в полный URI "http://www.acme.com/support/suppliers.html", а относительный URI в следующем фрагменте

```
<IMG src="../../../icons/logo.gif" alt="logo">
```

будет преобразован в полный URI "http://www.acme.com/icons/logo.gif".

В HTML URI используются для:

- ссылки на другие документы или ресурсы (см. элементы [A](#) и [LINK](#)).
- ссылки на внешние таблицы стилей или скрипты (см. элементы [LINK](#) и [SCRIPT](#)).
- включения в страницу изображений, объектов или апплетов (см. элементы [IMG](#), [OBJECT](#), [APPLET](#) и [INPUT](#)).
- создания изображений-карт (см. элементы [MAP](#) и [AREA](#)).
- отправки форм (см. [FORM](#)).
- создания документов с использованием кадров (см. элементы [FRAME](#) и [IFRAME](#)).
- ссылок на внешние источники (см. элементы [Q](#), [BLOCKQUOTE](#), [INS](#) и [DEL](#)).
- ссылок на соглашения о метаданных, описывающих документ (см. элемент [HEAD](#)).

Подробнее об URI см. в разделе о типах [URI](#).

2.2 Что такое HTML?

Чтобы представить информацию для глобального использования, нужен универсальный язык, который понимали бы все компьютеры. Языком публикации, используемым в World Wide Web, является HTML (HyperText Markup Language - язык разметки гипертекстов).

HTML дает авторам средства для:

- публикации электронных документов с заголовками, текстом, таблицами, списками, фотографиями и т.д.
- загрузки электронной информации с помощью щелчка мыши на гипертекстовой ссылке.
- разработки форм для выполнения транзакций с удаленными службами, для использования в поиске информации, резервировании, заказе продуктов и т.д.
- включения электронных таблиц, видеоклипов, звуковых фрагментов и других приложений непосредственно в документы.

2.2.1 Краткая история HTML

Язык HTML был разработан Тимом Бернерс-Ли во время его работы в CERN и распространен браузером Mosaic, разработанным в NCSA. В 1990-х годах он добился особенных успехов благодаря быстрому росту Web. В это время HTML был расширен и дополнен. В Web очень важно использование одних и тех же соглашений HTML авторами Web-страниц и производителями. Это явилось причиной совместной работы над спецификациями языка HTML.

HTML 2.0 (ноябрь 1995, см. [\[RFC1866\]](#)) был разработан под эгидой Internet Engineering Task Force (IETF) для упорядочения общепринятых положений в конце 1994 года. HTML+ (1993) и HTML 3.0 (1995, см. [\[HTML30\]](#)) - это более богатые версии языка HTML. Несмотря на то, что в обычных дискуссиях согласие никогда не было достигнуто, эти черновики привели к принятию ряда новых свойств. Усилия Рабочей группы World Wide Web Consortium по HTML в упорядочении общепринятых положений в 1996 привели к версии HTML 3.2 (январь 1997, см. [\[HTML32\]](#)). Изменения по отношению к HTML 3.2 перечислены в [Приложении А](#)

Большинство людей признают, что документы HTML должны работать в различных браузерах и на разных платформах. Достижение совместимости снижает расходы авторов, поскольку они могут разрабатывать только одну версию документа. В противном случае возникает еще больший риск, что Web будет представлять собой смесь личных несовместимых форматов, что в конечном счете приведет к снижению коммерческого потенциала Web для всех участников.

В каждой версии HTML предпринималась попытка отразить все большее число соглашений между работниками и пользователями этой индустрии, чтобы усилия авторов не были потрачены впустую, а их документы не стали бы нечитаемыми в короткий срок.

Язык HTML разрабатывался с той точки зрения, что все типы устройств должны иметь возможность использовать информацию в Web: персональные компьютеры с графическими дисплеями с различным разрешением и числом цветов, сотовые телефоны, переносные устройства, устройства для вывода и ввода речи, компьютеры с высокой и низкой частотой и т.д.

2.3 HTML 4.0

В HTML 4.0 вводятся механизмы таблиц стилей, скриптов, кадров, внедрения объектов, улучшенная поддержка разных направлений письма и направления справа налево, таблицы с большим количеством возможностей и новые свойства форм, обеспечивая лучшие возможности доступа для людей с физическими недостатками.

2.3.1 Интернационализация

Эта версия HTML разработана с помощью экспертов в области интернационализации, так что документы можно писать на любом языке и легко передавать их по всему миру. Это достигается за счет использования [\[RFC2070\]](#), относящегося к интернационализации HTML. Важным шагом стало принятие стандарта ISO/IEC:10646 (см. [\[ISO10646\]](#)) в качестве набора символов для документов HTML. Это наиболее содержательный стандарт в мире, в котором решены вопросы представления национальных символов, направления письма, пунктуации и других языковых вопросов. HTML теперь предоставляет лучшую

поддержку различных языков в одном документе. Это обеспечивает более эффективное индексирование документов для поисковых машин, типографию высшего качества, преобразование текста в речь, более удобные переносы и т.д.

2.3.2 Доступность

Поскольку сообщество Web растет, и возможности и умения его членов различаются, очень важно, чтобы основные технологии соответствовали потребностям. Язык HTML разработан так, чтобы сделать Web-страницы более доступными для пользователей с физическими недостатками. В HTML 4.0 имеются следующие дополнения, продиктованные соображениями доступности:

- усилено разделение структуры и представления документа, что побуждает использовать таблицы стилей вместо элементов и атрибутов представления языка HTML.
- улучшены формы, включена возможность назначения клавиш доступа, возможность семантической группировки управляющих элементов формы, семантической группировки вариантов в тэге [SELECT](#) и активные метки.
- добавлена возможность разметки текстового описания включенного объекта (с помощью элемента [OBJECT](#)).
- введен новый механизм действия изображений-карт на стороне клиента (элемент [MAP](#)), который позволяет авторам интегрировать изображения и текстовые ссылки.
- альтернативный текст для изображений, включаемых с помощью [IMG](#), обязателен.
- добавлена поддержка атрибутов [title](#) и [lang](#) во всех элементах.
- добавлена поддержка элементов [ABBR](#) и [ACRONYM](#).
- более широкий диапазон целевых устройств (телетайп, шрифт Брайля и т.д.) для использования в таблицах стилей.
- улучшены таблицы, включена поддержка заголовков, групп столбцов и механизмов для упрощения невизуального представления документа.
- добавлены длинные описания таблиц, изображений, кадров и т.д.

Авторы, разрабатывающие страницы с учетом доступности, получают не только эту возможность, но также и некоторые другие: хорошо разработанные документы HTML с разделенными структурой и представлением легче адаптируются к новым технологиям.

Примечание. Подробнее о разработке доступных документов HTML Вы можете прочесть в [\[WAIGUIDE\]](#).

2.3.3 Таблицы

Новая модель таблиц в HTML основана на [\[RFC1942\]](#). Теперь авторы имеют большую власть над структурой и компоновкой таблицы (например, группы столбцов). Возможность дизайнеров рекомендовать ширину столбцов позволяет агентам пользователей отображать данные таблицы постепенно (по мере получения) и не ждать всю таблицу до создания изображения.

Примечание. Во время написания этого документа некоторые средства разработки документов в формате HTML широко использовали для форматирования страниц таблицы, что вызывало проблемы совместимости.

2.3.4 Составные документы

В HTML теперь имеется стандартный механизм для внедрения объектов и приложений в документы HTML. Элемент [OBJECT](#) (а также более специфичные элементы, его приемники, [IMG](#) и [APPLET](#)) обеспечивает механизм включения в документ изображений, видео-файлов, звуковых файлов, математических выражений, специализированных приложений и других объектов. Он также позволяет авторам указывать иерархию или альтернативный способ создания изображения для агентов пользователей, не поддерживающих указанный способ создания изображения.

2.3.5 Таблицы стилей

Таблицы стилей упрощают разметку HTML и существенно снижают участие языка HTML

в представлении документа. Они предоставляют авторам и пользователям возможность управлять представлением документов - шрифтами, выравниванием, цветами и т.д. Информацию о стиле можно указать для отдельных элементов или групп элементов, в документе HTML или во внешних таблицах стилей.

Механизмы связи таблиц стилей с документами не зависят от языка таблиц стилей.

До появления таблиц стилей возможности управления созданием изображения у авторов были ограничены. В HTML 3.2 был включен ряд атрибутов и элементов для управления выравниванием, размером шрифта и цветом текста. Авторы также использовали для компоновки страниц таблицы и изображения. Поскольку на обновление браузеров у пользователей уйдет довольно долгое время, эти средства еще будут использоваться в течение какого-то времени. Однако поскольку таблицы стилей обеспечивают более мощные механизмы представления, World Wide Web Consortium существенно сократит число элементов и атрибутов представления в HTML. В этой спецификации элементы и атрибуты, которые могут быть впоследствии исключены, помечены как "нежелательные". Они сопровождаются примерами полдостижения того же эффекта с помощью других элементов или таблиц стилей.

2.3.6 Скрипты

С помощью скриптов авторы могут создавать динамичные Web-страницы (например, "интеллектуальные формы", изменяющиеся по мере заполнения их пользователем) и использовать HTML как средство построения сетевых приложений.

Механизмы, обеспечивающие включение скриптов в документы HTML, не зависят от языка скриптов.

2.3.7 Печать

Иногда авторы хотят упростить для пользователей печать текущего документа. Если документ является частью другого документа, отношения между ними можно описать с помощью элемента HTML [LINK](#) или языка описания ресурсов (Resource Description Language - RDF) W3C (см. [RDF](#)).

2.4 Создание документов в формате HTML 4.0

Авторам и разработчикам для работы с HTML 4.0 рекомендуется ознакомиться со следующими общими принципами.

2.4.1 Разделение структуры и представления

HTML происходит из SGML, который всегда был языком определения структурной разметки. По мере развития HTML все большее количество его элементов и атрибутов для представления заменяется другими механизмами, в частности, таблицами стилей. Опыт показывает, что отделение структуры документа от аспектов его представления снижает стоимость обслуживания широкого диапазона платформ, носителей и т.д. и упрощает изменение документов.

2.4.2 Универсальность доступа к Web

Чтобы сделать свой Web-сервер доступным для всех пользователей, особенно для пользователей с физическими недостатками, авторы должны предполагать, как их документы могут отображаться на различных платформах: речевых браузерах, программах чтения азбуки Брайля и т.д. Мы не рекомендуем авторам ограничивать творческий процесс, но рекомендуем предусматривать альтернативные методы подачи информации. HTML предлагает ряд таких механизмов (например, атрибут [alt](#), атрибут [accesskey](#) и т.д.)

Авторам также следует иметь в виду, что к их документам могут обращаться пользователи с другой конфигурацией компьютеров. Для корректной интерпретации документов авторам следует включать в свои документы информацию о языке и направлении письма в тексте, о кодировке документа и прочую подобную информацию.

2.4.3 Помощь агентам пользователей в последовательном создании изображений

При тщательной разработке таблиц и использовании новых возможностей HTML 4.0 авторы могут ускорить отображение документов агентами пользователей. Авторы могут прочесть здесь о том, как создавать таблицы для последовательного представления (см. элемент [TABLE](#)). Разработчики могут получить информацию об алгоритмах последовательного представления в [замечаниях о таблицах](#) в приложении.

3 Об SGML и HTML

В этом разделе документа Вы познакомитесь с SGML и узнаете о его отношении к HTML. Полное описание SGML не входит в этот стандарт (см. [ISO8879](#)).