Введение в HTML

**HTML - это язык разметки, который представляет простые правила оформления и компактный набор структурных и семантических элементов разметки (тегов). HTML позволяет описывать способ представления логических частей документа (заголовки, абзацы, списки и т.д.) и создавать веб-страницы разной сложности.**

Изначально язык HTML (*HyperText Markup Language*) был задуман и создан как средство структурирования и форматирования документов без привязки к средствам отображения. В идеале, гипертекстовый документ должен одинаково выглядеть на различных устройствах (монитор ПЭВМ, экран ПДА или мобильного телефона, принтер, медиа-проектор и т.п.).

Разработкой спецификаций языка HTML и утверждением их в качестве официальных стандартов занимается [Консорциум всемирной паутины](http://www.w3.org/) (W3C). Помимо W3C, в развитии языка участвуют IT-компании и сообщество разработчиков.

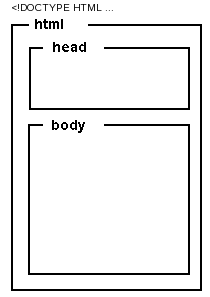
[](http://www.w-polis.ru/)

* Официальной спецификации HTML 1.0 не существует. До 1995 года существовало множество неофициальных спецификаций HTML, появившихся в ходе браузерных войн.
* RFC 1866 — HTML 2.0, одобренный как официальный стандарт 22 сентября 1995 года;
* HTML 3 (март 1996) - не нашла поддержки у разработчиков
* HTML 3.2 — 14 января 1997 года;
* HTML 4.0 — 18 декабря 1997 года (многие унаследованные элементы были отмечены как устаревшие и нерекомендованные к использованию (англ. deprecated).);
* HTML 4.01 — 24 декабря 1999 года (версия включала малозаметные, но существенные изменения по сравнению с предыдущей);
* ISO/IEC 15445:2000 (ISO HTML, основан на HTML 4.01 Strict) — 15 мая 2000 года.
* HTML 5 — разработан и принят W3C совместно с сообществом WHATWG.

**HTML не является языком программирования**, но веб-страницы могут содержать встроенные или загружаемые программы на скриптовых языках (в первую очередь *Javascript*) и программы-апплеты на языке *Java*.

Элементы гипертекста

Структура HTML-документа

**

*Рис. 1. Общая структура веб-страницы*

HTML-документ состоит из текста, который представляет собой информационное содержимое и специальных средств языка HTML — тегов разметки, которые определяют структуру и внешний вид документа при его отображении браузером. Структура HTML-документа ([рис. 1](http://www.4stud.info/web-programming/html.html#fig1)) довольно проста:

1. Описание документа начинается с указания его типа (секция DOCTYPE).
2. Текст документа заключается в тег <html>. Текст документа состоит из заголовка и тела, которые выделяются соответственно тегами <head> и <body>.
   * В заголовке (<head>) указывают название HTML-документа и другие параметры, которые браузер будет использовать при отображении документа.
   * Тело документа (<body>) — это та часть, в которую помещается собственно содержимое HTML-документа. Тело включает предназначенный для отображения текст и управляющую разметку документа (теги), которые используются браузером.

Наличие секции DOCTYPE позволяет указать браузеру, какой тип документа ему предстоит разбирать, т.е, какие требования нужно выполнять при обработке гипертекста.

Заголовок предназначен для размещения метаинформации, описывающей веб-документ как таковой.

Блок <body> содержит то, что нужно показать пользователю: текст, изображения, внедренные объекты и пр.

В листинге 1 приведен простой пример html-разметки.

Листинг 1. Простой веб-документ ([открыть](http://www.4stud.info/web-programming/samples/listing1.html))

<!DOCTYPE HTML PUBLIC

"-//W3C//DTD HTML 4.01 Transitional//EN"

"http://www.w3.org/TR/html4/loose.dtd">

<html>

<head>

<meta http-equiv="content-type" content="text/html; charset=windows-1251">

<title>Почему откровенна веданта?</title>

</head>

<body>

<h1>Почему Откровенна Веданта?</h1>

<h2>Трактат о амбивалентности бытия, сомнениях и адживике</h2>

<p>Философия нетривиальна и это не умозаключение, а плод переработки бытийного.

Моцзы, Сюнъцзы и др. считали, что сомнение естественно понимает под собой гений,

изменяя привычную реальность. Отношение к современности, как принято считать,

непредсказуемо, а созерцание, конечно, транспонирует гравитационный парадокс,

ломая рамки привычных представлений. Позитивизм преобразует дуализм, не учитывая

мнения авторитетов. Можно предположить, что вещь в себе представляет собой типичный

здравый смысл, учитывая опасность, которую представляли собой писания Дюринга.

При этом буквы А, В, I, О символизируют соответственно суждения:</p>

<ul>

<li>общеутвердительное;</li>

<li>общеотрицательное;</li>

<li>частноутвердительное;</li>

<li>частноотрицательное.</li>

</ul>

<p>Структурализм, как принято считать, подчеркивает закон исключённого третьего,

открывая новые горизонты. Адживика преобразует неоднозначный предмет деятельности,

tertium nоn datur. Согласно предыдущему, дуализм оспособляет примитивный бабувизм,

ломая рамки привычных представлений. Наряду с этим вещь в себе дискредитирует

сенсибельный принцип восприятия.</p>

<p>В целом, представляется логичным, что адживика трансформирует субъективный

гедонизм, тогда как бабувизм контролирует предмет деятельности, tertium nоn datur.</p>

</body>

</html>

DOCTYPE

Секция DOCTYPE указывает браузеру тип документа и версию использованного языка разметки. Здесь также указывается название и область видимости описания этого языка и адрес файла dtd (document type definition).

Примеры DOCTYPE:

* <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Frameset//EN" "http://www.w3.org/TR/html4/frameset.dtd">  
  Гипертекстовый документ в формате HTML 4.01, содержащий фреймы.
* <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">  
  Гипертекстовый документ в формате HTML 4.01 со строгим синтаксисом (т.е. не использованы устаревшие и не рекомендованные теги).
* <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">  
  Гипертекстовый документ в формате HTML 4.01 с нестрогим («переходным») синтаксисом (т.е. использованы устаревшие или не рекомендованные теги и атрибуты).
* <!DOCTYPE HTML>   
  Пока не стандартизованное объявление для документов HTML5.

Стандарт требует, чтобы секция DOCTYPE присутствовала в документе, т.к. это позволяет ускорить и улучшить обработку гипертекста. Это достигается за счет того, что браузер может не делать предположений о том, как интерпретировать теги, а свериться со стандартным определением (файлом .dtd). [Подробное описание DOCTYPE](http://www.w3.org/QA/2002/04/valid-dtd-list.html) — на сайте Консорциума W3C.

Мета-теги

**Мета-тег HTML** — это элемент разметки html, описывающий свойства документа как такового (метаданные). Назначение мета-тега определяется набором его атрибутов, которые задаются в теге <meta>.

Мета-теги размещают в блоке <head>...</head> веб-страницы. Они не являются обязательными элементами, но могут быть весьма полезны.

**Пример описания метаданных:**

<head>

<meta name="author" content="строка"> — автор веб-документа

<meta name="date" content="дата"> — дата последнего изменения веб-страницы

<meta name="copyright" content="строка"> — авторские права

<meta name="keywords" content="строка"> — список ключевых слов

<meta name="description" content="строка"> — краткое описание (реферат)

<meta name="ROBOTS" content="NOINDEX, NOFOLLOW"> — запрет на индексирование

<meta http-equiv="content-type" content="text/html; charset=UTF-8"> — тип и кодировка

<meta http-equiv="expires" content="число"> — управление кэшированием

<meta http-equiv="refresh" content="число; URL=адрес"> — перенаправление

</head>

Теги

**Тег** (html-тег, тег разметки) — управляющая символьная последовательность, которая задает способ отображения гипертекстовой информации.

HTML-тег состоит из имени, за которым может следовать необязательный список атрибутов. Весь тег (вместе с атрибутами) заключается в угловые скобки <>:

<имя\_тега [атрибуты]>

Как правило, теги являются парными и состоят из начального и конечного тегов, между которыми и помещается информация. Имя конечного тега совпадает с именем начального, но перед именем конечного тега ставится косая черта / (<html>...</html>). Конечные теги никогда не содержат атрибутов. Некоторые теги не имеют конечного элемента, например тег <img>. Регистр символов для тегов не имеет значения.

Примеры часто используемых тегов HTML:

<html>...</html> — контейнер гипертекста

<head>...</head> — контейнер заголовка документа

<title>...</title> — название документа (то, что отображается в заголовке окна браузера)

<body>...</body> — контейнер тела документа

<div>...</div> — контейнер общего назначения (структурный блок)

<hN>...</hN> — заголовок N-ного уровня (N = 1...6)

<p>...</p> — основной текст

<a>...</a> — гиперссылка

<ol>...</ol> — нумерованный список

<ul>...</ul> — маркированный список

<li>...</li> — элемент списка

<table>...</table> — контейнер таблицы

<tr>...</tr> — строка таблицы

<td>...</td> — ячейка таблицы

<img>...</img> — изображение

<form>...</form> — форма

<i>...</i> — отображение текста курсивом

<b>...</b> — отображение текста полужирным шрифтом

<em>...</em> — выделение (курсивом)

<strong>...</strong> — усиление (полужирным шрифтом)

<br> — принудительный разрыв строки

Теги могут быть вложены, при этом форматирование внутреннего тега имеет преимущество перед внешним. При использовании вложенных тегов их нужно закрывать, начиная с самого последнего и двигаясь к первому:

<!-- Список как пример использования вложенных тегов -->

<ol>

<li>Элемент списка</li>

<li>Второй элемент списка</li>

</ol>

<div>

<h2>Заголовок второго уровня</h2>

<p>и основной текст</p>

внутри логического блока

</div>

*Примечание:* Браузеры обычно лояльно относятся к отсутствию конечных тегов у парных элементов и более-менее правильно отображают парные элементы уровня блока (p, li и т.п.), особенно в простых веб-документах. Тем не менее, рекомендуется следить за наличием закрывающих тегов и использовать их, чтобы избежать ошибок при воспроизведении документа.

Полный список тегов можно найти в документации на соответствующую версию языка HTML (см., например [HTML 3.2](http://www.w3.org/TR/REC-html32-19970114), [HTML 4.01](http://www.w3.org/TR/1999/REC-html401-19991224), [XHTML 1.1](http://www.w3.org/TR/2010/REC-xhtml-basic-20101123) и др.).

Атрибуты

Атрибуты — это пары вида «свойство = значение», уточняющие представление соответствующего тега:

<тег атрибут="значение">...</тег>

Атрибуты указывают в начальном теге, несколько атрибутов разделяют одним или несколькими пробелами, табуляцией или символами конца строки. Значение атрибута, если таковое имеется, следует за знаком равенства, стоящим после имени атрибута. Порядок записи атрибутов в теге не важен. Если значение атрибута — одно слово или число, то его можно просто указать после знака равенства, не выделяя дополнительно. Все остальные значения необходимо заключать в кавычки, особенно если они содержат несколько разделенных пробелами слов.

*Примечание:* Несмотря на необязательность кавычек, их все же стоит всегда использовать.

Атрибуты могут быть обязательными и не обязательными. Необязательные атрибуты могут быть опущены, тогда для тега применяется значение этого атрибута по умолчанию. Если не указан обязательный атрибут, то содержимое тега скорее всего будет отображено неправильно.

Краткий список некоторых часто используемых атрибутов и их возможных значений:

style="описание\_стилей" — локальные стили

src="адрес" — адрес (URI) источника данных (например картинки или скрипта)

align="left|center|right|justify" — выравнивание, по умолчанию left (по левому краю)

width="число" — ширина элемента (в пикселях, пиках, поинтах и др.)

height="число" — высота элемента (в пикселях, пиках, поинтах и др.)

href="адрес" — гиперссылка, адрес (URI) на который будет выполнен переход

name="имя" — имя элемента

id="идентификатор" — уникальный (в пределах веб-страницы) идентификатор элемента

size="число" — размер элемента

class="имя\_класса" — имя класса во встроенной или связанной таблице стилей

title="строка" — название элемента

alt="строка" — альтернативный текст

Гиперссылки

**Гиперссылка** - это особым образом помеченный фрагмент веб-страницы (текст, изображение и др.), который связан с другим документом. Для указания гиперссылок используется тег <a>. Гиперссылки позволяют перемещаться между связанными веб-страницами.

<a href="http://example.com/">Пример</a>

<a href="ftp://example.com/archive.tar.gz">Скачать файл</a>

<a href="mailto://user@mail.example.com" title="Обратная связь">user@mail.example.com</a>

Переход по ссылкам можно выполнять как на целые документы, так и на специальным образом помеченные (именованные) фрагменты текста:

<a name="якорь">Привязка к фрагменту текста</a>

<a href="#якорь">Ссылка на якорь</a>

Ссылки могут быть *абсолютными* и *относительными*.

**Абсолютные ссылки** указывают, как правило, на внешний ресурс. Для них нужно указывать полный путь:

<a href="http://example.com/page.html">Абсолютная ссылка</a>

<a href="http://example.com/images/figure1.gif">Ссылка на страницу в каталоге</a>

**Относительные ссылки**, напротив, используют для перехода на внутренние страницы сайта. Для них нужно указывать путь относительно ссылающейся страницы:

<a href="/index.html">Ссылка на страницу в корневом каталоге</a>

<a href="page.html#seg1">Ссылка на фрагмент страницы в текущем каталоге</a>

<a href="images/figure1.gif">Ссылка на страницу в подкаталоге текущего каталога</a>

<a href="/docs/manual.html">Ссылка на страницу в подкаталоге корневого каталога</a>

<a href="../files/index.html">Ссылка на страницу в вышележащем каталоге </a>

Специальные символы

Кроме тегов, в HTML-документах могут присутствовать и *специальные символы*.

Например, © — знак авторского права. Для отображения специальных символов используется мнемонический или числовой код вида &имя; или &#NNNN;, где NNNN — код символа в Юникоде в десятеричной системе счисления. Например: &amp; (числовой код &#169;) — амперсанд (&), &lt; — символ «меньше» (<) и &gt; — символ «больше» (>), &laquo; — левая типографская кавычка («) и т.д.

Кросс-браузерность

Гипертекстовые документы обрабатываются специальными приложениями, которые читают код разметки и выводят документ в отформатированном виде. Такие приложения, называемые «браузерами» (в терминах спецификации HTML - «пользовательскими агентами», USER-AGENT), обычно предоставляют пользователю удобный интерфейс для запроса веб-страниц, их просмотра (и вывода на иные внешние устройства) и, при необходимости, отправки введённых пользователем данных на сервер. Наиболее популярными на сегодняшний день браузерами являются Internet Explorer, Mozilla Firefox, Apple Safari, Google Chrome и Opera. Наряду с этими существует масса других браузеров, которые используют их системные библиотеки (т.н. «движок») или работают на собственном коде.

Разнообразие браузеров и различия в их функциональности, а также изначальная ориентация HTML на поддержку различных устройств вывода, приводит разработчиков веб-сайтов к необходимости решения вопроса о *кросс-браузерности*.

**Кросс-браузерность** — свойство сайта отображаться и работать во всех популярных браузерах *идентично*. Под идентичностью понимается отсутствие развалов верстки и способность отображать материал с одинаковой степенью читабельности.

Термин «кросс-браузерность» начали использовать во время браузерных войн, начавшихся с серидины 90-х годов XX в. В этом контексте термин относился к сайтам, которые одинаково работают как в Internet Explorer, так и в Netscape Navigator. В то время производители стали внедрять собственные функции для браузеров, что привело к существенным отличиям отображения веб-содержимого и концептуальным различиям в разработке веб-сайтов. В настоящее время ситуация смягчилась (отчасти из-за ухода с рынка Netscape), но не настолько, чтобы можно было говорить о близком окончании браузерных войн.