

Web、网页、浏览器

Web

Web (World Wide Web) 即全球广域网，也称为万维网。

我们常说的 web端 就是网页端。

网页

网页是构成网站的基本元素。网页主要由文字、图像和超链接等元素构成。当然，除了这些元素，网页中还可以包含音频、视频以及Flash等。

我们在浏览器上输入网址后，打开的任何一个页面，都是属于网页。

浏览器

浏览器是网页运行的平台，常见的浏览器有谷歌 (Chrome)、Safari、火狐 (Firefox)、IE、Edge、Opera等。

Web标准

W3C组织

W3C: World Wide Web Consortium, 万维网联盟组织，用来制定web标准的机构（组织）。

W3C 万维网联盟是国际最著名的标准化组织。1994年成立后，至今已发布近百项相关万维网的标准，对万维网发展做出了杰出的贡献。

W3C 组织就类似于现实世界中的联合国。

为什么要遵循WEB标准呢？因为很多浏览器的浏览器内核不同，导致页面解析出来的效果可能会有差异，给开发者增加无谓的工作量。因此需要指定统一的标准。

Web 标准

Web标准：制作网页要遵循的规范。

Web标准不是某一个标准，而是由W3C组织和其他标准化组织制定的一系列标准的集合。

1、Web标准包括三个方面：

- 结构标准 (HTML)：用于对网页元素进行整理和分类。
- 表现标准 (CSS)：用于设置网页元素的版式、颜色、大小等外观样式。
- 行为标准 (JS)：用于定义网页的交互和行为。

根据上面的Web标准，可以将 Web前端分为三层，如下。

2、Web前端分三层：

- HTML (HyperText Markup Language)：超文本标记语言。从**语义**的角度描述页面的**结构**。相当于人的身体组织结构。

- CSS（Cascading Style Sheets）：层叠样式表。从**审美**的角度美化页面的**样式**。相当于人的衣服和打扮。
- JS：JavaScript。从**交互**的角度描述页面的**行为**。相当于人的动作，让人有生命力。

浏览器的组成

浏览器分成两部分：

- 1、渲染引擎（即：浏览器内核）
- 2、JS 引擎

1、渲染引擎（浏览器内核）

浏览器所采用的「渲染引擎」也称之为「浏览器内核」，用来解析 HTML与CSS。渲染引擎决定了浏览器如何显示网页的内容以及页面的格式信息。

渲染引擎是浏览器兼容性问题出现的根本原因。

渲染引擎的英文叫做 Rendering Engine。通俗来说，它的作用就是：读取网页内容，计算网页的显示方式并显示在页面上。

常见浏览器的内核如下：

浏览器	内核
chrome	Blink
欧鹏	Blink
360安全浏览器	Blink
360极速浏览器	Blink
Safari	Webkit
Firefox 火狐	Gecko
IE	Trident

备注：360的浏览器，以前使用的IE浏览器的Trident内核，但是现在已经改为使用 chrome 浏览器的Blink内核。

另外，移动端的浏览器内核是什么？大家可以自行查阅资料。

2、JS 引擎

也称为 JS 解释器。用来解析网页中的 JavaScript 代码，对其处理后再运行。

浏览器本身并不会执行JS代码，而是通过内置 JavaScript 引擎（解释器）来执行 JS 代码。JS 引擎执行代码时会逐行解释每一句源码（转换为机器语言），然后由计算机去执行。所以JavaScript 语言归为脚本语言，会逐行解释执行。

常见浏览器的JS 引擎如下：

浏览器	JS 引擎
chrome / 欧鹏	V8
Safari	Nitro
Firefox 火狐	SpiderMonkey (1.0-3.0) / TraceMonkey (3.5-3.6) / JaegerMonkey (4.0-)
Opera	Linear A (4.0-6.1) / Linear B (7.0-9.2) / Futhark (9.5-10.2) / Carakan (10.5-)
IE	Trident

HTML的概述

HTML的概念

HTML 全称为 HyperText Markup Language，译为**超文本标记语言**。

HTML 不是一种编程语言，是一种描述性的**标记语言**。

作用：HTML是负责描述文档**语义**的语言。

概念：超文本

所谓的超文本，有两层含义：

- (1) 图片、音频、视频、动画、多媒体等内容，成为超文本，因为它们超出了文本的限制。
- (2) 不仅如此，它还可以从一个文件跳转到另一个文件，与世界各地主机的文件进行连接。即：超级链接文本。

概念：标记语言

HTML 不是一种编程语言，是一种描述性的**标记语言**。这主要有两层含义：

(1) **标记语言是一套标记标签**。比如：标签 `<a>` 表示超链接、标签 `` 表示图片、标签 `<h1>` 表示一级标题等等，它们都是属于 HTML 标签。

说的通俗一点就是：网页是由网页元素组成的，这些元素是由 HTML 标签描述出来，然后通过浏览器解析，就可以显示给用户看了。

(2) 编程语言是有编译过程的，而标记语言没有编译过程，HTML标签是直接由浏览器解析执行。

HTML是负责描述文档语义的语言

HTML 格式的文件是一个纯文本文件（就是用txt文件改名而成），用一些标签来描述语义，这些标签在浏览器页面上是无法直观看到的，所以称之为“超文本标记语言”。

接下来，我们需要学习 HTML 中的很多“标签对儿”，这些“标签对儿”能够给文本不同的语义。

比如，面试的时候问你，`<h1>` 标签有什么作用？

- 正确答案：给文本增加主标题的语义。
- 错误答案：给文字加粗、加黑、变大。

关乎“语义”的更深刻理解，等接下来我们学习了各种标签，就明白了。

HTML的专有名词

- 网页：由各种标记组成的一个页面就叫网页。
- 主页(首页)：一个网站的起始页面或者导航页面。index
- 标记：比如 `<p>` 称为开始标记，`</p>` 称为结束标记，也叫标签。每个标签都规定好了特殊的含义。
- 元素：比如 `<p>内容</p>` 称为元素。
- 属性：给每一个标签所做的辅助信息。
- XHTML：符合XML语法标准的HTML。
- DHTML：dynamic，动态的。`javascript + css + html` 合起来的页面就是一个 DHTML。
- HTTP：超文本传输协议。用来规定客户端浏览器和服务端交互时数据的一个格式。
- SMTP：邮件传输协议。
- FTP：文件传输协议。

书写第一个 HTML 页面

VS Code 执行代码需要代码在一个文件夹中。

我们打开 VS Code 软件，新建一个文件，名叫 `test.html`（注意，文件名是 `test`，后缀名是 `html`），保存到本地。

紧接着，在文件里，输入 `html:5`，然后按一下键盘上的 `Tab` 键，就可以自动生成如下内容：

```
<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="UTF-8">
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<title>Document</title>
</head>
<body>

</body>
</html>
```

上面的内容，就是 html 页面的骨架。我们在此基础上，新增几个标签，完整代码如下：

```
<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="UTF-8">
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<title>Document</title>
</head>
<body>
<h3>我是三级标题</h3>
<img src="" alt="">
<a href="https://www.baidu.com">我是超链接，可以点击一下</a>
</body>
</html>
```

标签写完之后，我们用 chrome 浏览器打开上面这个 `test.html` 文件，看看页面效果：

到此，第一个简单的 HTML 页面就写完了。是不是很有成就感？

HTML结构详解

HTML标签通常是成对出现的（**双边标记**），比如 `<div>` 和 `</div>`；也有少部分单标签（**单边标记**），如：`
`、`<hr />` 和 `` 等。

属性与标记之间、各属性之间需要以空格隔开。属性值以双引号括起来。

html骨架标签分类

标签名	定义	说明
<code><html></html></code>	HTML标签	页面中最大的标签，称为根标签
<code><head></head></code>	文档的头部	注意在head标签中我们必须要设置的标签是title
<code><title></title></code>	文档的标题	让页面拥有一个属于自己的网页标题
<code><body></body></code>	文档的主体	元素包含文档的所有内容，页面内容基本都是放到body里面的

1、文档声明头

任何一个标准的HTML页面，第一行一定是一个以 `<!DOCTYPE` 开头的语句。这一行，就是文档声明头，即 DocType Declaration，简称DTD。

DTD可告知浏览器文档使用哪种 HTML 或 XHTML 规范。

在HTML5中极大的简化了DTD，也就是说HTML5中就没有XHTML了。

HTML5的DTD（文档声明头）如下：

```
<!DOCTYPE html>
```

2、页面语言 lang

下面这行标签，用于指定页面的语言类型：

```
<html lang="en">
```

最常见的语言类型有两种：

- en：定义页面语言为英语。
- zh-CN：定义页面语言为中文。

3、头标签 head

html5 的比较完整的骨架：

```
<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="UTF-8">
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
<meta name="Keywords" content="厉害很厉害" />
<meta name="Description" content="极客时间是中国领先的互联网技术培训公司。" />
<title>Document</title>
</head>
<body>

</body>
</html>
```

面试题：

- 问：网页的head标签里面，表示的是页面的配置，有什么配置？
- 答：字符集、关键词、页面描述、页面标题、IE适配、视口、iPhone小图标等等。

头标签内部的常见标签如下：

- `<title>`：指定整个网页的标题，在浏览器最上方显示。
- `<base>`：为页面上的所有链接规定默认地址或默认目标。
- `<meta>`：提供有关页面的基本信息。
- `<link>`：定义文档与外部资源的关系。

meta 标签：

meta表示“元”。“元”配置，就是表示基本的配置项目。

常见的几种 meta 标签如下：

(1) 字符集 charset：

```
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
```

字符集用meta标签中的 `charset` 定义，`charset`就是character set（即“字符集”），即**网页的编码方式**。

字符集(Character set)是多个字符的集合。计算机要准确的处理各种字符集文字，需要进行字符编码，以便计算机能够识别和存储各种文字。

上面这行代码非常关键，是必须要写的代码，否则可能导致乱码。比如你保存的时候，meta写的和声明的不匹配，那么浏览器就会出现乱码。

utf-8是目前最常用的字符集编码方式，常用的字符集编码方式还有gbk和gb2312等。关于“编码方式”，我们在下一段会详细介绍。

(2) 视口 viewport：

```
<meta name="viewport" content="width=device-width, initial-scale=1.0">
```

`width=device-width`：表示视口宽度等于屏幕宽度。

(3) 定义“关键词”：

举例如下：

```
<meta name="Keywords" content="网易, 邮箱, 游戏, 新闻, 体育, 娱乐, 女性, 亚运, 论坛, 短信" />
```

这些关键词，就是告诉搜索引擎，这个网页是干嘛的，能够提高搜索命中率。让别人能够找到你，搜索到你。

(4) 定义“页面描述”：

meta除了可以设置字符集，还可以设置关键字和页面描述。

只要设置Description页面描述，那么百度搜索结果，就能够显示这些语句，这个技术叫做**SEO**（search engine optimization，搜索引擎优化）。

设置页面描述的举例：

```
<meta name="Description" content="极客时间是中国领先的互联网技术培训公司。" />
```

上面的几种 `<meta>` 标签都不用记，但是另外还有一个 `<meta>` 标签是需要记住的：

```
<meta http-equiv="refresh" content="3;https://www.baidu.com">
```

上面这个标签的意思是说，3秒之后，自动跳转到百度页面。

title 标签：

用于设置网页标题：

```
<title>网页的标题</title>
```

title标签也是有助于SEO搜索引擎优化的。

base标签：

```
<base href="/">
```

base 标签用于指定基础的路径。指定之后，所有的 a 链接都是以这个路径为基准。

4、<body> 标签

`<body>`：用于定义HTML文档所要显示的内容，也称为主体标签。我们所写的代码必须放在此标签内。

`<body>` 标签的属性有：

- `bgcolor`：设置整个网页的背景颜色。
- `background`：设置整个网页的背景图片。
- `text`：设置网页中的文本颜色。
- `leftmargin`：网页的左边距。IE浏览器默认是8个像素。
- `topmargin`：网页的上边距。
- `rightmargin`：网页的右边距。
- `bottommargin`：网页的下边距。

计算机编码介绍

计算机，不能直接存储文字，存储的是编码。

计算机只能处理二进制的数数据，其它数据，比如：0-9、a-z、A-Z，这些字符，我们可以定义一套规则来表示。假如：A用110表示，B用111表示等。

ASCII码：美国发布的，用1个字节（8位二进制）来表示一个字符，共可以表示 $2^8=256$ 个字符。美国的国家语言是英语，只要能表示0-9、a-z、A-Z、特殊符号。

ANSI编码：每个国家为了显示本国的语言，都对ASCII码进行了扩展。用2个字节（16位二进制）来表示一个汉字，共可以表示 $2^{16}=65536$ 个汉字。例如：

中国的ANSI编码是GB2312编码（简体），对6763汉字进行编码，含600多特殊字符。另外还有GBK（简体）。

日本的ANSI编码是JIS编码。

台湾的ANSI编码是BIG5编码（繁体）。

GBK：对GB2312进行了扩展，用来显示罕见的、古汉语的汉字。现在已经收录了2.1万左右。并提供了1890个汉字码位。K的含义就是“扩展”。

Unicode编码（统一编码）：用4个字节（32位二进制）来表示一个字符，想法不错，但效率太低。例如，字母A用ASCII表示的话一个字节就够，可用Unicode编码的话，得用4个字节表示，造成了空间的极大浪费。A的Unicode编码是0000 0000 0000 0000 0000 0000 0100 0000

UTF-8（Unicode Transform Format）编码：根据字符的不同，选择其编码的长度。比如：一个字符A用1个字节表示，一个汉字用3-4个字节表示。

毫无疑问，开发中，都用UTF-8编码吧，准没错。

中文能够使用的字符集两种：

- 第一种：UTF-8。UTF-8是国际通用字库，里面涵盖了所有地球上所有人类的语言文字，比如阿拉伯文、汉语、鸟语.....
- 第二种：GBK（对GB2312进行了扩展）。gb2312 是国标，是中国的字库，里面仅涵盖了汉字和一些常用外文，比如日文片假名，和常见的符号。

字库规模： UTF-8（字很全） > gb2312（只有汉字）

重点1：避免乱码

我们用meta标签声明的当前这个html文档的字库，一定要和保存的文件编码类型一样，否则乱码（重点）。

拿 sublime 编辑器举例，当我们不设置的时候，sublime默认类型就是UTF-8。而一旦更改为gb2312的时候，就一定要记得设置一下sublime的保存类型： 文件→ set File Encoding to → Chinese Simplified(GBK)。VS Code 的道理一样。

重点2：UTF-8和gb2312的比较

保存大小：UTF-8（更臃肿、加载更慢） > gb2312（更小巧，加载更快）

总结：

- UTF-8：字多，有各种国家的语言，但是保存尺寸大，文件臃肿；
- gb2312：字少，只用中文和少数外语和符号，但是尺寸小，文件小巧。

列出2个使用情形：

- 1) 某公司是做日本动漫的，经常出现一些日语动漫的名字，网页要使用UTF-8。如果用gb2312将无法显示日语。
- 2) 某公司就是中文网页，极度的追求网页的显示速度，要使用gb2312。如果使用UTF-8将每个汉字多一个byte，所以5000个汉字，多5kb。

我们是怎么查看网页的编码方式的呢？在浏览器中打开网页，右键，选择“查看网页源代码”，找到meta标签中的charset属性即可。