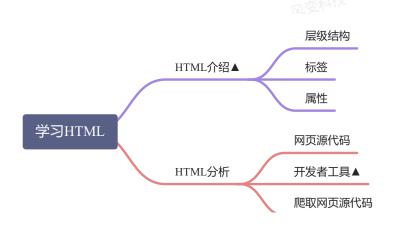
第1课 学习HTML

一、课程结构导图



注: ▲为重点知识点。

二、知识点讲解

2.1 HTML 介绍

概念: HTML是用来描述网页的一种语言,英文全称是Hyper Text Markup Language,也称为超文本标记语言。HTML 文档主要由HTML 元素组成。HTML 元素由标签以及标签中的内容构成。

示例: (<!-- 注释内容-->为HTML 注释)

```
<!-- 声明为HTML5文档-->
1 <!DOCTYPE html>
                                 <!-- HTML文档(起点) -->
2 <html>
3 <head>
                                 <!-- 文档头(起点)-->
        <meta charset="UTF-8">
                                <!-- 设置编码-->
        <title>风变一下,就能学到</title><!-- 设置网页标题-->
                                 <!-- 文档头(终点)-->
    </head>
                                 <!-- 文档体(起点)-->
    <body>
        风变一下
                                 <!-- <p>标签定义文字段落-->
                                 <!-- 文档体(终点)-->
    </body>
                                 <!-- HTML文档 (终点) -->
10 </html>
```

2.1.1 HTML 层级结构

代码解释:最外层是一个<html></html>标签,把所有的代码包含其中,在<head></head>标签中,又包含了两个标签<meta>和<title>。在<body></body>标签中,包含了标签。

2.1.2 标签

概念: 标签用于标记文本信息,指用尖括号(<>和</>)括起来的字母和英文,一般标签形式有两大类:闭合标签和自闭合签。闭合标签,常见于HTML代码中,一般成对出现(有开始标签<>,也有结束标签</>),如:<title>...</title>是标题标签,<div>...</div>是块标签。

自闭合标签,只有一个尖括号</>(斜杠/可省略),标签的内容都在尖括号中展示,比如:是图片标签,<input/>是input标签。

示例:

常见的HTML标签:

1120H210/775.	$B\mathcal{N}BMMMMMMMMMM$				
HTML常用标签					
标签	作用	标签	作用		
<html></html>	定义html文档	<h1>、<h2>、<h3></h3></h2></h1>	定义标题		
<head></head>	定义文档头部	>	定义段落		
<body></body>	定义文档主体		定义图片		
<a>	定义超链接		定义有序列表		
<audio></audio>	定义音频		定义无须列表		
<button></button>	定义按钮	<	定义单个列表条目		
<div></div>	定义块区域				

2.1.3 属性

概念:在HTML 文档中,属性存放在尖括号<>中,都是以赋值形式存在,即:属性=属性值。常见的属性有:class、id、href、style。

风变科技

双变科技

常见HTML属性与用法			
属性	用法		
class	为html元素定义一个或多个类名(classname)		
id	定义元素的唯一id		
href	用来定义链接		
style	规定元素的行内样式(inline style)		
	—————————————————————————————————————		

示例:

```
1 <!DOCTYPE html>
2 <head>
      <meta charset="UTF-8"> <!-- charset是一个属性 -->
4 </head>
5 <style>
      /*定义了class属性为style_1的样式*/
      .style_1 { width: 100px; }
      /*定义了class属性为style 2的样式*/
       .style_2 { width: 50px;
10 </style>
11 <body style="background:#a8c7e2"> <!-- style是一个属性 -->
       <img class='style_1' src="https://res.pandateacher.com/风变科技logo1.png" ><!-- class是</pre>
12
13
      <img class='style_2' src="https://res.pandateacher.com/forchange_blue_logo.png" >
      <img class='style_2' src="https://res.pandateacher.com/forchange_black_logo.png" >
14
15
      <br>
      <img class='style_1' src="https://res.pandateacher.com/风变科技logo2.png" >
       <img class='style_2' src="https://res.pandateacher.com/panda_black_logo.png" >
17
      <img class='style_2' src="https://res.pandateacher.com/panda_white_logo.png" >
18
19 </body>
20 </html>
```

注意:为了方便反复调用,部分内容的尺寸、颜色等设置会放到<head>标签中的<style>标签里。在<style>标签中,class属性的样式用点'.',id属性用并号 '#'。

2.2 HTML分析

我们可以通过两种方式查看网页源代码(注:所有演示都是在谷歌浏览器下进行。),一种是直接打开网页源代码查看,另一种是打开开发者工具进行查看。

注意两者的一个区别,当网页是静态网页的时候,二者基本相同;当网页是动态网页时,开发者工具中的代码是浏览器渲染之后显示的内容,包括网页源代码,还有藏在Network 面板中的一些数据(第4~5课会介绍动态网页)。

2.2.1 网页源代码

1. 随机打开一个网站,在网页任意地方点击鼠标右键,然后点击"查看网页源代码"。



2. 点击"查看网页源代码"之后,会弹出一个新窗口(如下图),该代码便是网页源代码。

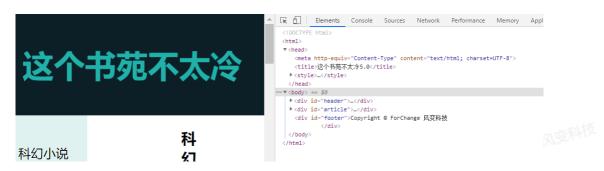
```
→ C • view-source:https://localprod.pandateacher.com/python-manuscript/crawler-html/spider-men5.0.html
    <html>
        <head>
             <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
             <title>这个书苑不太冷5.0</title>
                 text-decoration: none;
            }
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
            body {
                 margin: 0;
                 width:100%;
                 height: 100%;
                 background-color:#0c1f27;
color:#20b2aa;
                 text-align:center;
                 padding:15px;
                 line-height:60px:
```

2.2.2 开发者工具

1. 随机打开一个网页,右键点击【检查】选项,Windows电脑的同学可直接按F12或Fn+F12。



2. 随即, 网页下方(或右方)弹出一个子窗口, 这便是浏览器的开发者工具。



HTML的层级关系可通过这些小三角形进行查阅,每一个可以展开和合上的小三角形里包含的内容,都是一个层级,就像电脑中一层一层的文件夹。

2.2.3 爬取网页源代码

示例:

```
1 import requests # 调用requests 库
2 res = requests.get('https://localprod.pandateacher.com/python-manuscript/crawler-html/spider-men5.0.html') # 获取
3 code = res.text # 将返回的Response 对象转化为字符串格式
4 print(code) # 打印Response 对象转化的字符串格式
```

注意:使用requests.get() 爬取的是网页源代码的内容,当网页是动态网页时,爬取到的代码和开发者工具中的代码是不一样的。