

# 互联网基本原理和HTML入门

第1天课堂笔记（本课程共2天）

班级：北京前端训练营1期

讲师：邵山欢

日期：2016年3月21日

邵山欢老师

微博：@邵山欢

QQ : 179427026

E-mail : shaoshanhuan@163.com

爱前端官网 ：[www.iqianduan.cn](http://www.iqianduan.cn)

# 目录

[互联网基本原理和HTML入门 1](#_Toc8522)

[目录 2](#_Toc18516)

[一、从“上网”说开去 3](#_Toc20374)

[1.1 上网就是请求数据 3](#_Toc30303)

[1.2 HTTP协议 5](#_Toc30536)

[1.3 服务器 6](#_Toc21483)

[1.4 浏览器 6](#_Toc8794)

[二、HTML初步 7](#_Toc17081)

[2.1 纯文本 7](#_Toc21637)

[2.2 HTML是负责描述文本语义的语言 10](#_Toc17556)

[三、Sublime安装和使用方法 13](#_Toc18817)

[3.1 安装 13](#_Toc11040)

[3.2 使用 14](#_Toc9725)

[四、HTML的基本骨架 16](#_Toc17256)

[4.1 文档声明头DTD 17](#_Toc14734)

[4.2 命名空间 19](#_Toc17050)

[4.3 字符集 19](#_Toc8427)

[4.4 关键字和页面描述 21](#_Toc9674)

[4.5 title 22](#_Toc21291)

[五、HTML页面的特点 23](#_Toc1211)

[5.1 HTML对换行、TAB缩进、空格不敏感 23](#_Toc14824)

[5.2 空白折叠现象 24](#_Toc6136)

[六、h和p 25](#_Toc13165)

[6.1 h系列标签 25](#_Toc29904)

[6.2 p标签 26](#_Toc5974)

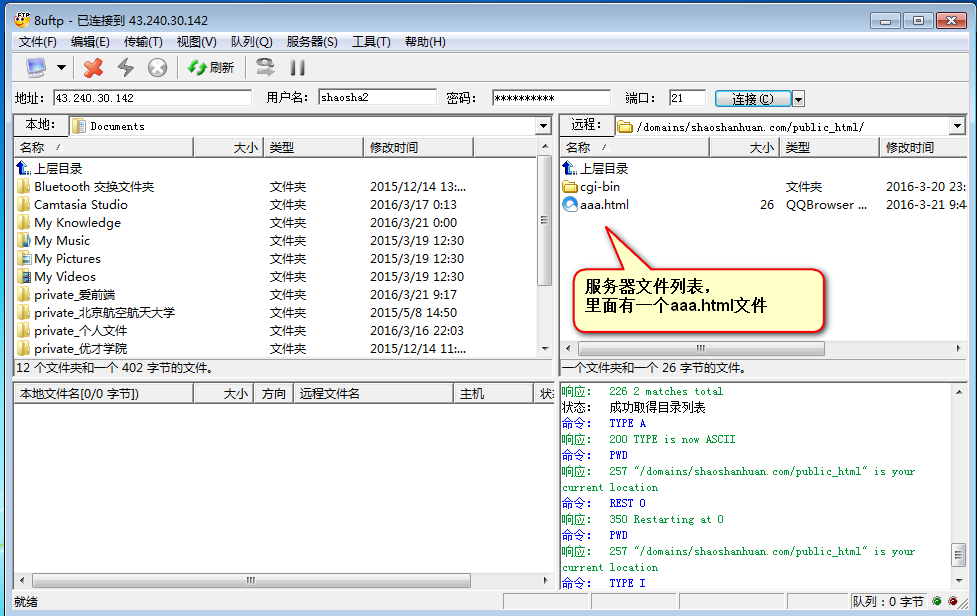
[七、作业 27](#_Toc21985)

# 一、从“上网”说开去

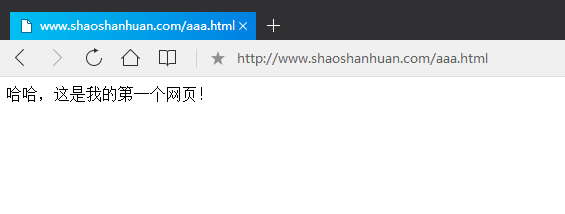
## 1.1 上网就是请求数据

老师现在进行一个互联网原理的**整体感知**的教学，你注意，整体感知中，一些具体的细节，可能令你没有安全感（很多的知识盲点）。没有关系，老师进行的是“整体感知”，就是为了让你宏观地学习一些东西，先不要在意细节。

网页是真实的物理存在，我们现在来看一下老师购买的一个服务器的操作。



所以我们就能通过网址www.shaoshanhuan.com**/aaa.html**来查看这个页面。



我们来看看文件夹的情况：



www.shaoshanhuan.com/ccc/ddd.html

小问题：我们想访问www.shaoshanhuan.com**/haha/xixi/hehe.html**

那么请问，我们的网页应该放在哪里？

**答案：放在haha文件夹中的xixi文件夹中里面**。

更深入的问题：我们平时输入网址，都是www.shaoshanhuan.com ， 没有后缀

实际上等价于访问了www.shaoshanhuan.com**/index.html**文件。

index是英语目录、列表的意思，这是默认的首页文件。

文件夹的默认文件，也是index.html

www.shaoshanhuan.com/aaa

等价于

www.shaoshanhuan.com/aaa/index.html

总结一下，一会儿我们马上展开研究：

也就是说，上网就是一个请求数据的过程。当我们输入一个网址www.shaoshanhuan.com/aaa/1.html的时候，我们的计算机将对远程服务器发出一个HTTP请求。我要请求你的aaa文件夹中的1.html文件。服务器响应了这个请求，将1.html这个网页文件，通过HTTP请求，传输到用户的计算机中。用户的浏览器，对这个文件进行渲染。

上网就是一个请求文件、回馈文件的过程，没有持久连接。

## 1.2 HTTP协议

HTTP大家肯定不陌生，一说网址，装X的人就爱说：

**http://**www.shaoshanhuan.com

HTTP的全称叫做**H**yper**t**ext **T**ransfer **Pr**otocal，超文本传输协议。干嘛用的？就是浏览器和服务器之间传输文件用的。（建议大家背诵简称的全名，面试的时候吓死面试官！）

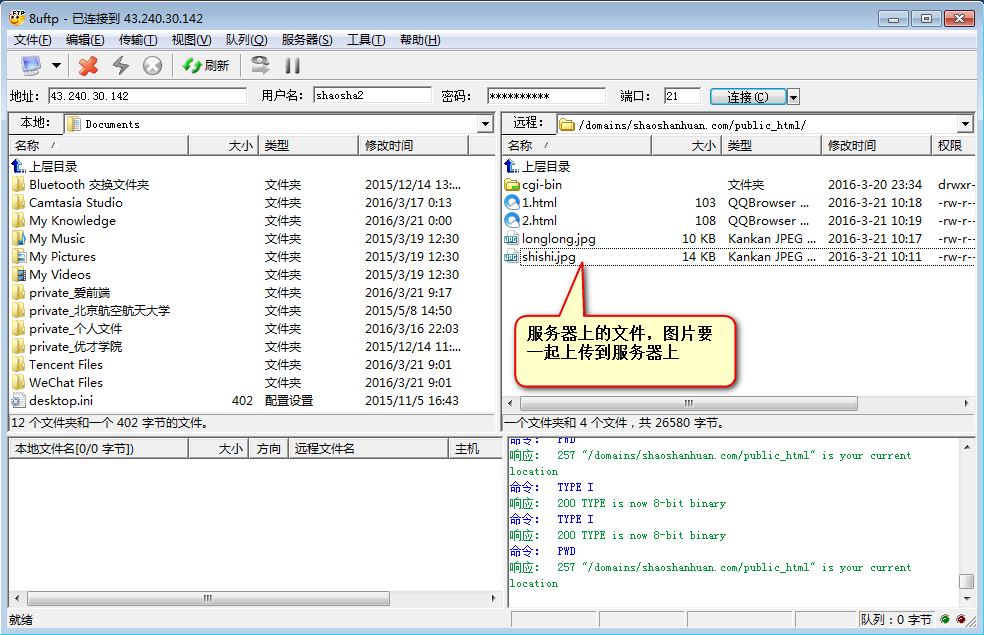
HTTP是有两部分，请求request、响应response。当你输入网址的时候，此时浏览器会发出一个HTTP请求，请求服务器上的响应页面。服务器收到请求之后，会再次通过HTTP将页面传输回来。

注意：

① 访问一个网页的时候，并不一定只有一次HTTP请求发出！如果页面上有图片、视频、音频，那么将产生更多的HTTP请求。

② 不光输入网址能够产生HTTP请求，我们点击超级链接的时候，也能够产生HTTP请求。

Ajax课程上我们拓展学习HTTP，现在先学这么多。



## 1.3 服务器

服务器server就是计算机，也有CPU、硬盘、内存，也安装软件。可能没有显示器、鼠标、键盘（用普通计算机远程管理它）。



服务器上存放文件的，服务器要24小时开机，不能断点，一旦断点，就无法访问网站了。

我们可以通过管理软件，来远程管理服务器，比如老师演示的8U-FTP。

我们这里承诺一下：4月4日前，我们找一天晚自习，给大家手把手教学域名的购买、服务器的租用。

在Ajax课程上，我们深入研究服务器，现在就学这些。

## 1.4 浏览器

浏览器browser，就是渲染网页的一个软件，安装在客户的电脑里面。

**HTTP请求的发起、接收，都是由浏览器来完成的。**

粗略的介绍一下浏览器的品牌（CSS课程上，我们将深入学习各个浏览器的差别）：



浏览器有版本的差别，比如IE有6、7、8、9、10，我们在后面的课程将学习浏览器兼容的知识。

浏览器都有临时文件夹，比如IE的临时文件夹就是：

C:\Users\**Danny**\AppData\Local\Microsoft\Windows\Temporary Internet Files

你请求的所有网页文件，都是存放在这个临时文件夹中的。所有的网页都是在计算机本地进行渲染的。所以，有些时候，第一次打开网页速度慢，第二次打开网页速度快，这是因为第一次已经把文件传输过来了，就不用传输了。

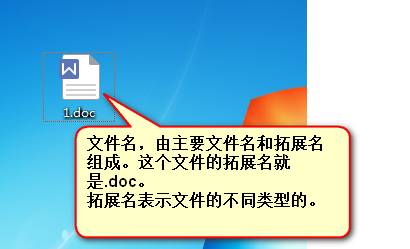
# 二、HTML初步

HTML的全称叫做Hypertext Markup Language，超文本标记语言，是网页的文件格式。网页的格式，还有PHP、JSP、ASP等等，那么HTML是最最基本的网页文件格式。至于他们的区别，日后的课程会讲解。

## 2.1 纯文本

我们现在先来学习什么是“纯文本”。

拓展名的含义：



不同的拓展名： .mp3音乐文件， .jpg 图片文件 ， .docx 文档

要让我们的操作系统，显示出文件的拓展名。

|  |  |
| --- | --- |
|  |  |

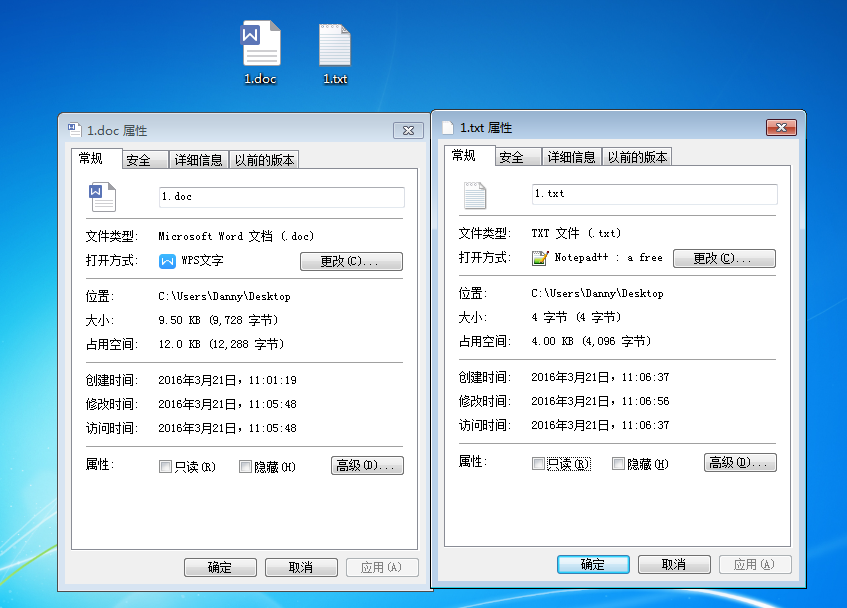
win8、win10没有这么复杂，自己琢磨琢磨吧。

我们来看一个实验：

1.doc和1.txt中， 都是只有“哈哈”两个字。但是doc文件，尺寸很大；而txt文件，尺寸只有4字节。

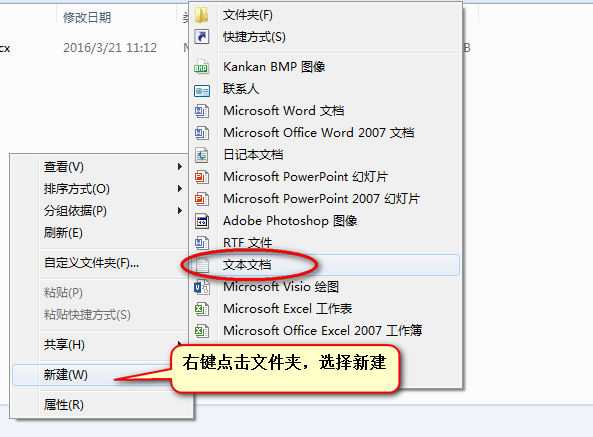
这是因为txt文件是纯文本文件，里面只有文字，没有其他的任何东西，没有样式、没有字号、没有颜色。

而doc文件，里面存放着页边距、行高、颜色、字号、字体这些信息。



**HTML文件是纯文本文件。**

我们这么创建一个网页：



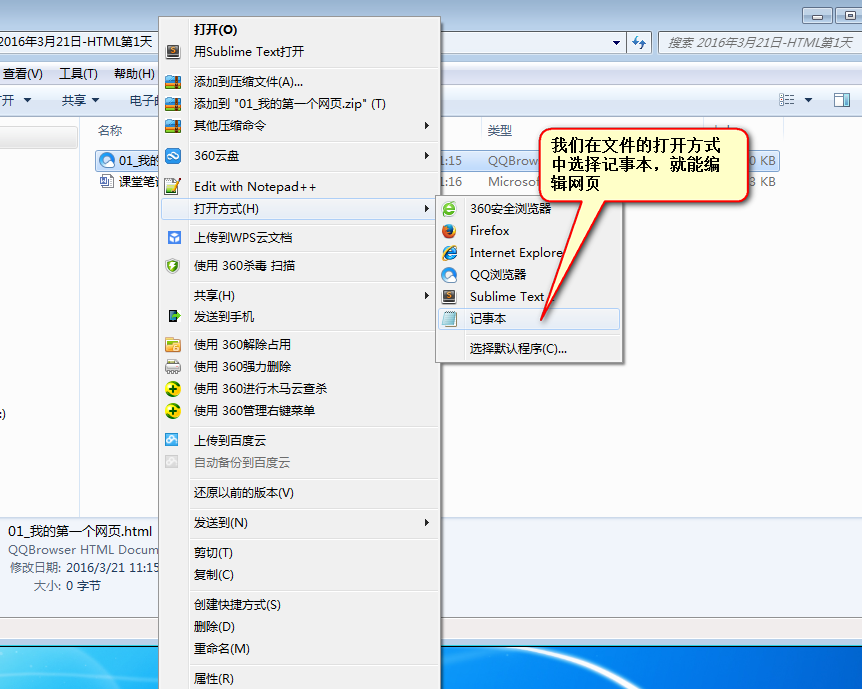
创建出来一个txt文件：



强行更改txt这个拓展名为html：



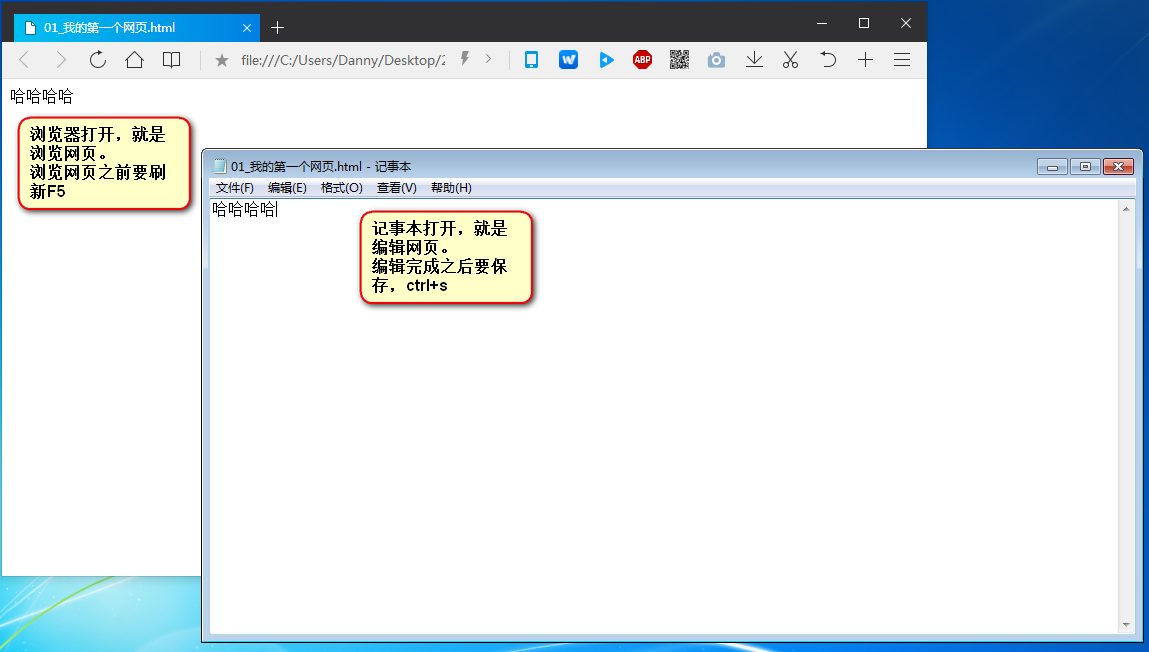
这个html文件就是一个网页文件了。



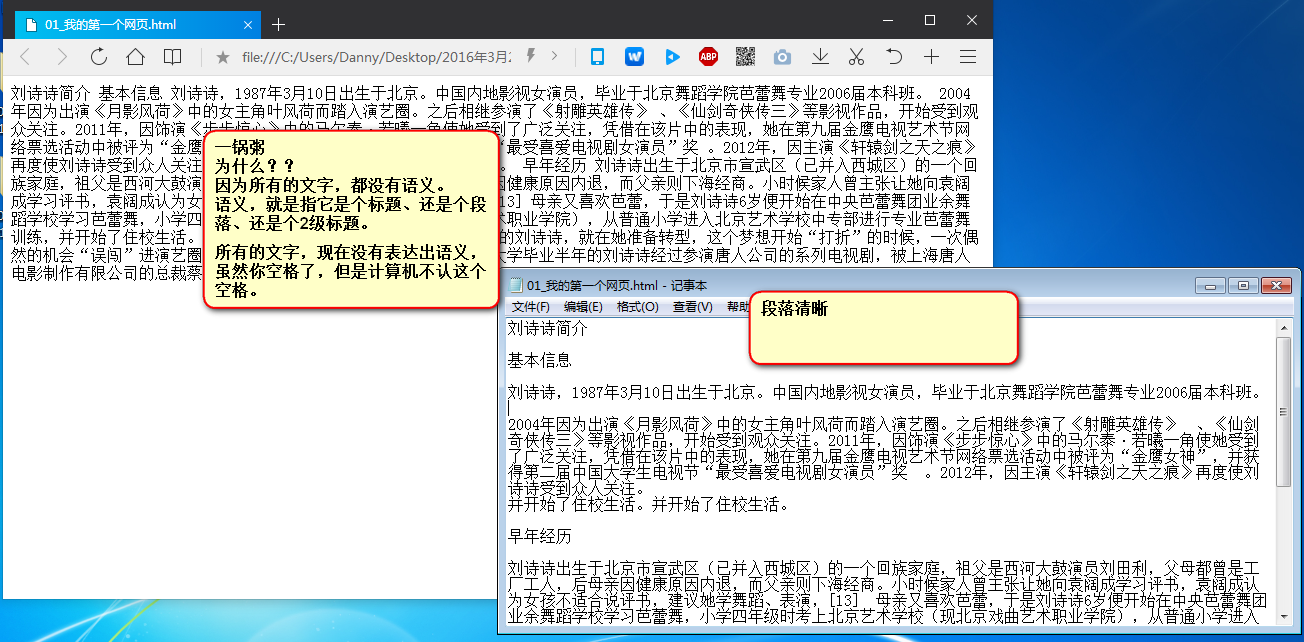
不要给我提什么DreamWeaver了！因为HTML文件是纯文本文件，所以任何的纯文本编辑器，都能够编写网页！

HTML的制作，不依赖于任何编辑器。

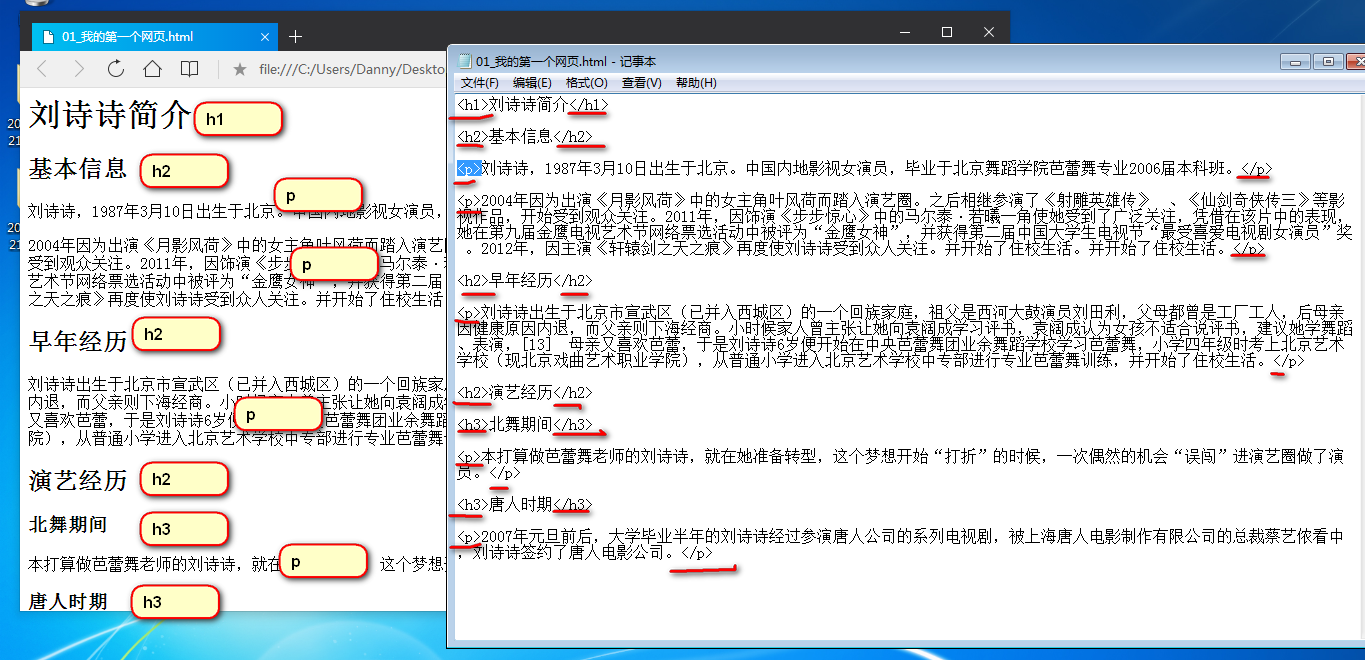
网页的制作和浏览，分别使用记事本和浏览器：



## 2.2 HTML是负责描述文本语义的语言



我们用标签对儿来添加语义：



比如下面的文字：

|  |
| --- |
| 1. 刘诗诗简介 |

它是**没有任何语义**的文字。

而：

|  |
| --- |
| 1. **<h1>**刘诗诗简介**</h1>** |

给这几个字添加了**“1级标题”**的语义。

比如下面的文字，

|  |
| --- |
| 1. 刘诗诗，1987年3月10日出生于北京。中国内地影视女演员，毕业于北京舞蹈学院芭蕾舞专业2006届本科班。 |

也是**没有任何语义**的文字。

而：

|  |
| --- |
| 1. **<p>**刘诗诗，1987年3月10日出生于北京。中国内地影视女演员，毕业于北京舞蹈学院芭蕾舞专业2006届本科班。**</p>** |

给这些文字增加了**“段落”**的语义。

HTML不过如此，就是通过一对儿一对儿的标签，来给文本增加语义的语言。

<p> 叫做起始标签

</p> 叫做结束标签

页面渲染的时候，<h1>这些标签，是不会显示在页面上的，这就是“超”字的含义：一些文本，就是文本；而一些文本（标签），**是描述别的文本语义的文本，不会照搬显示在页面上**。 这种文件就是超文本文件。

也解释了什么是“标记”，就是一对儿一对儿的标签对儿。

超文本标记语言HTML： **H**yper**t**ext **M**arkup **L**anguage。

你加上h1标签之后，文本就是主标题的语义，浏览器此时会默认将h1的文字变得大、黑、粗，这个是浏览器的默认样式，和语义是无关的。浏览器会默认给一些语义的文字加上合适的样式。

但是，一定要记住,任何HTML标签，都和样式无关的，只能表达语义。

面试的时候，人家问h1什么意思啊？

**~~错误的答案：让字变得大、黑、粗。~~**

正确的答案：让文字加上1级标题的语义。

# 三、Sublime安装和使用方法

## 3.1 安装

再次强调，做网页和用什么软件无关。任何的纯文本编辑器，都能够制作网页；相反的，任何的能够制作网页的软件，吹破天了，本质上都是纯文本编辑器。

现在比较有名的制作网页的软件：sublime、WebStorm、Atom、DreamWeaver、NotePad++、Editplus、frontpage。

sublime是英语宏大的、壮丽的意思，冷门词汇。

sublime这个编辑器近些年非常流行，能够编辑多种语言，Java、PHP、C语言都能编写。

官网http://www.sublimetext.com/

这个编辑器非常好用：

1） 能够高亮显示代码，比如自动识别我们的标签，用不同颜色表示。



2） 自动感应

3） 有非常丰富的插件，并且可以自己写插件，Python实现的。

问题是，sublime的下载安装，极其简单，但是默认的下载好的纯净版，需要安装很多插件、汉化、破解才能使用。现在老师为了方便大家，提供一个傻瓜版，就是老师已经继承了很多插件、汉化、破解了给你使用。当你的能力足够牛逼的时候，我们教你自己配置纯净版的sublime。

傻瓜版本的安装方法：

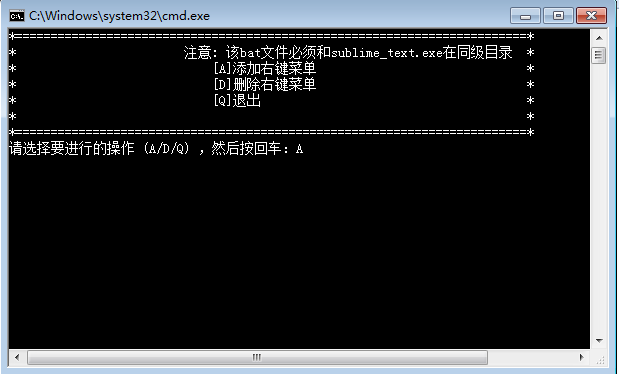


解压缩，就能用。所以，你别瞎解压，放到你的存放软件的文件夹里面！！

如果想要方便的把文件用sublime打开，那么用：

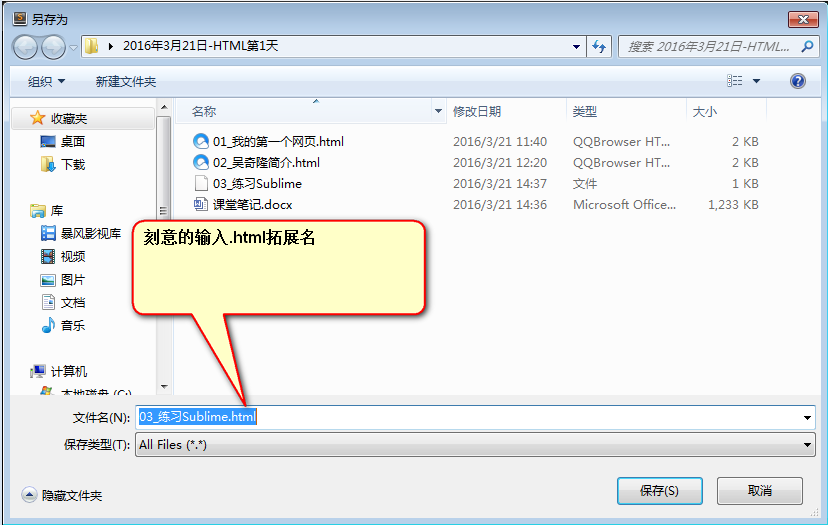


**输入A**：



## 3.2 使用

sublime能够编辑很多很多语言！所以，在新建了一个文件之后，一定要记得先保存，保存的时候，要刻意地加上.html后缀。



Sublime的常见快捷键：

Ctrl+N 新建文件

Ctrl+S 保存功能。这里提一嘴，Sublime不怕突然掉电。它不是实时保存，而是实时缓存。



Ctrl+F 查找

Ctrl+Z 撤销。Sublime提供无限次的撤销。

Ctrl+C 复制

Ctrl+V 粘贴

输入标签名，然后按tab键，能自动生成标签对儿。

Ctrl+Shift+D 复制当前行

Ctrl+X 删除当前行（和剪切是一个快捷键）

Ctrl+鼠标滚轮

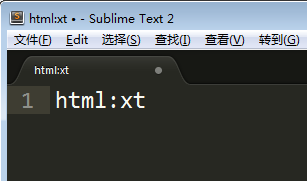
按住鼠标滚轮，拖拽，就可以产生多行光标。

支持快速输入法，比如

p\*10 然后按tab键，就能生成10个p标签。

# 四、HTML的基本骨架

直接在sublime中，输入：



然后按tab键，如果tab键不管用，那么就按ctrl+E键，就会自动展开为：

|  |
| --- |
| 1. <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"> 2. <html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en"> 3. <head> 4. <meta http-equiv="Content-Type" content="text/html;charset=UTF-8"> 5. <title>HTML骨架演示</title> 6. </head> 7. <body> 9. </body> 10. </html> |



抽象一下：

|  |
| --- |
| 1. <!DTD > 2. <html> 3. <head> 5. </head> 6. <body> 8. </body> 9. </html> |

接下来我们就要一行一行的讲解HTML基本骨架的含义。

## 4.1 文档声明头DTD

就是HTML第一行语句：

|  |
| --- |
| 1. <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"> |

所谓的DTD就是**D**oc**T**ype **D**efinition 文档类型定义, 也有人翻译为**D**oc**T**ype **D**eclartion 文档类型声明

HTML文件的第一行，都是这个DTD，干什么用的？这一行语句，**告诉浏览器我是什么版本的HTML文件**。

HTML、CSS这两个语言，官网维护者就是总部位于美国的**W3C组织**，简称W3。全称World Wide Web Consortium国际万维网联盟。官网：https://www.w3.org/

中文的网站就是w3scholl.com.cn，

里面的各种DTD手册：<http://www.w3school.com.cn/tags/tag_doctype.asp>

HTML从诞生到现在，最新的版本HTML5。不是所有的浏览器都兼容哦，现在还有一部分人使用的浏览器过旧，所以现在很多公司还在使用HTML的上一个版本，HTML4.01。

HTML4.01又有三个小版本：

HTML4.01 Strict 严格版：不能使用font、b、u、i等等的废弃标签，不能使用框架集，结构和样式分离。

HTML4.01 Transitional 过渡版（通用版）：没有那么多限制，可以使用font等废弃标签，不能使用框架集

HTML4.01 Frameset：框架集版：可以使用框架集

讲一下历史，在2007、08年之前，人们制作一个红色的标题：

|  |
| --- |
| 1. <h1><font color="red">我是一个主标题</font></h1> |

现在font标签已经被废弃了，改用CSS来描述页面的样式。

W3C事儿多，发现HTML还不够严格。比如，标签到底是大写字母还是小写字母呀？

|  |
| --- |
| 1. <H1></H1> |

是对的么？

再比如，属性可不可以用单引号？

|  |
| --- |
| 1. <a href=’2.html’></a> |

是对的么？

**所以，W3C为了解决这个问题，推出了XHTML版本。X表示extensional“拓展的”HTML。**

**在XHTML1.0中，严格规定了标签必须是小写，所有的属性都必须用双引号封闭啊，必须有结尾反斜杠……**

XHTML1.0版本中，延续了HTML4.01的3个小版本：

XHTML1.0 Strict 严格版：不能使用font等等的废弃标签，不能使用框架集，结构和样式分离。

XHTML1.0 Transitional 过渡版（通用版）：没有那么多限制，可以使用font等废弃标签，不能使用框架集

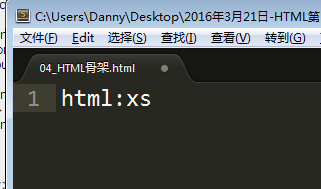
XHTML1.0 Frameset：框架集版：可以使用框架集

总结一下，一共有6种DTD。HTML4.01有3种，XHTML1.0有3种。6种版本，就有6种DTD。

按严格程度：

XHTML1.0 Strict > HTML4.01 Strict > XHTML1.0 transitionl > HTML4.01 transitionl

Sublime生成DTD，只需要更改冒号后面的词语，xs就能生成XHTML1.0 Strict版本。



严格版本中，font、b、u、i标签都是废弃的。**而这些标签我们将来要使用它做一些“小心思”、“小创意”，所以不能一棒子打死。**

如果没听懂，别废话，老老实实用html:xt就行了。只要记住，一共有6种。

HTML5是新的版本，不在设立strict、transitional、frameset等版本了。XHTML1.0随着HTML5的诞生，也没有了。这就是W3C，在啪啪啪自己打脸。

## 4.2 命名空间

在DTD下面一行，就是html标签对儿：

|  |
| --- |
| 1. <html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en"> 2. </html> |

html是一个标签，xmlns是它的一个属性， xml:lang也是它的一个属性

|  |
| --- |
| 1. <html 属性1=”值” 属性2=”值”> 2. </html> |

键值对儿，键就是属性的名字，值就是属性的值。简称K-V对儿， k就是key键， v就是value值的意思。

xmlns这个属性是干嘛的呢？xmlns的全称，叫做XML NameSpace， HTML是XML中的一种，但是我们现在不准备展开说这事儿。Ajax课上会详细介绍XML这个东西。NameSpace叫做命名空间。很简单，就是你的文章里面的h1表示的是1级标题的，别人的是不是也是1级标题？所以就需要有人统一一下标准，使用的就是一个固定的网址http://www.w3.org/1999/xhtml

xml:lang="en" 表示所有的标签的语言都是英语，language

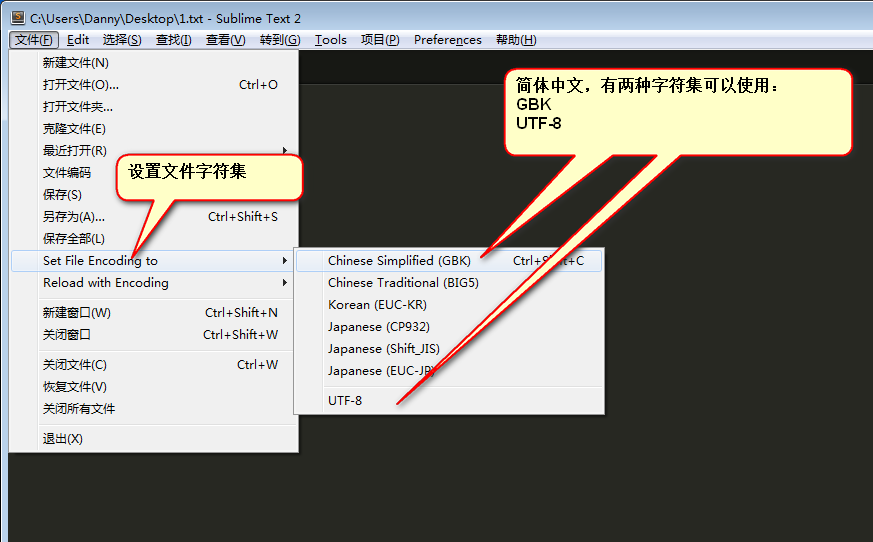
这里都是固定写法，没有别的值。

## 4.3 字符集

简体中文可以使用的字符集就两个，分别叫做GBK、UTF-8。

GBK是国标的一个意思，也称为“gb2312”，中国人制定的。

UTF-8是国际的一个字符集标准。



我们做了一个实验，都是同一个汉字，但是：

GBK 每一个汉字2字节

UTF-8 每一个汉字3字节



字符集就相当于活字印刷术里面的字的模型。

老王家的字多，所以描述“哈”字，可能是“第3个柜子里面第2层第4行第8列”

老李家的字少，所以描述“哈”字，可能是“第9层第3行第1列”

GBK中只有汉字简体（不包括少数民族文字）、绝大部分繁体字、少量片甲名、少量符号

UTF-8中有世界上所有国家中的所有文字和符号。

我们在HTML网页中，用meta来设置字符集，meta是“元”、“初始”的意思，表示元设置、初始设置。

|  |
| --- |
| 1. <meta http-equiv="Content-Type" content="text/html;charset=UTF-8"> |

这个meta是一个单标签，没有成对儿。一会儿在img标签中，我们细讲这事儿。

Sublime将默认保存utf-8字符集。

所以，HTML页面中，有两种字符集设置：

|  |
| --- |
| 1. <meta http-equiv="Content-Type" content="text/html;charset=**UTF-8**"> |

和

|  |
| --- |
| 1. <meta http-equiv="Content-Type" content="text/html;charset=**gb2312**"> |

不用背，只需要记忆红色部分。

工作的时候用哪种？

1） 如果公司的网页中，没有出现怪异符号、没有出现其他国家文字（除了英语），那么要使用gb2312。因为文件尺寸小，每一个字就少1个字节。1000字就是1kb，5000字就是5kb。

2） 如果公司的网页中，有少数民族文字、韩语、阿拉伯语，那么要使用UTF-8。因为文字全。

|  |  |  |
| --- | --- | --- |
|  | 文字全不全 | 汉字几字节 |
| UTF-8 | 全 | 3 |
| gb2312 | 不全 | 2 |

UTF-8

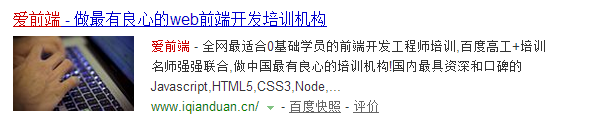


点破了一道面试题：怎么样让网页打开速度变快？

我们现在知道了方法1：就是让网页的字符集变为gb2312，能够显著降低页面尺寸。

方法2、方法3、方法4……以后再说。

## 4.4 关键字和页面描述



浏览器抓取我们的页面，会在旁边显示一小段页面描述。

|  |
| --- |
| 1. <meta name="description" content="爱前端-全网最适合0基础学员的前端开发工程师培训，百度高工+培训名师强强联合，做中国最有良心的培训机构！国内最具资深和口碑的Javascript,HTML5,CSS3,Node,React,Redux,Angular,Reflux,flux,vue,backbone,工程化,设计模式,性能优化,模块化,ES6,jQuery培训，帮助更多学员实现IT梦想，找到高薪工作的培训机构，推出《JavaScript设计模式》一书，《爱前端视频教程》等被大家誉为最经典的web前端视频教程"> |

抽象出来语法：

|  |
| --- |
| 1. <meta name=”description” content=”页面描述” > |

description 描述的意思， content内容。

这个东西显著提升SEO， SEO就是Search Engine Optimization，搜索引擎优化。就是少花钱，让搜索引擎能带来更多的流量，让更多人点击。

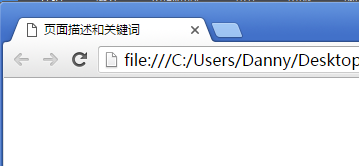
可以设置页面关键字，搜索引擎会抓取这些关键字：

|  |
| --- |
| 1. <meta name="keywords" content="手机,电脑,冰箱,彩电"> |

搜索引擎就知道我们的网站是干什么的了，就能相应的排名提升。

## 4.5 title

|  |
| --- |
| 1. <title>页面描述和关键词</title> |



# 五、HTML页面的特点

## 5.1 HTML对换行、TAB缩进、空格不敏感

|  |
| --- |
| 1. <body> 2. <h1>我是一个主标题</h1> 3. <p>我是一个段落</p> 4. </body> |

等价于：

|  |
| --- |
| 1. <body> 2. <h1>我是一个主标题</h1> 3. <p>我是一个段落</p> 4. </body> |

等价于：

|  |
| --- |
| 1. <body><h1>我是一个主标题</h1><p>我是一个段落</p></body> |

也就是说，HTML现在你就感觉到了，不仅仅负责描述语义，还描述了页面的层次。

层次是什么，就是看标签谁包裹谁，和缩不缩进无关！

下面的缩进版本的HTML代码，非常清楚，一下子就能够看出来谁包裹谁：

|  |
| --- |
| 1. <body> 2. <h1>我是一个主标题</h1> 3. <p>我是一个段落</p> 4. <div> 5. <div> 6. <div> 7. <div> 9. </div> 10. </div> 11. </div> 12. </div> 13. </body> |

不缩进版本，可读性不强，不许这么写代码：

|  |
| --- |
| 1. <body> 2. <h1>我是一个主标题</h1> 3. <p>我是一个段落</p> 4. <div> 5. <div> 6. <div> 7. <div> 9. </div> 10. </div> 11. </div> 12. </div> 13. </body> |

将页面制作完毕之后，会把代码进行压缩。容铭老师会给大家讲解工程化的技术的时候，有压缩技术：



所以这里的著名面试题，有了新的答案：“如何提升页面加载速度？” “压缩代码，将HTML压缩为1行”

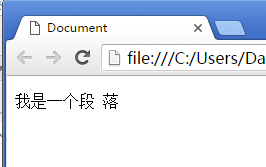
## 5.2 空白折叠现象

HTML中的文字，不管有多少个空格，多少个换行，多少个缩进，都会被压缩为一个空格。

文字，现在有很多的换行、空格、缩进：

|  |
| --- |
| 1. <p> 2. 我是一个段 3. 落 4. </p> |

折叠为了一个空格：



# 六、h和p

## 6.1 h系列标签

h表示header，标题的意思。一共有六个，有不同的语义。

h1 1级标题

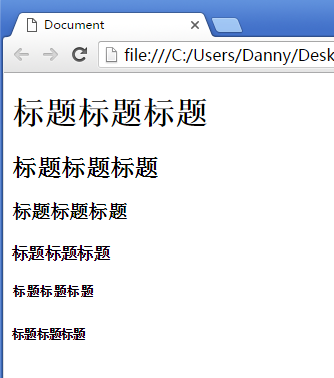
h2 2级标题

h3 3级标题

h4 4级标题

h5 5级标题

h6 6级标题



一般来说，页面上只能有一个h1标签，其他标签个数不定。实际上这个不是W3C的规则，而是搜索引擎的规则。

h1标签中的内容，权重非常高，搜索引擎会特别注意抓取里面的文字。搜索引擎如果看见页面有多个h1，视为作弊，降低你的权重。

h标签是个文本级标签。

## 6.2 p标签

p表示paragraph，段落。

|  |
| --- |
| <p>文字</p> |

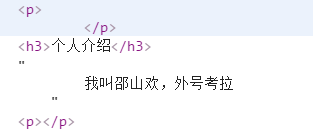
这个就是一个段落。

p标签你一定要注意一件事儿，p里面只能放文字、图片、表单元素。不能放其他东西。p是一个文本级标签。

错误的写法：

|  |
| --- |
| 1. <p> 2. <h3>个人介绍</h3> 3. 我叫邵山欢，外号考拉 4. </p> |

浏览器会这么解析：



正确的写法：

|  |
| --- |
| 1. <h3>个人介绍</h3> 2. <p> 3. 我叫邵山欢，外号考拉 4. </p> 5. <p> 6. 我喜欢HTML、CSS 7. </p> |

这里实际上揭示了一个事情，就是标签的级别。我们HTML将所有的标签都进行了分类，只有两类：

1） 容器级标签 ： 里面谁都可以装，甚至可以包裹和自己一样的标签。

2） 文本级标签 ： 里面只能放文字、图片、表单元素，和其他的文本级标签。

# 七、作业

快速讲一下img和a，明天会细讲

img是图片。

|  |
| --- |
| <img src=”路径” /> |

img是英语image图片的缩写，src是source资源的缩写。

这个标签不是一个对儿。

|  |
| --- |
| 1. <p>文字</p> |

p是给别人增加语义的，需要包裹文字。而img自己就是图片语义，不需要包裹谁。所以就是自封闭标签，也叫作单标签。

|  |
| --- |
| 1. <img src=”” **/>** |

明天介绍相对路径、绝对路径。

预习一下超级链接

a是anchor锚的意思

|  |
| --- |
| 1. <a href=”页面地址.html”>链接文字</a> |

自己折腾HTML标签，用img、a标签做一个小网站。所有的样式都不要有，裸奔。