# Xpath选择器

## 用途：

不仅可以用来在HTML文档中选择节点,还可以在XML文档中选择节点（手机自动化中会用到）

## 特点

语法非常像Linux系统中的路径，所以又叫做路径表达式

## 语法结构

### 绝对路径表示法：

从根节点（html）开始，路径分割符是/（斜杠，类似于linux中路径分割符）

如：/html/body/div/span

表示对应路径下的span元素

### 相对路径表示法：

以//开头后面加元素名称, 用法类似于css中的后代选择器，可以把//对应css中空格

如：//span

表示html文档下所有的span元素

如：//div//span

表示方法类似于CSS中的 div span

### 绝对路径与先对路径混合使用：

如： //div/p/span

p和span的位置确定的，分别属于div元素的子元素与孙元素

## 通配符 \*

与css中一致，星号（\*），表示匹配所有元素

用法：//div/\* 表示匹配div下面所有元素

等价于css中的 div>\*

## 属性选择器

### 使用场景

选择具有某个属性（值）的元素

### 表示方法

//\*[@属性] 表示选择具有某个属性的元素

//\*[@属性=”属性值“] 表示选择具有某个属性值的元素，属性值必须加引号（单引号双引号都可以）

示例：

//\*[@style] 选择所有具有style属性的元素

//p[@spec='len2'] 选择所有具有spec 值为 len2 的元素

Tips:在xpath中没有表示id和class的特殊方法,id 、class 也是属性

如：

//div[@id="food"]

//div[@class="cheese"]

根据部分属性选择元素的方法有

//\*[contains(@属性,”value“)]

表示属性包含value

//\*[starts-with(@属性,”value“)]

表示属性以value开头

## 子元素选择-单选

### 使用场景

选择属于其父元素的第n个某个 类型的子元素

### 表示方法

//\*[n] 表示选择\*的父元素下面第n个\*类型元素 等价于 \*:nth-of-type(n)

或//\*[position()=n]

示例：

//p[2], 选择p的父元素下的第二个p元素，等价于 p:nth-of-type(2)

### 倒数下标表示方法

#### 使用场景：

选择属于其父元素的倒数第n个某个 类型的子元素

#### 表示方法：

//\*[last()-n] 表示选择\*的父元素下面倒数第n+1个\*类型元素 等价于 \*:nth-last-of-type(n)

或//\*[position()=last()-n]

#### 示例：

//span[last()-1] 属于其父元素的倒数第二个span 等价于 span:nth-last-of-type(1)

//\*[@id="food"]/span[last()] #food元素下倒数第1个span

//\*[last()-1] 父元素下的倒数第二个元素（不限定类型）

## 子元素选择-多选

### 使用场景

选择属于其父元素的第m到n之间 某个类型的子元素

### 表示方法

//\*[position()>1] 选择属于其父元素的第一个以后的\*元素

//\*[position()<last()-1] 选择属于其父元素的倒数第二之前的\*元素

示例：

//\*[@id="food"]/\*[position() > last()-3] 选择#food下倒数第四个到最后一个元素

## 组选择

### 使用场景：

用于多组xpath表达式组合来选择元素的情况

### 表示方法

<s1>|<s2>

<s1>和<s2> 是两组xpath选择器

示例：

//p | //button 选择所有的p元素, button元素，等价于 p, button

//p | //span | //div 选择所有的p元素, span元素和div元素，等价于 p, span,div

Tips: css和xpath的表达式不能混用

## 相邻兄弟选择

### 使用场景

根据同级元素选择其他的同级元素

### 表示方法

//\*/following-sibling::\* 选择\*后面的兄弟元素 等价于 \*~\*

//\*/preceding-sibling::\* 选择\*前面的兄弟元素

示例：

//\*[@id='food']/span/following-sibling::p 选择span元素后面的兄弟元素p

//\*[@id='food']/p/preceding-sibling::span 选择p元素前面的兄弟元素span

### 扩展

//\*[@id='food']/span/following-sibling::p[1]

span元素后面的第一个兄弟元素p

//\*[@id='food']/span/following-sibling::p[last()]

span元素后面的最远一个兄弟元素P

//\*[@id='food']/span/following-sibling::p[position()>1]

span元素后面的第2到最后一个兄弟元素p

//\*[@id='food']/p/preceding-sibling::span[1]

p元素前面的第一个兄弟元素span

//\*[@id='food']/p/preceding-sibling::span[last()]

p元素前面的最远的一个兄弟元素span

//\*[@id='food']/p/preceding-sibling::span[position()>1]

p元素前面的第2个到最前面的一个兄弟元素span

## 父元素选择

### 使用场景：

根据子元素定位父元素

### 表示方法

//\*/.. 选择\*的父元素，类似于linux中表示上层路径的方法

示例：

//p[@class="special"]/..

选择//p[@class="special"]的父元素

//p[@class="special"]/../..

选择//p[@class="special"]的父元素的父元素

# 上节课作业代码

**from** selenium **import** webdriver  
**import** time  
  
*#如果环境变量path中有路径，可以不用写全*driver = webdriver.Chrome()  
*# 别忘了设置*driver.implicitly\_wait(10)  
  
driver.get(**'https://www.51job.com/'**)  
  
  
*#进入高级搜索*driver.find\_element\_by\_css\_selector(**'.more'**).click()  
  
*#职位搜索python*driver.find\_element\_by\_id(**'kwdselectid'**).send\_keys(**'python'**)  
  
*#选择城市*driver.find\_element\_by\_id(**'work\_position\_input'**).click()  
  
  
time.sleep(1)*#让页面有时间发生变化，  
  
#清除已经选择的地区*eles=driver.find\_elements\_by\_css\_selector(**'#work\_position\_click\_multiple\_selected span.ttag'**)  
**for** ele **in** eles:  
 ele.click()  
  
*#选择杭州*driver.find\_element\_by\_id(**'work\_position\_click\_center\_right\_list\_category\_000000\_080200'**).click()  
  
*#点击确定按钮*driver.find\_element\_by\_id(**'work\_position\_click\_bottom\_save'**).click()  
time.sleep(1)  
  
*#点击空白取消下拉菜单*driver.find\_element\_by\_css\_selector(**'#saveSearchForm+div>label'**).click()  
  
*#选择职能*driver.find\_element\_by\_id(**'funtype\_click'**).click()  
time.sleep(1)  
driver.find\_element\_by\_id(**'funtype\_click\_center\_right\_list\_category\_0100\_0100'**).click()  
time.sleep(1)  
driver.find\_element\_by\_id(**'funtype\_click\_center\_right\_list\_sub\_category\_each\_0100\_0106'**).click()  
driver.find\_element\_by\_id(**'funtype\_click\_bottom\_save'**).click()  
time.sleep(1)  
  
*#公司性质*driver.find\_element\_by\_id(**'cottype\_list'**).click()  
driver.find\_element\_by\_css\_selector(**'#cottype\_list .ul span[title="外资（欧美）"]'**).click()  
time.sleep(1)  
  
*#选择工作年限*driver.find\_element\_by\_id(**'workyear\_list'**).click()  
driver.find\_element\_by\_css\_selector(**'#workyear\_list .ul span[title="1-3年"]'**).click()  
  
*#确定搜索*driver.find\_element\_by\_css\_selector(**'.btnbox>.p\_but'**).click()  
  
*#判断页面是否加载好*driver.find\_element\_by\_css\_selector(**'.mainleft'**)  
  
*#获取职位信息：*eles=driver.find\_elements\_by\_css\_selector(**'#resultList div.el'**)  
  
**for** ele **in** eles[1:]:  
 *# if ele.get\_attribute('class').contain('title'):  
 # continue  
 # print(ele.text)* msg=ele.text.split(**'\n'**)  
 print(**'|'**.join(msg))  
  
driver.quit()