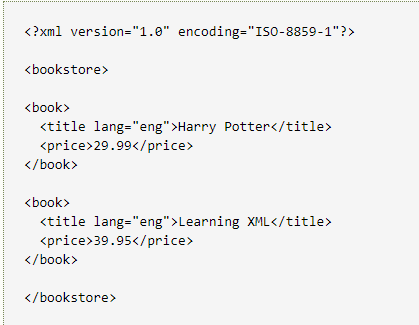
1. XPath语法当初是被用于设计搜索XML文档的，不过也可以很好的运用在HTML文档当中。Scrapy支持使用Xpath去过滤下载的Web节点。Xpath主要操作的是HTML节点，或者节点中的属性，或者节点中的文本。
2. Xpath语法为基本原理是使用***路径(Path)***来选择文档中的节点。
3. Xpath含有超过100个内建的函数。
4. XPath有7种类型的节点：元素，属性，文本，命名空间，处理指令，注释以及文档节点。
5. XPath使用路径表达式在XML文档中选取节点，节点是通过沿着路径或者step来选取的。
6. 一般常用的路径表达式：

|  |  |  |
| --- | --- | --- |
| 表达式 | | 用途 |
| Nodename | | 选取此节点的所有子节点 |
| / | | 从根节点选取 （如果路径起始于/,则所表示的路径为绝对路径） |
| // | | 从匹配选择的当前节点选择节点，从不考虑他们的位置 |
| . | | 选取当前节点 |
| .. | | 选取当前节点的父节点 |
| @属性名 | | 选取属性 |
| 谓语 (Predicates) Note: 位置采用[]进行定位。 采用[]对属性值进行筛选。 | | |
| /bookstore/book[1] | | 选取bookstore下面的一个人book子元素 |
| /bookstore/book[last()] | | 从根目录开始选取bookstore元素下面的最后一个book子元素 |
| /bookstore/book[last()-1] | | 从根目录开始选取bookstore元素下面的倒数第二个book子元素 |
| /bookstore/book[position()<3] | | 从根目录开始选取bookstore元素下面的前2个book子元素 |
| //title[@lang] | | 从根目录开始不管位置，选取拥有属性lang的元素 |
| //title[@lang=’eng’] | | 从根目录开始不管位置，选取拥有属性lang的元素，并且该值等于eng的元素 |
| /bookstore/book[price>35.00] | | 从根目录开始找寻bookstore下面的价格大于35的book子元素 |
| /bookstore/book[price>35.00]/title | | 从根目录开始寻找bookstore元素下价格大于35的book子元素下的title子元素。 |
| XPath通配符 | | |
| \* | 匹配任何元素节点 | |
| @\* | 匹配任何属性节点 | |
| Node() | 匹配任何类型的节点 | |

 示例：

对于此XML文件(可以使用在线XPath测试器：https://www.freeformatter.com/xpath-tester.html) ：

* bookstore 表示选取所有bookstore元素的子元素，子元素但是不包含孙元素，只遍历一层。
* /bookstore 表示从根节点开始选取bookstore的子节点。
* /bookstore/book 表示从根节点开始选取bookstore下面的所有book的子节点
* //book表示选取所有book元素不管他们所处的位置
* bookstore//book 选择所有bookstore下面的book的后代元素，不管是子，孙还是更深层数
* //@lang 从根节点开始，不管层数，选取下面名为lang的属性