XPath 语法总结

XPath 是用于在 XML 或 HTML 文档中定位元素的语言，以下是其核心语法：

1. **基本路径表达式**
   1. /：从根节点开始选取
   2. //：从匹配选择的当前节点选择文档中的节点，而不考虑它们的位置
   3. .：选取当前节点
   4. ..：选取当前节点的父节点
   5. @：选取属性
2. **节点选择**
   1. 标签名：直接写标签名选取所有该标签节点，如 book 选取所有 <book> 节点
   2. \*：匹配任何元素节点
   3. node()：匹配任何类型的节点
   4. @\*：匹配任何属性节点
3. **谓语（Predicates）**

用于筛选节点，写在方括号中：

* 1. book[1]：选取第一个 <book> 元素
  2. book[last()]：选取最后一个 <book> 元素
  3. book[position() < 3]：选取前两个 <book> 元素
  4. book[@lang]：选取拥有 lang 属性的 <book> 元素
  5. book[@lang='en']：选取 lang 属性值为 en 的 <book> 元素
  6. book[price>30]：选取 price 子元素值大于 30 的 <book> 元素

1. **选取多个路径**

使用 | 运算符可以选取多个路径：

* 1. //book | //cd：选取所有 <book> 和 <cd> 元素

1. **轴（Axes）**

轴定义了所选节点与当前节点之间的树关系：

示例：//book/child::title 等价于 //book/title，//book/attribute::lang 等价于 //book/@lang

* 1. ancestor：选取当前节点的所有祖先（父、祖父等）
  2. descendant：选取当前节点的所有后代（子、孙等）
  3. following：选取文档中当前节点之后的所有节点
  4. preceding：选取文档中当前节点之前的所有节点
  5. parent：选取当前节点的父节点
  6. child：选取当前节点的所有子节点（默认轴，可省略）
  7. attribute：选取当前节点的所有属性

1. **函数**
   1. text()：选取节点的文本内容，如 //book/title/text() 选取所有 <title> 元素的文本
   2. contains(string1, string2)：检查 string1 是否包含 string2，如 //a[contains(@href, 'example')]
   3. starts-with(string1, string2)：检查 string1 是否以 string2 开头
   4. last()：返回节点列表中的最后一个位置
   5. position()：返回当前节点在节点列表中的位置

掌握这些基础语法，就能应对大多数 XPath 定位需求。