**环境安装**

pip install lxml

**流程**

实例化etree对象，并将页面源码存入其中

使用etree对象的xpath方法结合xpath表达式获取数据

**实例化对象**

**用本地的html文件实例化对象**

from lxml import etree

tree=etree.parse('./try.html')

**用互联网上的页面源码实例化对象**

from lxml import etree

tree=etree.HTML()

**Xpath方法的作用及xpath表达式**

**Xpath方法(返回都是列表)**

tree.xpath(xpath表达式)

**xpath表达式**

#表示从根标签(最外层标签)开始一级一级定位a标签(/:表示下一级,用在开头是根标签)

/html/div/a

#表示从根标签(最外层标签)不管层级定位a标签(//:表示下多级,用在开头是全局定位)

/html//a

#全局搜索a标签且class为song的内容(标签名[@属性名=属性值]:可以用属性定位标签)

//a[@class='song']

#全局搜索a标签且是第3个(标签名[索引]:可以选择是第几个该标签,从1开始)

//a[3]

#定位该a标签且获取其标签之间的文本内容(标签名/text():其标签直系文本内容；标签名//text():其标签下所有的文本内容)

/html/div/a/text()

#定位img都src属性的内容(标签名/@属性名:其标签的该属性的值 )

//img/@src