Beautiful Soup--第三方HTML/XML解析库

安装库：

pip install beautifulsoup4

导入模块：

from bs4 import BeautifulSoup

相关类：

bs4.BeautifulSoup

对应属性/方法：

\_\_init\_\_(str,parser)：利用获取的http主体字符串，构造一个BeautifulSoup对象



TAG：返回Tag对象

find\_all(tag[,attr][,limit][,recursive])：搜索所有子孙节点，找到符合标签和属性的完全内容，返回ResultSet对象。recursive为False时，只搜索子节点；limit限制结果数量；attrs限制搜索的属性值

find(tag[,attr])：找到符合标签和属性的第一个子孙节点，返回Tag对象

get\_text()：返回所有内容

相关类：

bs4.element.ResultSet

对应方法：

相关类：

bs4.element.Tag

对应方法：

TAG：返回子Tag对象

name：标签名称

attrs：获取属性列表

string：获取标签内容。返回NavigableString

contents：返回所有子节点的列表。末节标签可以获得NavigableString列表

children：返回所有子节点的迭代器

descendants：返回所有子孙节点的生成器(属于迭代器)

parent：获取父节点

parents：获取所有直系祖宗节点的生成器

previous\_sibling：返回上一个亲兄弟节点

next\_sibling：返回下一个亲兄弟节点

previous\_siblings：返回之前的所有亲兄弟节点

next\_siblings：返回之后的所有亲兄弟节点

previous\_element：上一个解析节点

next\_element：下一个解析节点

previous\_elements：所有后续节点(子孙节点/兄弟节点)

next\_elements：所有前面的节点(兄弟节点及其子孙节点)

get(attr)：获得属性值

相关类：

bs4.element.NavigableString

对应方法：

replace\_with(str)：替换内容中的字符串