BS4

- beautifulsoup4
  - 解析html, xml 提取内容
  - 安装
    - bs4 pip install beautifulsoup
    - 解析器
      - 标准库 html.parse 内置 , 稍慢, 宽容度高
      - lxml 快, 宽容度高/ 安装最新版 pip / pip install lxml
  - 构造解析
    - 构造
      - soup = BeautifulSoup(html, 'html.parser')
      - soup = BeautifulSopu(open('page.html'), 'html.parser')
    - 对象分类
      - 标签: bs4.element.Tag
      - 名称: name
      - 属性:
        - tag['属性名称']
        - tag.attrs
          - ['key']
          - .get('key')
      - 可导航字符串
    - 导航 DOM
    - 搜索树
      - soup.find_all()
        - '标签' name参数, 要查找的标签名
        - 关键字 例如: id='id' / class_='link2' / id=True
        - 属性 attrs = {'key':'value'} soup.find_all(attrs={'class':'message', 'id':'xyz'})
        - string 文本
        - 限制数量 limit=值
      - soup.find()
      - soup.select('css选择器')

以上内容整理于 幕布