

第一次实验报告

吴沛琳 电院2203 522030910190

1. 实验概览

学习request请求的基本过程，了解HTML网页设计语言，学习HTML的结构和基本标签；基于三个练习了解BeautifulSoup使用方法

2. 实验环境

docker: sjtucmic/ee208

3. 练习题

1. 利用findAll函数查找所有<a>标签，获取其中href属性并保存到urlset中
2. 利用findAll函数查找所有标签，获取其中src属性并保存到urlset中
3. 根据网页源代码发现需求内容的源码特点（根节点满足<div class='wrap'>），利用findAll函数获得所有对应根节点，每一个根节点对应图片、超链接和标题，进一步搜索节点并获得相关属性保存到zhihulist中

4. 代码运行结果

三小问结果分别见codes/result1.txt, codes/result2.txt, codes/result3.txt

5. 分析与思考

思考题：超链接类型有：绝对路径、相对路径、javascript脚本等，在实践过程中还有获取过直接链接到文件（.zip文件等）的超链接

