

股票数据定向爬虫

功能描述

- 目标：获取上交所和深交所所有股票的名称和交易信息
 - 输出：保存到文件中
 - 技术路线：requests-bs4-re
-

候选数据网站的选择

新浪股票：<http://finance.sina.com.cn/stock/>

百度股票：<https://gupiao.baidu.com/stock/>

候选数据网站的选择

选取原则：股票信息静态存在于HTML页面中，无js代码生成，没有Robots协议限制

选取方法：浏览器F12, 源代码查看等等

选取心态：不要纠结于某个网站，多找信息源尝试

程序的结构设计

- 从东方财富网获取股票列表
- 根据股票列表逐个到百度股票获取个股信息
- 将结果存储到文件