

数据集说明及爬取意义

一、爬取意义：

通过新浪新闻收集、处理和分析数据来获得热点从而获得有用的信息和见解，帮助企业做出更明智的决策和策略。

二、数据集字段：

1. url：新闻的链接地址
2. title：新闻的标题
3. content：新闻的内容

这个数据集是从新浪新闻的各个版块国内、国际、军事、科技、财经、股票、金融等爬取过来。各个数据存储在名为“**news_collection.xlsx**”的 **xlsx** 文件中，数据集中的每一行表示一个单独的新闻链接，新闻标题和新闻内容信息。