

题目：淘宝售后数据爬虫系统

摘 要

现在互联网已经充满了人们的生活，大数据时代已经到来，在充斥着庞大数据量的万物互联时代中，怎样去有效地提取数据并分析是一个新的问题,本毕业设计首先介绍了在大数据环境下网络爬虫的重要性,接着介绍了网络爬虫的概念,工作原理,工作流程,网页爬行策略,python 在编写爬虫领域的优势,最后采用爬虫技术，智能化的提取出飞扬男装专卖店的售后数据，进行分析、统计，为商家优化商品结构，提高服务质量做支撑。

本系统采用 Jieba 分词以及 Wordcloud 生成词云的方式，快速有效地提取出售后数据的关键词并反馈给商家，采用 Selenium 调用 chrome 浏览器进行淘宝页面的动态渲染，可以使得页面加载完成之后再获取淘宝的售后数据信息，通过动态 IP 代理的方式防止淘宝的频繁验证，通过极验 API 破解淘宝的滑动验证码，通过 numpy 的数字分析破解淘宝的数字验证码，可以非常有效的绕过淘宝的反爬机制。

本系统与飞扬男装专卖店进行对接，进行了淘宝数据的爬取，针对专卖店的售后数据进行了特征分析，分析出了 20-35 岁人群喜爱 T 恤、衬衫、卫衣、休闲裤、牛仔裤等商品，35 岁-55 岁人群喜爱格子衬衫、POLO 衫、西装裤等商品，最畅销的商品是短袖 T 恤，每件利润平均 35 元，利润最高的商品是羽绒服，每件利润平均 300 元，主要在华北地区畅销。

飞扬男装专卖店对该系统经过两个月的使用以及店铺的自我改善，月好评率从 88%提升至 91%，月均销量从 3.5 万提升到 3.7 万，好评关键词为“质量好”、“性价比高”、“穿着效果好”，说明本系统对淘宝商家有着实际有效的应用。

关键词：淘宝售后；电子商务；爬虫；Selenium；Python

Abstract

Now the Internet has been full of people's life, and the era of big data has come. In the era of Internet of all things full of huge amount of data, how to effectively extract data and analyze it is a new problem. This graduation design first introduces the importance of web crawler in the big data environment, and then introduces the concept of web crawler, Working principle, workflow, web page crawling strategy, Python's advantages in the field of compiling reptiles. Finally, using reptile technology, intelligent extraction of the after-sales data of Feiyang men's clothing store, analysis and statistics, to optimize the commodity structure for businesses and improve the quality of service.

In this system, Jieba word segmentation and wordcloud are used to generate word cloud, which can quickly and effectively extract the key words of the data after sale and feed back to the merchants. Selenium is used to call Chrome browser for dynamic rendering of Taobao page, so that the after-sale data information of Taobao can be obtained after the page is loaded, and the frequent verification of Taobao can be prevented by dynamic IP proxy It can effectively bypass the anti climbing mechanism of Taobao by using API to break the sliding verification code of Taobao and numpy to break the digital verification code of Taobao.

The system is connected with Feiyang men's clothing store, and the Taobao data is crawled. According to the characteristics analysis of the after-sale data of the store, it is found that the 20-35-year-old people like T-shirt, shirt, sweater, casual pants, jeans and other commodities, while the 35-55-year-old people like plaid shirt, polo shirt, suit pants and other commodities. The best-selling commodity is short sleeve T-shirt, with an average profit of 35 per piece Yuan, the most profitable commodity is down jacket, with an average profit of 300 yuan per piece, which is mainly sold in North China.

After two months of use and self-improvement of the system, the monthly praise rate of Feiyang men's clothing store increased from 88% to 91%, the monthly average sales volume increased from 35000 to 37000, and the praise keywords were "good quality", "high cost performance" and "good wearing effect", indicating that the system has practical and effective application for Taobao merchants..

Key Words: Taobao after sales; e-commerce; spider; Selenium; Python

目 录

引 言	1
第 1 章 绪论	2
1.1 开发工具和环境介绍	2
Pycharm 简介	2
Web.py 简介	2
Python 简介	2
Mysql 简介	2
Selenium 简介	2
Numpy 简介	3
Vue.js 简介	3
1.2 名词解释	3
网络爬虫	3
售后数据	3
电子商务	3
第 2 章 需求分析	5
2.1 系统功能分析	5
2.1.1 系统的主要功能和特色	5
2.2 可行性研究	5
2.2.1 技术可行性	5
2.2.2 经济可行性	5
2.2.3 操作可行性	5
2.3 需求分析	6
2.3.1 问题现状分析与课题价值	6
2.3.2 课题难点及解决方式	6
2.3.3 功能需求分析	7
第 3 章 系统总体设计	15
3.1 系统功能设计目标	15

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：

<https://d.book118.com/317152035114006156>