EE208 Final Project Report

Integrated Search Engine for Elctronic Commodities

电子商品集成搜索引擎

董世文　方少恒　杨弘博　周李韬

GROUP 14

SJTU F1803016

2020 年 1 月 11 日

目录

前言

Part I

Crawler

Chapter 1

京东爬虫

1.1 网页 URL 爬取

1.2 商品信息提取

dsw

1.3 商品评论标签提取

yhb

1.4 商品评分计算

yhb

Chapter 2

苏宁爬虫

2.1 网页 URL 爬取

yhb

2.2 商品信息爬取

fsh

2.3 商品评论标签爬取

yhb

2.4 商品评分计算

yhb

Part II

Index & Search

Chapter 3

构建索引

dsw

Chapter 4

图片匹配

4.1 LOGO 匹配

fsh

4.2 LSH 匹配

fsh

Chapter 5

商品检索与排序

5.1 按相关性检索

5.2 按属性检索

Part III

Web Front-end

Chapter 6

Web 框架

6.1 web.py 配置

6.2 Bootstrap 框架

Chapter 7

网页布局

7.1 搜索主页

7.2 商品信息陈列

7.3 商品过滤功能

Appendix

写作分配

Acknowledgements

总结