**中国科学技术大学软件学院**

**软件工程实验项目环节**

**开题报告**

**项 目 名 称：基于 Spark 的景点推荐系统**

**成 员 名 单： 褚思源 林靖清 王炜康**

**导 师： 袁宇丹**

**工 程 领 域： 推荐系统**

**研 究 方 向： Java 框架技术、大数据**

**结 题 时 间： 2020.07.20**

**中国科学技术大学软件学院**

**填表日期：2020 年 7 月 19 日**

**项目总结报告**

1. 引言
2. 编写目的

写项目总结报告的目的是为了能够使我们在以后的项目开发中吸取过去的经验，不断规范开发流程以及提高开发效率。

2、背景

项目名称：基于 Spark 的景点推荐系统

随着因特网的快速发展，用户可以享受更加便利地服务。同时，面对海量数据，精准地获得目标信息也变得更加困难。一方面，用户面临对大量的信息束手无策，经常会迷失在大量的信息空间中；另一方面，网站也失去了与用户的联系，不能与用户建立稳固的合作关系。在这种背景下推荐系统（Recommender Systems）应运而生，它根据用户的趣爱好推荐符合用户兴趣爱好的对象。

本课题计划开发一个景点推荐系统。系统是一个基于 SSM 框架的 web 项目，通过 Scrapy 框架爬取景点数据，并通过 Spark 引擎进行数据处理、实现推荐算法。

其中后台管理系统对用户信息和景点信息进行管理，如添加删除景点信息和完成用户信息的完善。推荐系统通过在网站系统获取到用户的点击事件（如用户对某各景点的评分、用户的对某个景点详情页的浏览次数）记录信息，并根据根据该信息做出相应的处理，将推荐结果存入到数据库中，web 前端通过查询数据库将推荐的景点展示给用户。

3、小组成员贡献

|  |  |  |  |
| --- | --- | --- | --- |
| 姓名 | 学号 | 项目中的分工 | 贡献比 |
| 林靖清 | SA19225242 | 主要负责推荐系统 | 33.3% |
| 王炜康 | SA19225378 | 主要负责爬虫 | 33.3% |
| 褚思源 | SA19225123 | 主要负责网站 | 33.3% |

4、参考资料

项目文档

需求分析文档

概要设计文档

编码规范

详细设计文档

接口设计文档

用户手册

项目安装手册

项目技术论文

二、开发成果

1、产品

项目是基于 Spark 的景点推荐系统，利用爬取的信息构建旅游网站，供用户查看、评论等。根据用户对景点的评分，为他推荐他可能感兴趣的景点。

1. 产品功能

构建的旅游网站可以让用户查看景点信息，包括介绍、位置、热度、图片等等，同时用户可以对景点进行打分和评论。推荐系统可以根据用户对景点的评分，为他推荐他可能感兴趣的景点。

1. 基本流程

使用 Scrapy-Redis 搭建分布式爬虫系统、使用 Vue + SSM 搭建旅游网站、使用 ALS 协同过滤算法构建推荐系统。

1. 进度

按照预期完成了爬虫系统、网站、推荐系统的开发。

1. 费用

购买服务器

三、开发工作评价

1、生产效率评价

基本按时完成开发任务

2、产品评价

（1）爬取的数据不够多，还可以加入更多的景点和评论；

（2）网站用户基数少，推荐系统评分矩阵比较稀疏。

3、技术方法评价

构建了 Scrapy-Redis 分布式爬虫系统，提高了爬虫效率。使用 Vue + SSM 搭建网站，界面简介清晰，并且功能丰富。推荐系统使用 ALS 协同过滤算法，使用评分矩阵比较直观。

四、项目经验总结

1、项目生产周期

项目周期包括三个阶段：项目的准备阶段、项目实施阶段、项目总结阶段。在项目生命周期的初始阶段，最关键的工作是明确项目的概念和制定计划，并且在之后的阶段里要严格按照指定的计划执行，按时完成。

2、开发团队

在项目中有许多相互依赖、相互影响的工作，因此，一些工作的结果往往直接影响到下一步的工作。在项目管理的过程中，沟通必须非常真实，管理层应该及时劝阻或者惩罚任何传递不正确沟通信息的项目成员，因为只有准确的数据，才能使项目评估总结工作准确无误。获得准确绩效评估的方法是和关键项目成员一起采取阶段性评估，来总结整个项目的进展。在开展的阶段性评估中，讨论已完成、未完成，以及即将实施的工作，并做出正确的评估，参会的人员通常有外来服务商，他们可起到仲裁作用，并能提出有用的建议。

3、需求分析

需求分析的内容是针对待开发软件提供完整、清晰、具体的要求，确定软件必须实现哪些任务。具体分为功能性需求、非功能性需求与设计约束三个方面。

功能性需求即软件必须完成哪些事，必须实现哪些功能，以及为了向其用户提供有用的功能所需执行的动作。

非功能性需求，作为对功能性需求的补充，软件需求分析的内容中还应该包括一些非功能需求。主要包括软件使用时对性能方面的要求、运行环境要求。

设计约束，一般也称做设计限制条件，通常是对一些设计或实现方案的约束说明。

4、调研

调研用户领域的组织结构、岗位设置和职责定义，从功能上区分有多少个子系统，划分系统的大致范围，明确系统的目标。

调研每个子系统所需的工作流程、功能与处理规则。

对调研的内容事先准备，针对不同管理层次的用户询问不同的问题，列出问题清单。

对与用户沟通的情况及时总结归纳，整理调研结果，找出新的疑点，初步构成需求基线。