**河北传媒学院**

**本科毕业论文开题报告**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 专业 | 数据科学与大数据技术 | 班级 | | 2020级本科班 |
| 姓名 | 路天祎 | 学号 | | 2008090023 |
| 指导教师 | 刘爱珍 | | | |
| 题目 | 基于Spark的河北省旅游景区推荐系统的设计与实现 | | | |
| 一、课题的来源及意义  1.来源  随着国家不断发展和进步，人们物质水平也在不断提高。与此同时，人们也更加趋于追逐精神生活的富足。满足自身精神层面富足的方式有很多，而大多数人会选择旅游的方式来进行精神层面的放松。 旅游可以适当满足自身的情感需求，放松身心，使精神层面解除紧绷的状态。但是，在如今快节奏的时代里，人们的假期大都是有限并短暂的，并不足以让人们充分了解景区信息，制定自身旅游计划，充分满足自身需求。  同时，河北省作为一个拥有丰富自然和人文景观资源的地区，也吸引了大量的游客。然而，由于游客数量庞大和景区众多，游客如何选择合适的景区成为一个难题。因此，设计并实现一套基于Spark的河北省旅游景区推荐系统，将能够提供游客个性化的推荐服务，帮助游客更好地规划旅游行程，提升旅游的体验。  Spark技术具有快速处理大规模数据、支持多种语言、具有灵活性和拓展性、并且能应用于实时数据流场景的特点。因此通过Spark技术，可以实现对旅游数据的实时监测。基于此，本设计旨在通过数据爬取、数据存储和Spark技术实现一个基于Spark的便捷的、高效的河北热门旅游景区推荐系统。  2.意义  设计并实现一套基于Spark的河北省旅游景区推荐系统，将能够提供游客个性化的推荐服务，帮助游客更好地规划旅游行程，提升旅游的体验；并且还能在有效的时间内，弥补自己对旅游景区的不了解的不足，制定令人满意的旅游计划，迎接愉快的度假。  二、国内外发展现状  1.国外发展现状：  在国外，旅游推荐系统已经得到了广泛的应用。一些主流的旅游网站，如TripAdvisor、Expedia和Booking.com等，已经在大力推广和运用旅游推荐系统。这些推荐系统主要基于协同过滤、内容过滤和混合过滤等技术，通过对用户的历史浏览记录、评价及喜好偏好进行分析，从而给出符合用户需求的推荐产品。  2.国内发展现状：  随着互联网的发展和旅游行业的兴盛，国内的旅游推荐系统也逐渐得到了关注和发展。目前，携程、去哪儿等大型在线旅游平台及一些小型旅游网站，也主要通过协同过滤、内容过滤和混合过滤等技术实现推荐系统的应用。然而，目前国内旅游系统也有一些瑕疵。例如对于景区选择性推荐还不够完善。由于景区特色的不同，其适应的旅游季节也会有所不同。因此，对于景区选择性推荐问题还需不断完善。  综上所述，旅游景区推荐系统在国内外都得到了广泛的关注和研究。通过spark和协同过滤的技术，这些系统能够提供个性化的旅游推荐服务，提升用户的旅游体验。而基于spark的河北省旅游景区推荐系统也借鉴了国内外所使用的良好算法，例如协同过滤算法，也可以弥补景区选择性推荐不够完善的问题。  三、主要内容及目标  1.主要内容：  python爬虫获取河北省旅游景区信息，对所获取的数据进行清洗等，并将其存储到数据库中；  通过问卷调查的方式获取用户的年龄、喜欢的主题、旅游作息、旅游月份以及其他相关信息；  使用spark技术、django框架、协同过滤算法等进行系统的搭建和实现；  2.主要目标：  1）设计目标  该平台具备用户功能和管理员功能，其中，用户功能包括用户的注册登录、景区推荐、景区浏览、景区筛选、景区点赞和收藏、制定个性化旅游计划、个人信息管理等功能。管理员功能包括管理员登录、旅游景区管理、评论管理、用户管理、旅游计划管理等功能。  2）设计内容  首先通过爬虫爬取旅游网站的景区信息。  然后通过问卷调查的方式获取用户的年龄、喜欢的主题、旅游作息、旅游月份以及其他相关信息。  本系统分为用户功能和管理员功能；  ①用户功能  注册登录：用户在使用本系统进行景区预约、订票等操作时需要进行注册以及登录操作，以便更好地制定个人旅游规划；  景区推荐：根据调查问卷收集到的用户喜好，以及登录用户的个人信息、搜索习惯等，通过推荐算法，为登录用户提供个性化的景区推荐。添加推荐算法的主要目标是为了提高用户体验，推荐更符合用户个人需求的旅游景区；  景区浏览：用户可以查看景区的详情介绍、图片、地址、主题、评分 、评论等信息，以便帮助用户更好的了解景区；  景区筛选：用户可以根据景区主题、月份、距离远近等进行筛选，以便用户更快捷的获取符合个人预期的景区；  景区点赞和收藏：用户可以对自己感兴趣的景区进行点赞或收藏操作，以便用户制定个性化旅游计划时，可以以此作为参考；  制定个性化旅游计划：用户可根据个人点赞、浏览或者收藏的景区，制定个性化旅游规划；  个人信息管理：用户对个人基本信息的完善及后续修改。例如:用户名称、密码等；  ②管理员功能  1）管理员登录：管理员通过输入用户名、密码等进入本系统后台，登录时验证账号密码，系统随机生成验证码，防止恶意登录；  2）旅游景区管理：此处可以管理通过爬虫进行数据采集的景区数据，另外可以对旅游景区进行新增修改等操作；  3）评论管理：管理员管理用户对景区评论信息；  4）用户管理：管理员在此处查看所注册本系统的用户基本信息；  5）旅游计划管理：管理用户所制定的旅游计划；  四、设计方法、手段、及主要任务  1.运用python爬虫技术获取网站的旅游景区信息，并将数据存储到Mysql中，对数据进行处理。  2.运用django框架，进行系统代码的编写和执行，并且使用协同过滤算法和spark技术进行景区推荐。  3.测试代码，确定程序最终完成效果。  五、进度计划安排  1. 选题开题  选题阶段：2023年09月04日—2023年10月31日  开题阶段：2023年11月01日—2023年11月15日  2. 设计制作  初步设计阶段：2023年11月16日—2023年12月15日  整体设计阶段：2023年12月16日—2024年01月15日  完成系统设计：2024年01月16日—2024年02月10日  3. 撰写论文  论文初稿：2024年02月11日—2024年2月底  论文二稿：2024年03月01日—2024年3月31日  论文终稿：2024年04月01日—2024年4月20日  4. 毕业答辩  毕业答辩时间：2024年4月底  六、可行性分析与已具备条件  （1）可行性分析  1、需求分析：  ①功能需求：  该系统主要是给具有在河北省内旅游需求的人士提供的，该系统主要提供以下功能：  旅游景区筛选：用户可通过该功能快速的确定符合自己需求的景区。  个人旅游计划界面：用户可以通过查看自己点赞、浏览的景区信息，制定个性化旅游规划。  ②数据需求：  旅游景区数据基本信息包括旅游景区名称、景区评分、景区票价、景点分类等。为了更好的方便用户的查询和筛选；  获取用户的年龄、喜欢的主题、旅游作息、旅游月份以及其他相关信息。以便系统根据推荐算法推荐用户可能喜欢的景区。  2、技术可行性：  项目使用流行的开源的应用框架django,框架可以实现应用的整个搭建和开发过程。  数据获取方面：Python是免费的而且开源的编程语言，降低时系统开发和维护的成本，同时Python爬虫技术发展至今，已经具有非常成熟的技术架构，而且Python拥有许多优秀的第三方库，可以用它顺利的爬取本项目需要的数据，并且存储在搭建好的数据库中。  数据储存方面：MySQL，运行速度快，使用成本低，与其他大型数据库的设置和管理相比，其复杂程度较低，易于使用，完全符合本系统的构想。  技术方面：Spark 可以用于构建推荐系统，例如基于协同过滤的推荐算法。通过分析用户行为和喜好数据，可以为用户提供个性化的产品或内容推荐。  3、法律可行性：  ①隐私保护：当涉及用户数据收集和分析时，特别需要遵守相关的隐私法律法规，例如欧盟的GDPR以及其他国家或地区的个人数据保护法规。  ②版权和知识产权：如果推荐系统使用了其他人创造的内容，例如图片、文字或音频，应确保获得了合法的版权或授权，并遵守相关法律法规。  ③歧视和平等原则：在推荐景区时，需要避免任何形式的歧视行为或偏见，确保推荐系统的运作是公正和平等的。  4、经济可行性：  ①市场需求：确定河北省旅游景区目标受众和了解潜在市场需求。了解目标用户的特点、旅游偏好和消费能力，以确保推荐系统能够满足他们的需求。  ②收入来源：  广告：与相关的旅游商家合作，将他们的广告展示在推荐系统中，从中获得收入。  ③成本分析：  开发成本：评估系统开发、测试、部署和维护所需的人力资源和技术投入成本。  基础设施成本：考虑Spark集群的硬件和软件资源需求，并评估相应的成本。  运营成本：包括服务器租赁、数据存储、网络带宽、营销费用、人员培训和运维成本等。  数据获取成本：获取和更新旅游景区相关数据的成本，例如景区介绍、评价和用户反馈等。  ④收益预测：根据市场需求、定价策略和收入来源估计推荐系统的收入，并与成本进行比较，以确定经济可行性。  （2）已具备的实验条件  本人具有充足的技术储备,熟悉python开发工具以及Django框架的熟练运用,项目整体设计功能较为完善,具备开发旅游平台所需的全部条件。  七、主要参考文献  [1]李艳杰.MySQL数据库下存储过程的设计与应用[J].信息技术与信息化.2021  [2]胡阳.Django企业开发实战[M].人民邮电出版社  [3]张凡.基于Web的旅游服务平台的设计与实现.[D].软件工程.2018  [4]綦慧，徐晓慧.基于Web系统的大数据搜索技术的实现和优化.[A].信息学.2021  [5]白秀军.基于大数据的枣庄市智慧旅游服务平台构建研究.[D].旅游管理.2018 | | | | |
| 选题是否合适： 是 否  课题能否实现： 能 不能  指导教师（签字）  年 月 日 | | | 选题是否合适： 是 否  课题能否实现： 能 不能  指导小组组长（签字）  年 月 日 | |