

本科毕业设计（论文）工作记录本

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 学 院 | 计算机科学与技术学院 | | | | 专业 | | 计算机科学与技术 | | |
| 学生姓名 | 孟政元 | | | | 学号 | | 21140821 | | |
| 指导教师 | 王康平 | | | | 职称 | | 讲师 | | |
| 合作导师 |  | | 职称 |  | 单位 | | 计算智能 | | |
| 设计（论文）题目 | | 网易云音乐用户数据的分析和可视化技术研究与实现 | | | | | | | |
| 题 目 性 质 | | 工程类（实际项目） | | | |  | | 研究类（基础研究） |  |
| 工程类（自选项目） | | | | √ | | 研究类（应用研究） |  |
| （注：单选，题目性质在相应栏内填写“√”,与课题论证书一致） | | | | | | | |
| 起 止 日 期 | | 2018 年 1 月 1 日 —— 2018 年 5 月 14 日（共 14 周） | | | | | | | |
| 指导教师签字 | |  | | | | | | | |
| 研究室主任签字 | |  | | | | | | | |

**注：**专业一定要填写全称。[专业名称：计算机科学与技术、计算机科学与技术（网络与信息安全）、物联网工程、理科试验班（唐敖庆计算机班）]

吉林大学教务处制

工 作 记 录 表

|  |
| --- |
| **毕业实习（调研）**  时间： 2018 年 3 月 3 日 —— 2018 年 3 月 10 日（第 1 周——第 2 周）  地点：  完成任务情况： |
| **毕业实习（调研）**  时间： 2018 年 3 月 11 日 —— 2018 年 3 月 17 日（第 2 周——第 3 周）  地点：  完成任务情况： |

注：共四类表格，可根据需要复制各类表格并填写；要求每周都有工作记录。

工 作 记 录 表

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **查阅资料目录**   |  |  |  |  | | --- | --- | --- | --- | | 序号 | 文献名称 | 文献来源（卷期页） | 阅读时间 | | 1 | 大数据时代下的数据可视化研究 | 河北大学, 2014 | 2018年 03月 06日（ 第 1 周） | | 2 | 数据可视化的研究与发展 | 计算机工程, 2002, 28(8):1-2 | 2018年 03月 06日（ 第 1 周） | | 3 | 数据可视化技术及其应用 | 红外与激光工程, 2001, 30(5):339-342 | 2018年 03月 06日（ 第 1 周）） | | 4 | 数据可视化技术及其应用展望 | 全国自动化新技术学术交流会会议论文集(一). 2005 | 2018年 03月 06日（ 第 1 周） | | 5 | 数据可视化在数据挖掘中的应用 | 计算机应用与软件, 2003, 20(11):71-73 | 2018年 03月 06日（ 第 1 周） | | 6 | 基于Web的数据可视化技术初探 | 福建电脑, 2007(12):58-59 | 2018年 03月 06日（ 第 1 周） | | 7 | 大数据时代背景下的数据可视化应用研究 | 电子制作, 2013, 47(5x):118-118 | 2018年 03月 06日（ 第 1 周） | | 8 | 可视化库D3.js的应用研究 | 信息技术与信息化, 2015(2):107-109 | 2018年 03月 06日（ 第 1 周） | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |

注：共四类表格，可根据需要复制各类表格并填写；要求每周都有工作记录。

工 作 记 录 表

|  |
| --- |
| **答 疑**  时间： 2018 年 1 月 1 日 10 时 （第 1 周）  地点：线上  形式：面谈( ) 电话（ ）邮件（ √） QQ（ ） 微信（ ） 其他（ ）  内容提要：和指导老师进行题目的初步探讨，和老师交换了关于题目的意见和见解， 根据老师的要求进行题目研究的细节的具体化，为下一步进行技术上的选择做好准备。 |
| **答 疑**  时间： 2018 年 1 月 8 日 10 时 （第 2 周）  地点：线上  形式：面谈( ) 电话（ ）邮件（ √） QQ（ ） 微信（ ） 其他（ ）  内容提要：把自己关于题目的看法和具体的实现细节向老师做汇报，同时向老师讲述了自己关于论文在技术上的实现相关的选择，老师对这些进行了解之后做出了相关的考虑，并决定了论文的基本方向和技术选择。 |

注：共四类表格，可根据需要复制各类表格并填写；要求每周都有工作记录。

工 作 记 录 表

|  |
| --- |
| 时间： 2018 年 1 月 15 日 （第 3 周）  记事：   1. 根据论文的大纲和与老师讨论的结果确定论文的实现细节 2. 确定了论文需要爬取的数据 3. 确定了采用Node.js，JavaScript作为基本开发语言 4. 确定了采用D3来进行数据的可视化 5. 确定了采用MongoDB来作为数据库存储从网络上爬取的数据 |
| 时间： 2018 年 3 月 5 日 （第 4 周）  记事：  根据需要爬取的数据，编写了初步可用的爬虫程序，并且使用爬虫程序爬取了少许的数据，但是在爬取的过程中遇到了下面的问题：   1. 爬取的数据比较的复杂，需要根据需要进行相应的取舍 2. 爬取的速度过快，触发了网易云音乐的反爬虫机制，导致爬取失败 |

注：共四类表格，可根据需要复制各类表格并填写；要求每周都有工作记录。

工 作 记 录 表

|  |
| --- |
| 时间： 2018 年 3 月 12 日 （第 5 周）  记事：针对前面出现的问题，寻求规避网易云音乐反爬虫机制的方案，初步设想了一下几个方案：   1. 降低爬取速度，花费更长的时间来进行数据的爬取 2. 购买代理IP，通过使用多个IP，每个IP慢速爬取来防止被屏蔽 3. 自己搭建代理服务器集群   在综合考虑了各个方案之后，决定自己搭建代理服务器集群来解决问题 |
| 时间： 2018 年 3 月 19 日 （第 6 周）  记事：为了搭建自己的代理服务器集群，在网上查阅了很多的关于HTTP代理的资料，详细了解了HTTP代理的内部细节，并且根据查阅的资料编写了一个简单的，能够达到自己要求的代理程序，并且采取了制作镜像的方式部署到了多个服务器上， 组成了一个可以快速爬取数据的代理服务器集群，帮助爬虫程序规避反爬机制，快速的爬取到足量的数据 |

注：共四类表格，可根据需要复制各类表格并填写；要求每周都有工作记录。

工 作 记 录 表

|  |
| --- |
| 时间： 2018 年 3 月 26 日 （第 7 周）  记事：在爬取的足量的数据之后，需要对存储在数据库中的数据进行初步的处理和分析，但是数据库所在的服务器配置并不高，所以需要对数据库做优化，减少查询的时间。  数据存储在自己搭建的MongoDB数据库中，为了对数据库进行优化， 在网上阅读了MongoDB的官方文档，对MongoDB的基本特性有了大致的了解，结合程序需要用到的功能，对数据库进行了优化，加快了数据查询的速度，为之后的分析做好了准备。 |
| 时间： 2018 年 4 月 2 日 （第 8 周）  记事：在之前爬取到的数据和对数据库进行了优化的基础上，编写程序对数据进行分析。经过相应的处理之后，待到了下一步进行可视化需要的各项数据：   1. 用户总数 2. 男女用户， 性别不明的用户的数量 3. 用户的年龄信息 4. 用户的地域信息 5. 用户之间相互关注的信息 |

注：共四类表格，可根据需要复制各类表格并填写；要求每周都有工作记录。

工 作 记 录 表

|  |
| --- |
| 时间： 2018 年 4 月 9 日 （第 9 周）  记事：确定数据的可视化要基于D3来实现，在网上检索了与D3相关的资料，了解了数据可视化当前的基本情况，以及基于D3的数据可视化当前的发展状态以及常用的方法和工具，并进行了相应的学习。  在学习D3可视化的过程中，发现基于地图来进行可视化可以更加直观的展示数据的某些特征在地理位置上的分布情况，而且具有更好的交互性，所以打算将基于地图的可视化技术用于项目中，并且基于此进行了一系列的准备和研究。 |
| 时间： 2018 年 4 月 16 日 （第 10 周）  记事：在对D3的学习的基础上，通过从网络上获取的中国地图的数据，通过使用D3和SVG在网页上绘制除了中国地图和各个省份的图形，并且添加了鼠标的交互，为之后带入实际数据做准备。 |

注：共四类表格，可根据需要复制各类表格并填写；要求每周都有工作记录。

工 作 记 录 表

|  |
| --- |
| 时间： 2018 年 4 月 23 日 （第 11 周）  记事：在获取了分析结果和数据的可视化的基础准备的基础上，将分析的数据结果带入可视化中进行展现，并且对程序进行调试，解决了一些可视化过程中遇到的适配问题，并且添加了一些用户交互动作，让可视化更加易于使用，易于观察。 |
| 时间： 2018 年 4 月 30 日 （第 12 周）  记事：根据爬取到的用户之间互相关注的数据，构建了用户关系有向图对应的矩阵，并且在经过初步处理之后采用floyd算法计算了各个用户之间的最短路径，分析了用户相互关注的基本特征。 |

注：共四类表格，可根据需要复制各类表格并填写；要求每周都有工作记录。

工 作 记 录 表

|  |
| --- |
| 时间： 2018 年 5 月 7 日 （第 13 周）  记事：在前面的的数据爬取，数据存储，数据分析，数据可视化的基础上，根据之前已经确定的论文的大纲，进行了论文的初稿的撰写，并且进行了相应的排版的处理，检查没有错别字的其他的错误之后交给了老师进行检查，得到了老师的意见和建议。 |
| 时间： 2018 年 5 月 14 日 （第 14 周）  记事：在论文的初稿和老师的指导意见的基础上，对论文进行最终的修改，修改了论文中一些不合理的地方，添加了一些内容，让论文更加的完整的流畅，并且对论文的排版进行了最终的修改，对排版中不合理的地方进行了更正。 |