郑州大学信息管理学院毕业论文（设计）书面成绩评分表

（指导教师用表）

专业：信息管理与信息系统 班级：一班 姓名：张磊 学号：20133410139

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 评价内容 | 具体要求 | 分值 | 评分 | | | | |
| A | B | C | D | E |
| 调查论证 | 能独立查阅文献和从事其它调研；能正确翻译外文资料；能提出并较好地论述课题实施方案；有收集、加工各种信息及获取新知识的能力。 |  | 10 | 9 | 8 | 7 | ≤6 |
| 实验方案能力与实验技能 | 能正确设计实验方案，独立进行实验工作，如装置安装、调试、操作等。 |  | 20 | 18 | 16 | 14 | ≤  12 |
| 分析与解决问题的能力 | 能运用所学知识和技能去发现与解决实际问题；能正确处理实验数据；能对课题进行理论分析，得出有价值的结论。 |  | 20 | 18 | 16 | 14 | ≤  12 |
| 工作量、工作态度 | 按期圆满完成规定任务，工作量饱满，难度较大；工作努力，遵守纪律；工作作风严谨务实。 |  | 20 | 18 | 16 | 14 | ≤  12 |
| 论文  （设计）质量 | 立论正确，论述充分，结论严谨合理；实验正确，分析处理科学；文字通顺，技术用语准确，符号一致，编号齐全，书写工整规范，图表完备、简洁、正确；论文结果有应用价值。 |  | 20 | 18 | 16 | 14 | ≤  12 |
| 创 新 | 工作中有创新意识；对前人工作有改进或突破，或有独特见解。 |  | 10 | 9 | 8 | 7 | ≤  6 |
| 总 分 |  |  |  |  |  |  |  |

指导教师签名：

郑州大学信息管理学院毕业论文（设计）书面成绩评分表

（评阅人用表）

专业：信息管理与信息系统 班级：一班 姓名：张磊 学号：20133410139

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 评价内容 | 具体要求 | 分值 | 评分 | | | | |
| A | B | C | D | E |
| 选题质量与指导老师条件 | 选题符合专业培养目标，课题性质、难度、与份量能满足毕业设计（论文）综合训练的教学要求，指导老师有科研背景，指导学生数适当。 |  | 15 | 13 | 11 | 9 | 5 |
| 查阅资料与翻译文献 | 查阅文献有一定广泛性；翻译外文资料质量较好；有综合归纳资料的能力和自己见解。 |  | 10 | 8 | 7 | 5 | 3 |
| 论文  （设计）  质量 | 立论正确，论述充分，结论严谨合理；实验正确，分析处理科学；文字通顺，技术用语准确，符号一致，编号齐全，书写工整规范，图表完备、简洁、正确；论文结果有应用价值。 |  | 50 | 45 | 40 | 35 | 30 |
| 资料  完整性 | 附件资料系统完整，编排装订有序。 |  | 15 | 13 | 10 | 7 | 5 |
| 创 新 | 对前人工作有改进或突破，或有独特见解 |  | 10 | 8 | 7 | 5 | 3 |
| 总 分 |  |  |  |  |  |  |  |

评阅人签名：

郑州大学信息管理学院毕业论文（设计）答辩成绩评分表

（答辩委员会用表）

专业：信息管理与信息系统 班级：一班 姓名：张磊 学号：20133410139

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 评价内容 | 具体要求 | 分值 | 评分 | | | | |
| A | B | C | D | E |
| 报告内容 | 论述思路清晰；语言表达准确，概念清楚，论点正确；方法科学，分析归纳合理；结论严谨；论文结果有应有价值。 |  | 50 | 45 | 40 | 35 | 30 |
| 创 新 | 对前人工作有改进或突破，或有独特见解 |  | 10 | 9 | 8 | 7 | 6 |
| 答 辩 | 回答问题有立论根据，基本概念清楚。主要问题回答准确，有逻辑性。 |  | 30 | 27 | 24 | 21 | 18 |
| 报告时间 | 符合要求 |  | 10 | 9 | 8 | 7 | 6 |
| 总 分 |  |  |  |  |  |  |  |

答辩委员会委员： 、 、

、 、

郑州大学信息管理学院毕业设计（论文）成绩评定表

专业：信息管理与信息系统 班级：一班

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 姓名 | 张磊 | 学号 | 20133410139 | 总成绩 |  |
| 题目 | 网络引文的爬虫系统初探 | | | | |
| 指  导  教  师  评  语 |  | | | | |
| 评定成绩： 签名： 年 月 日 | | | | |
| 评  阅  人  评  语 |  | | | | |
| 评定成绩： 签名： 年 月 日 | | | | |
| 答  辩  小  组  评  语 |  | | | | |
| 评定成绩： 签名： 年 月 日 | | | | |

郑州大学信息管理学院毕业设计（论文）课题申请表

（由指导教师填写）

专业：信息管理与信息系统 时间：

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 课  题  情  况 | 课题名称 | 网络引文的爬虫系统初探 | | | | | | | |
| 教师姓名 | 郭强 | 职称 | 教授 | 学位 | | 博士 | 教研室 | 信息管理 |
| 课题来源 | A.科研 B.论文 C.教学 D.其它 | | | | | | | |
| 课题类别 | A.设计 B.论文 | | 实施地点 | | 郑州大学 | | | |
| 主  要  研  究  内  容 | 本文以万方数据库为例，讲述网络引文的爬取过程。首先介绍了网络爬虫的技术基础，然后研究了网络引文爬虫的一些关键问题如根URL的选取、URL消重等，最后实现了一个针对万方数据库的爬虫系统。 | | | | | | | | |
| 目  标  和  要  求 | 能够实现对网络引文的爬取和存储，并在爬取效率和数据准确性方面满足基本要求。 | | | | | | | | |
| 特  色 | 选题有现实意义，具有实用性。 | | | | | | | | |
| 教研  室审  题意  见 | 负责人签字： 年 月 日 | | | | | | | | |
| 学院  审批  意见 | 主管院长签字： 年 月 日 | | | | | | | | |

郑州大学信息管理学院毕业设计（论文）任务书

（由指导教师填写）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 学生姓名 | 张磊 | 学号 | 20133410139 | 教师姓名 | 郭强 | 职称 | 教授 |
| 题目 | 网络引文的爬虫系统初探 | | | | | | |
| 主要研究内容 | 本文以万方数据库为例，讲述网络引文的爬取过程。首先介绍了网络爬虫的技术基础，然后研究了网络引文爬虫的一些关键问题如根URL的选取、URL消重等，最后实现了一个针对万方数据库的爬虫系统。 | | | | | | |
| 研究与设计方法 | 系统分析和系统测试 | | | | | | |
| 主要技术指标（或研究目标） | 能够实现对网络引文的爬取和存储，并在爬取效率和数据准确性方面满足基本要求。 | | | | | | |
| 主要参考文献 | [1] 数据挖掘基础知识.[EB/OL].[2017-5-9].  http://www.xuexila.com/baikezhishi/1244762.html  [2] 赵茉莉.网络爬虫系统的研究与实现[D].成都:电子科技大学,2013:6~7  [3] 李海燕.网络舆情爬虫系统的设计与实现[D].厦门:厦门大学,2014:6~7,20  [4] 孙立伟,何国辉,吴礼发.网络爬虫技术的研究[J].电脑知识与技术,2010,6(15):4112~4114  [5] 李盛韬,余智华,程学旗,白硕.Web信息采集研究进展[J].计算机科学,2003(2):151~156  [6] 曾伟辉.支持AJAX的网络爬虫系统设计与实现[D].合肥:中国科学技术大学,2009:7~8  [7] node.js.[EB/OL].[2017-5-11].http://baike.baidu.com/item/node.js?fromtitle=nodejs  [8] 王越.基于nodejs的微博系统的设计与实现[D].成都:电子科技大学,2014:8~9  [9] 正则表达式.[EB/OL].[2017-5-15].http://www.jianshu.com/p/12e17ff09176  [10] 陈瑜芳.主题爬虫系统的研究[D].武汉:武汉理工大学,2010:13~14 | | | | | | |

郑州大学信息管理学院毕业设计（论文）计划进程表

（由学生填写）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 学生姓名 | 张磊 | 学号 | 20133410139 | 教师姓名 | 郭强 | 职称 | 教授 |
| 题目 | 网络引文的爬虫系统初探 | | | | | | |
| 时间 | 工作内容 | | | 完成情况 | | | |
| 2017.2.27－2017.3.5  2017.3.6－2017.3.19  2017.3.20－2017.3.26  2017.3.27－2017.5.14  2017.5.15－2017.5.28 | 选题  资料调研、开题  拟定与修改论文大纲  论文初稿  论文修改、定稿 | | | 顺利完成  完成国内外相关文献普查  顺利完成  完成  完成 | | | |

郑州大学信息管理学院毕业设计（论文）开题报告

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 题目名称 | 网络引文的爬虫系统初探 | | | | |
| 学生姓名 | 张磊 | 所学专业 | 信息管理与信息系统 | 班级 | 一班 |
| 1. 选题的目的和意义   近年来，计算机和信息技术有了突飞猛进的发展，产生了许多新概念和技术，如更高性能的计算机和操作系统、因特网（internet）、数据仓库（data warehouse）、神经网络等等。在市场需求和技术基础都满足的情况下，数据挖掘的概念和技术应运而生。  数据挖掘（Data Mining）旨在从大量的、不完全的、模糊的随机数据中提取出人们事先不知道的潜在的有价值的信息或知识。  网络的迅速发展使各行各业都开始采用计算机及相应的信息技术进行管理和运营。网络引文由于其传播方式的自由性、动态性，也逐渐成为论文参考文献的主要来源之一。为了能更方便的对论文中引用到的网络引文进行数据分析和数据挖掘，我们需要更方便的收集论文的网络引文，然而由于数据量较大、手工操作较为复杂，导致网络引文分析所需的数据供给不足，本文以此为出发点通过设计并实现一个网络引文爬虫系统来爬取文献信息，提供数据分析和数据挖掘所需的海量数据。 | | | | | |
| 1. 国内外研究综述   网络爬虫系统主要是用来在浩瀚网络中收集网页或特定类型的数据信息，然后从中解析出客户所需要的信息，并持久化到数据库系统。  目前在国内外存在很多互联网爬虫系统，很多企业也会利用爬虫系统定时爬取所需数据。  基于web的爬虫系统是最普遍的，从某个URL作为根路径，然后对整个互联网节点进行信息采集，适用于广泛的主题搜索应用。这类系统特点是采集数据量大，并且缺少针对性。代表系统有Mercator和Google。  另一种基于Agent的互联网爬虫系统，是利用Agent技术在信息采集的同时，对用户倾向进行观察和记录，然后不断调整数据采集策略。Agent爬虫系统能够智能、灵活地处理某些个性化信息采集。这类爬虫的代表系统有Amalthaca、美国爱荷华大学的ARACHNID。  还有一种是基于增量的采集系统，主要用来采集那些数据更新频繁的网页。工作原理是：在完成初次数据采集后，后续的采集只针对那些改变了的数据，对于未改变的数据将不再处理。增量采集方式可以提高数据爬取的效率，最大程度地提高网络和磁盘等资源的利用率。代表系统有：天网增量搜集系统、Chile Crawler。  三、毕业设计（论文）所采用的方法  系统分析和系统测试的研究方法。 | | | | | |
| 1. 主要参考文献与资料获得情况   主要参考文献如下：  [1] 数据挖掘基础知识.[EB/OL].[2017-5-9].  http://www.xuexila.com/baikezhishi/1244762.html  [2] 赵茉莉.网络爬虫系统的研究与实现[D].成都:电子科技大学,2013:6~7  [3] 李海燕.网络舆情爬虫系统的设计与实现[D].厦门:厦门大学,2014:6~7,20  [4] 孙立伟,何国辉,吴礼发.网络爬虫技术的研究[J].电脑知识与技术,2010,6(15):4112~4114  [5] 李盛韬,余智华,程学旗,白硕.Web信息采集研究进展[J].计算机科学,2003(2):151~156  [6] 曾伟辉.支持AJAX的网络爬虫系统设计与实现[D].合肥:中国科学技术大学,2009:7~8  [7] node.js.[EB/OL].[2017-5-11].  http://baike.baidu.com/item/node.js?fromtitle=nodejs  [8] 王越.基于nodejs的微博系统的设计与实现[D].成都:电子科技大学,2014:8~9  [9] 正则表达式.[EB/OL].[2017-5-15].http://www.jianshu.com/p/12e17ff09176  [10] 陈瑜芳.主题爬虫系统的研究[D].武汉:武汉理工大学,2010:13~14 | | | | | |
| 1. 指导教师审批意见   年 月 日 | | | | | |

郑州大学信息管理学院毕业设计（论文）中期检查表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 题目名称 | 网络引文的爬虫系统初探 | | | | |
| 设计人姓名 | 张磊 | 所学专业 | 信息管理与信息系统 | 班级 | 一班 |
| 1. 阶段性成果 2. 检索相关文献，学习毕业论文的相关结构与排版。 3. 根据检索到的参考文献，确定论文写作的大纲。 4. 通过系统分析和系统测试的方法，设计和实现了网络引文的爬虫系统。 5. 初步完成论文初稿。 6. 根据导师提出的修改意见，对论文初稿进行修改。 | | | | | |
| 1. 存在的问题 2. 研究目的不明确 3. 论文结构不恰当 4. 系统测试数据较少 5. 引文和网络引文的概念多余 | | | | | |
| 1. 拟采取的研究方法和可行性分析   拟采取系统分析和系统测试的研究方法。  这些研究方法适合实践性论文的写作，符合本文的特点，适用性强。 | | | | | |
| 1. 指导教师对劳动纪律、设计（论文）进展等方面的评语   指导教师签名： 年 月 日 | | | | | |

郑州大学信息管理学院答辩记录

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 学生姓名 |  | 学号 |  | 指导老师姓名 |  | 职称 |  |
| 题目 |  | | | | | | |
| 问题1：  答辩：  问题2：  答辩：  问题3：  答辩： | | | | | | | |

答辩委员会主席：

答辩委员： 、 、 、 、

20 年 月 日