**实验报告**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 姓名 | 安怡歌 吴佳慧 | | 学号 | 2021301041107 | 年级 | 2021级 |
| 成绩  （考核档次或分数） |  | | 指导教师 | 吴志强 | 专业 | 图书馆学 |
| 实验类型：  □独立实验课  R含实验的理论课  □计划外自选开放实验  □学生自主式开放实验  □大学生科技竞赛 | | | 实验日期：  20 24 年 2 学期，  第 7 周 | | 实验学时数：  学时 | |
| 相关课程：《数字图书馆原理与技术》 | | | | | | |
| 相关科研项目：无 | | | | | | |
| 实验名称 | | 实验二 数字资源的索引与检索实验 | | | | |
| 1. **预习部分（实验目的、实验基本原理等）**   实验原理：  1. 全文索引  数字图书馆存放的是规模化、非结构化的数字资源，通常使用的是针对这些资源内容的全文索引。全文索引的基本原理是：先定义一个词表，然后再资源文档中查找每个词条出现的频率和位置，把频率和位置信息按照词库的顺序进行归纳，这样就相当于对文档建立了一个以词表为目录的索引，当要查找某个词的时候就能快速定位到该词出现的位置  2. 信息检索  信息检索是图书馆为用户提供信息资源及其服务的前提。  经典的信息检索模型：   1. 布尔检索模型：是一个基于集合论和布尔代数的最简单的检索模型。查询则表示为索引项的布尔组合，用“AND、OR、NOT”连接起来，用“()”表示优先级；在检索匹配时，采用二值判定方式，一个文档当且仅当它能够满足布尔查询式时，它才能被检索出来。 2. 向量检索模型：是基于索引词词汇表V 的，文档d 和查询q 看成是索引词向量；检索过程就是计算文档向量与查询向量之间的相似度：可以根据相似度值的不同，对检索结果进行排序，还可以根据检索结果，做进一步的相关检索。   主要仪器设备（含必要的元器件、工具）：  实验目的：   1. 掌握数字图书馆数字资源的索引原理与方法。 2. 掌握数字图书馆数字资源的检索原理与方法。   实验内容：   1. 选择一个大型的搜索引擎（如百度），调研搜索引擎所使用的索引结构以及索引构建方法、所使用的检索模型以及检索流程。掌握数字图书馆全文索引、全文检索的相关原理、技术与基本流程。 2. 通过网络获取用于搜索引擎在构建索引以及检索时所使用的算法、工具、软件（或程序代码），重点分析如何利用它们的设计思想或者改造它们以用在数字图书馆的全文索引与检索中。 3. 通过对索引与检索基本原理的理解以及工具的调研，利用实验一“数字资源信息组织方法”的实验数据及结果，构建全文索引，并实现全文检索。 4. 设计UI界面，显示数字资源索引与检索实验过程中所涉及的流程、中间结果以及最终的实验结果。   实验要求：   1. 在调研基础上，分析描述数字资源全文索引与检索的基本流程 2. 分析索引、检索相关算法、工具或软件（程序代码）的基本功能，选择其中你认为适合的工具或代码作为本次实验的工具参考。 3. 利用实验一的实验数据及结果，完成大规模数字资源的全文索引与检索实验。 4. 总结在实验中遇到了哪些问题？如何解决的？本次实验有哪些收获？将上述问题及其心得体会总结成实验报告。实验报告以电子版方式，由学委收齐后，于本课程期末考试之前提交   目录  [1 理论知识 4](#_Toc163555359)  [1.1 大型搜索引擎——Google调研 4](#_Toc163555360)  [1.1.1 索引结构 4](#_Toc163555361)  [1.1.2 索引构建方法 4](#_Toc163555362)  [1.1.3 检索模型 4](#_Toc163555363)  [1.1.4 检索流程 4](#_Toc163555364)  [1.2 数字图书馆全文索引 4](#_Toc163555365)  [1.2.1 相关原理 4](#_Toc163555366)  [1.2.2 相关技术 4](#_Toc163555367)  [1.2.3 基本流程 4](#_Toc163555368)  [1.3 数字图书馆全文检索 4](#_Toc163555369)  [1.3.1 相关原理 4](#_Toc163555370)  [1.3.2 相关技术 4](#_Toc163555371)  [1.3.3 基本流程 4](#_Toc163555372)  [2 技术工具 5](#_Toc163555373)  [2.1 XXX 5](#_Toc163555374)  [2.2 XXX 5](#_Toc163555375)  [2.3 XXX 5](#_Toc163555376)  [2.4 XXX 5](#_Toc163555377)  [3 实验操作 5](#_Toc163555378)  [3.1 对分好类的测试集建立全文索引 5](#_Toc163555379)  [3.2 利用建立好的索引进行检索 5](#_Toc163555380)  [3.3 添加布尔检索 5](#_Toc163555381)  [4 UI设计 5](#_Toc163555382)  [4.1 制作可视化页面展示实验流程、中间结果及最终实验结果 5](#_Toc163555383) | | | | | | |

|  |  |
| --- | --- |
| **二、实验操作部分（可续页）**  *1. 选择一个大型的搜索引擎（如百度），调研搜索引擎所使用的索引结构以及索引构建方法、所使用的检索模型以及检索流程。掌握数字图书馆全文索引、全文检索的相关原理、技术与基本流程。* 1 理论知识1.1 大型搜索引擎——Google调研1.1.1 索引结构 Google：全球最大的搜索引擎之一，提供了广泛的搜索功能，包括网页搜索、图片搜索、新闻搜索、地图搜索等。  Google官方博客：https://blog.google/  Google技术文档：https://developers.google.com/search  Google Scholar：https://scholar.google.com/  ACM Digital Library：https://dl.acm.org/  IEEE Xplore：https://ieeexplore.ieee.org/  GitHub：https://github.com/ 1.1.2 索引构建方法1.1.3 检索模型1.1.4 检索流程1.2 数字图书馆全文索引1.2.1 相关原理1.2.2 相关技术1.2.3 基本流程1.3 数字图书馆全文检索1.3.1 相关原理1.3.2 相关技术1.3.3 基本流程 *2. 通过网络获取用于搜索引擎在构建索引以及检索时所使用的算法、工具、软件（或程序代码），重点分析如何利用它们的设计思想或者改造它们以用在数字图书馆的全文索引与检索中。* 2 技术工具2.1 XXX2.2 XXX2.3 XXX2.4 XXX *3. 通过对索引与检索基本原理的理解以及工具的调研，利用实验一“数字资源信息组织方法”的实验数据及结果，构建全文索引，并实现全文检索。* 3 实验操作3.1 对分好类的测试集建立全文索引3.2 利用建立好的索引进行检索3.3 添加布尔检索 *4. 设计UI界面，显示数字资源索引与检索实验过程中所涉及的流程、中间结果以及最终的实验结果。* 4 UI设计4.1 制作可视化页面展示实验流程、中间结果及最终实验结果 | |
| **三、实验效果分析（包括仪器设备等使用效果、实验完成情况）** | |
| **教师评语** |  |