**深 圳 大 学 实 验 报 告**

**课程名称： Java 程序设计**

**实验项目名称： 课程实验4：I/O和网络编程**

**学院： 计算机与软件学院**

**专业： 人工智能卓越班**

**指导教师： 潘微科**

**报告人： 邓瑞霖 学号： 2024150040 班级： 01**

**实验时间：2025年11月24日（周一）-2025年12月17日（周三）**

**实验报告提交时间： 2025年12月10日星期三**

**教务部制**

|  |
| --- |
| **实验目的与要求：**  **实验目的：**掌握I/O程序设计，能够读写本地文件等，掌握网络通信协议及相关程序设计。  **实验要求：**  **Part 1（30分）**  (1.1) 数据解析和统计。<https://snap.stanford.edu/data/web-Amazon.html> 网站上有很多Amazon的数据集供研究人员下载使用。本次实验使用Watches.txt.gz数据集，请下载后解压。格式说明请看网页上的“Data Format”部分。在报告中附上程序截图、运行结果（如每个输出文件前10行的截图等）和简要文字说明。  （i）使用Java语言读取解压后的文件（Watches.txt），并得到以下文件（10分）：  **review.txt**：每行2列，以分号作为分隔符，第1列是userID，第2列是productID，表示(user, product)二元组。该文件中不同行之间的顺序，按照userID从小到大排列，当userID相同时按照productID从小到大排列。  注：删除ID为unknown的记录；排序时按照字符串顺序。  （ii）使用Java语言根据review.txt进行计算，并得到以下文件（20分）：  **userNeighborhood.txt**：每行6列，以分号作为分隔符，第1列是userID，第2-6列是与该user最相似（根据相似度值）的5个user的userID，按相似度值从大到小排列（当相似度相同时，按照userID的字符串顺序从小到大排序），其中相似度是通过review.txt中的两列的信息计算得到的Jaccard index值。userNeighborhood.txt文件中不同行之间的顺序，按照第1列的userID从小到大排列。关于相似度的计算，要求使用多线程实现（5分）。  注：计算两个商品（product）之间的相似度（即Jaccard index）时，根据这两个商品所关联的用户的集合来计算——集合的交集的大小除以集合的并集的大小。计算两个用户（user）之间的相似度类似。  **Part 2（20分）**  (2.1) 实现句子中英互译（包括中译英和英译中）。要求使用DeepSeek、阿里通义、百度文心、腾讯混元或其他大模型API中的一个；  要求使用以下两个例句：  深圳大学1983年经教育部批准设立，肩负着为特区培养人才和为国家高等教育改革探路的光荣使命。  Shenzhen University was established in 1983 with the approval of the Ministry of Education, bearing the honorable mission of cultivating talent for the Special Economic Zone and pioneering reforms in China's higher education.  在报告中附上程序截图、运行结果和详细的文字说明。  **Part 3（30分）**  （3.1）利用套接字连接（TCP）编写一个简易在线聊天室程序。  该程序包括一个服务端（Server）和至少三个客户端（Client1, Client2, Client3）。所有客户端在启动后连接到服务端，并加入同一个公共聊天室。任何一个客户端发送一条文本消息，服务端在收到后，会将此消息转发给当前所有在线的其他客户端，从而实现群聊功能。具体要求如下：  服务端（Server）：   * 能够持续运行，监听指定端口，等待客户端连接。 * 能够处理多个客户端的连接请求，并为每个连接创建一个独立的线程。 * 当收到某个客户端发来的消息时，除了向发送者返回一个“消息已收到”的回执外，还需将此消息广播给所有其他在线的客户端。消息格式为[客户端名称]：消息内容。   客户端（Client）：   * 能够启动并连接到服务端。 * 具备两个线程：一个用于随时接收用户输入并发送给服务端，另一个用于持续接收并显示来自服务端的消息（包括其他客户端的发言和系统通知）。 * 客户端在发送消息后，能立即在本地屏幕上看到自己发送的消息，并在稍后收到服务端广播的同样消息。   报告要求：   * 附上系统架构示意图（展示一个服务端与多个客户端的连接及消息广播流向）。 * 提供关键的程序代码截图。 * 展示完整的运行结果截图（例如，展示三个客户端窗口和一个服务端窗口，模拟它们之间的多次对话）。 * 附上简要的文字说明，解释程序的工作流程和关键实现细节。   报告写作。要求：主要思路有明确的说明，重点代码有详细的注释，行文逻辑清晰可读性强，报告整体写作较为专业。（20分）  **说明：**  （1）本次实验课作业满分为100分，占总成绩的比例7%。  （2）本次实验课作业截至时间2025年12月17日（周三）21:59。  （3）报告正文：请在指定位置填写，本次实验不需要单独提交源程序文件。  （4）个人信息：WORD文件名中的“姓名”、“学号”，请改为你的姓名和学号；实验报告的首页，请准确填写“学院”、“专业”、“报告人”、“学号”、“班级”、“实验报告提交时间”等信息。  （5）提交方式：截至时间前，请在Blackboard平台中提交。  （6）发现抄袭（包括复制&粘贴整句话、整张图），**抄袭者和被抄袭者的成绩记零分。**  （7）延迟提交，不得分；如有特殊情况，请于截至日期之后的48小时内发邮件到panweike@szu.edu.cn，并在邮件中注明课程名称、作业名称、姓名、学号等信息，以及特殊情况的说明，我收到后会及时回复。  （8）期末考试阶段补交无效。 |

|  |
| --- |
| **Part 1（25分）**  (1.1) 数据解析和统计。<https://snap.stanford.edu/data/web-Amazon.html> 网站上有很多Amazon的数据集供研究人员下载使用。本次实验使用Watches.txt.gz数据集，请下载后解压。格式说明请看网页上的“Data Format”部分。在报告中附上程序截图、运行结果（如每个输出文件前10行的截图等）和简要文字说明。r  （i）使用Java语言读取解压后的文件（Watches.txt），并得到以下文件（10分）：  **review.txt**：每行2列，以分号作为分隔符，第1列是userID，第2列是productID，表示(user, product)二元组。该文件中不同行之间的顺序，按照userID从小到大排列，当userID相同时按照productID从小到大排列。  注：删除ID为unknown的记录；排序时按照字符串顺序。  （ii）使用Java语言根据review.txt进行计算，并得到以下文件（20分）：  **userNeighborhood.txt**：每行6列，以分号作为分隔符，第1列是userID，第2-6列是与该user最相似（根据相似度值）的5个user的userID，按相似度值从大到小排列（当相似度相同时，按照userID的字符串顺序从小到大排序），其中相似度是通过review.txt中的两列的信息计算得到的Jaccard index值。userNeighborhood.txt文件中不同行之间的顺序，按照第1列的userID从小到大排列。关于相似度的计算，要求使用多线程实现（5分）。  注：计算两个商品（product）之间的相似度（即Jaccard index）时，根据这两个商品所关联的用户的集合来计算——集合的交集的大小除以集合的并集的大小。计算两个用户（user）之间的相似度类似。   1. 首先读取文件Watches.txt和创建review.txt文件：   并创建List<TreeMap<String, String>> list存储每一位的productID和userID    创建两个标识符用于定位productId和userId：    使用Scanner解析文件      将list中的值进行排序，先按userId的大小排序，再按照productId的大小进行排序    写入文件：    得到的文件（部分）：    ii)读取review.txt中的内容    计算Jaccard相似度    采用多线程计算每个user的top-5相似用户      main函数：    运行后得到的文件userNeighborhood.txt（部分）    **Part 2（25分）**  (2.1) 实现句子中英互译（包括中译英和英译中）。要求使用DeepSeek、阿里通义、百度文心、腾讯混元或其他大模型API中的一个；  要求使用以下两个例句：  深圳大学1983年经教育部批准设立，肩负着为特区培养人才和为国家高等教育改革探路的光荣使命。  Shenzhen University was established in 1983 with the approval of the Ministry of Education, bearing the honorable mission of cultivating talent for the Special Economic Zone and pioneering reforms in China's higher education.  在报告中附上程序截图、运行结果和详细的文字说明。  选择阿里通义，在阿里通义中获得API Key    翻译部分：      json转义：    main函数：    输出：  在命令行中先输入javac Class/CH04/TongyiTranslateExample.java编译文件 在输入java Class.CH04.TongyiTranslateExample运行文件    **Part 3（30分）**  （3.1）利用套接字连接（TCP）编写一个简易在线聊天室程序。  该程序包括一个服务端（Server）和至少三个客户端（Client1, Client2, Client3）。所有客户端在启动后连接到服务端，并加入同一个公共聊天室。任何一个客户端发送一条文本消息，服务端在收到后，会将此消息转发给当前所有在线的其他客户端，从而实现群聊功能。具体要求如下：  服务端（Server）：   * 能够持续运行，监听指定端口，等待客户端连接。 * 能够处理多个客户端的连接请求，并为每个连接创建一个独立的线程。 * 当收到某个客户端发来的消息时，除了向发送者返回一个“消息已收到”的回执外，还需将此消息广播给所有其他在线的客户端。消息格式为[客户端名称]：消息内容。   客户端（Client）：   * 能够启动并连接到服务端。 * 具备两个线程：一个用于随时接收用户输入并发送给服务端，另一个用于持续接收并显示来自服务端的消息（包括其他客户端的发言和系统通知）。 * 客户端在发送消息后，能立即在本地屏幕上看到自己发送的消息，并在稍后收到服务端广播的同样消息。   报告要求：   * 附上系统架构示意图（展示一个服务端与多个客户端的连接及消息广播流向）。 * 提供关键的程序代码截图。 * 展示完整的运行结果截图（例如，展示三个客户端窗口和一个服务端窗口，模拟它们之间的多次对话）。 * 附上简要的文字说明，解释程序的工作流程和关键实现细节。   服务端（Service）代码：              Client1代码：          Client2:    Client3:    在命令框中编译并运行服务端和客户端：得到的结果如下：    模拟进行对话：    ++++++++++++++++++++++++++++++++++++++++++++++++++++++  **其他（例如感想、建议等等）。**  实验感想：  通过本次Java程序设计实验，我深刻体会到了理论与实践相结合的重要性，对I/O操作、网络编程以及API调用有了更加深入的理解。  技术收获与体会  Part 1：数据解析与统计  在处理Amazon数据集的过程中，我认识到大数据处理中数据清洗和排序的重要性。通过实现Jaccard相似度计算，不仅巩固了集合运算的知识，还学会了如何利用多线程优化计算密集型任务。特别是在处理用户-商品关系网络时，我体会到了算法效率对实际应用的关键影响。  Part 2：API调用与翻译服务  这部分实验让我接触到了现代软件开发中常见的API集成模式。通过调用阿里通义大模型API，我学会了HTTP请求的构造、JSON数据的处理以及异常处理机制。更重要的是，我认识到在AI时代，如何有效利用第三方服务来增强应用功能是每个开发者必备的技能。  Part 3：网络编程与实时通信  聊天室程序的实现让我对TCP/IP协议、多线程编程有了直观的认识。从简单的客户端-服务器模型到复杂的多用户实时交互，我体会到了网络编程中并发控制、资源管理和异常处理的重要性。这个项目让我真正理解了"网络即计算机"的含义。  遇到的挑战与解决方案：  1. 数据格式处理：在Part 1中，最初对数据格式理解不够准确，通过仔细阅读文档和多次调试才正确解析。  2. 多线程同步：在Part 3中，客户端列表的线程安全问题通过使用Collections.synchronizedList得到解决。  3. API集成：Part 2中API密钥管理和请求超时处理花费了不少时间，最终通过完善的错误处理机制实现稳定运行。 |

深圳大学学生实验报告用纸

|  |
| --- |
| 指导教师批阅意见：  成绩评定：  指导教师签字：  2025年 月 日 |
| 备注： |

注：1、报告内的项目或内容设置，可根据实际情况加以调整和补充。

2、教师批改学生实验报告时间应在学生提交实验报告时间后10日内。