



**综 合 实 验 报 告**

课程编号： C01067

课程名称： JAVA应用系统设计

学 号： 31701293

姓 名： 王文海

专业班级： 软件工程1702

所在学院： 计算学院

报告日期： 2019 年 12 月 30 日

**一、综合实践的主要目的：**

1. 进一步巩固和理解所学的专业基本理论知识和技术，培养学生运用所学的知识和技术能力，能够掌握软件应用系统的开发过程。
2. 通过实践，解决复杂工程问题。
3. 能够合理地组织数据、有效地存储和处理数据，正确地算法设计及进行算法分析和评价。
4. 能够针对所面对的复杂工程问题结合解决方案，选择恰当的工具，并且能够针对问题的解决过程有效的规划工具的协作方式

**二、实践主要内容：**

通过课程大作业实践，我完成了作品“基于jieba分词的主观题打分系统”，作品主要实现了对主观题问题答案和作答者回答进行在分词后，进行词频统计和向量化，最终根据向量间的余弦夹角值求出其相似度，最终得出作答者获得的分数。

**需求分析：**

1. **详细设计**

随着职业教育越来越鼓励学生对知识的综合运用，而非单纯的知识水平考试， 客观题的考察占比越来越低，方案设计等主观题的考察占比越来越高。传统的考 评软件，仅支持客观题自动评分，主观题仍需要老师人工打分。在教学智能化日 益提升的今天，函需一款软件将老师从批改试卷、作业这类重复性工作中完全解 放出来。

第一阶段：概要介绍文档

第二阶段：算法分工

第三阶段：项目详细方案

第四阶段：功能以及总体功能的结合

第五阶段：PPT制作

主要工具：jupyter，jieba库，pandas库

主要算法：余弦相似度

具体算法实现过程为：

1、分词

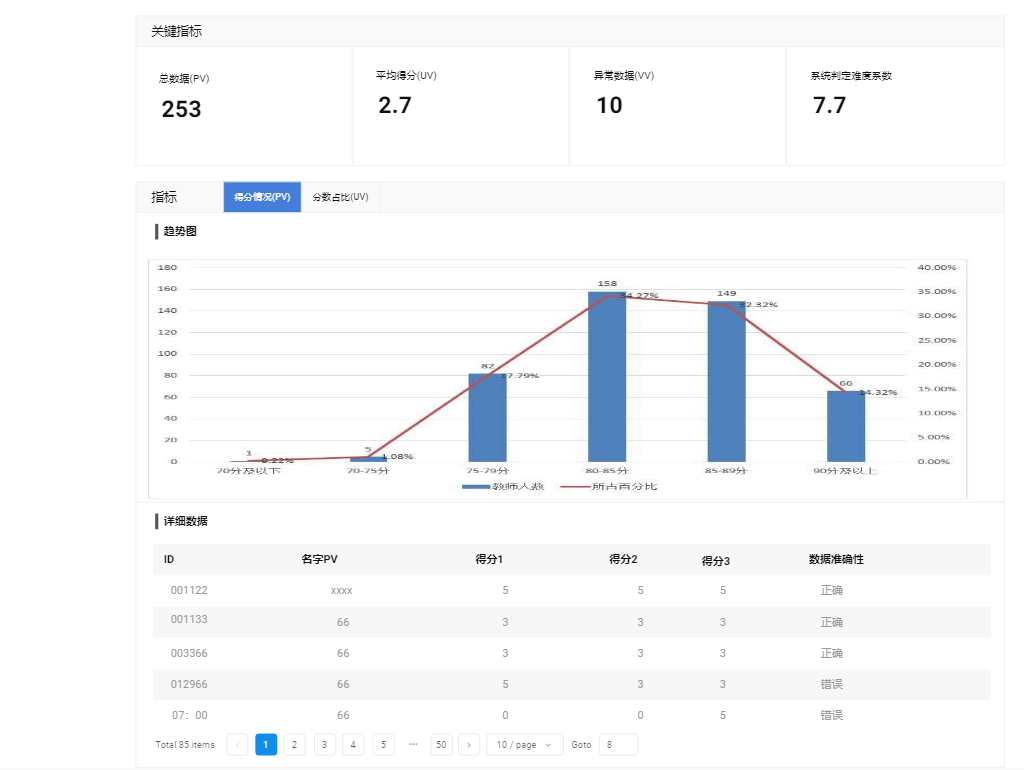
2、列出所有词，并按顺序进行标号

3、按出现顺序对两个想要比较的字符串中的每个分词进行标号匹配

4、对每个分词进行独热编码，统计出现次数

5、余弦相似度计算得出结果

1. **界面设计**

****

****

1. **应用开发关键技术**

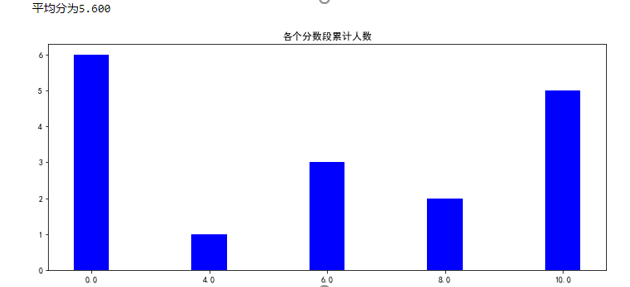
（1）能够阐明自己设计的系统的评分机制；

（2）考虑算法的效率，包括时间和空间性能；

（3）针对测试样例（部分提供给学生），统计参赛人员设计的算法得到的 学生各题分数与教师提供分数之间的绝对差值，所有学生的各题分数绝对差值最 小者为优。测试为开放性测试，不仅限于提供的样例数据。

**三、实践过程中的创新点**

在执行程序并且获得分数统计后，系统会对已有的数据进行分析，绘制图标，进行数据的可视化处理，便于使用者获取一些关键信息。



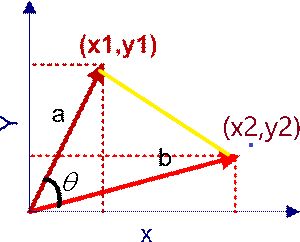
**四、能够合理地组织数据、有效地存储和处理数据，正确地算法设计及进行算法分析和评价。**

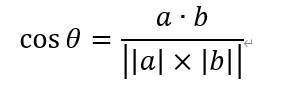
主要通过excle储存和处理数据，并且在程序执行结束后会对excle进行写操作并保存。



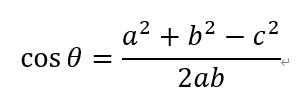
核心算法：余弦相似度

余弦相似度用向量空间中两个向量夹角的余弦值作为衡量两个个体间差异的大小。余弦值越接近1，就表明夹角越接近0度，也就是两个向量越相似，这就叫"余弦相似性"。

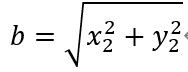
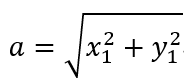




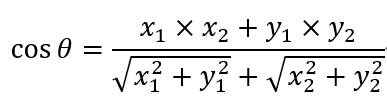
根据三边长度求余弦值：



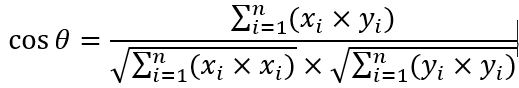
可知三边长度为：



代入计算可得：



扩展至N维空间向量中：



具体算法实现过程为：

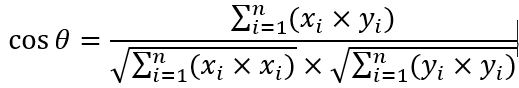
1、分词

2、列出所有词，并按顺序进行标号

3、按出现顺序对两个想要比较的字符串中的每个分词进行标号匹配

4、对每个分词进行独热编码，统计出现次数

5、余弦相似度计算得出结果



复杂度：



**五、项目总结。**

一、项目基本情况：

在这几个星期，项目的所有组员团结合作，每一个文档，PPT，算法的分析和解决，都有每一个小组成员日夜奋斗的身影。最后项目的算法，概要文档，详细文档，项目介绍PPT都已顺利完成，小组中每一个成员都功不可没。这次参赛对我们小组的每一个成员来说意义都十分重大，小组中有些是经常参加各种比赛的，经验丰富的成员，也有一些是第一次接触团队合作完成项目的成员。对于经验丰富的成员来说，这一次项目无疑是一个增加阅历的机会。而对于第一参加这种团队项目的成员来说，这次项目的意义更为重大。首先，他们体会到了小组合作完成项目的每一个流程，对一个项目需要干什么，自己要如何融入到一个项目中都有了一定的了解，更重要的是，这一次阅历对于他们来说更为重要，这也会是他们团队合作的最重要的一步。

二、建设中的工作情况（最好给每一个小标题都起一个煽情的名字）

首先，我们结合了四个人的理解，选择了一个最适合每一个人的选题，虽然这是最早的一步，但是考虑到小组人员的能力情况，我们给了足够多的时间让每一个人都能理解每一个项目的考察方向，从而做出最优的选择。

第二步，我们开始着手概要简介，我们在进行分工后，每个人对自己的任务都进行了制作并出色完成任务。

第三步，我们开始着手算法的实现，这一步是我们小组重点讨论的部分，在明确了业务要求之后，我们更具每一个成员的能力，做了大概的分工。在实现过程中，有成员遇到了代码方面的问题，只要有成员提出了相应的问题，组员中的所有人东欧会对他代码中的问题进行解答。这一个过程在整个环节中虽然占了最多的时间，但我认为，这也是做能体现出一个团队的的各步骤，成员们都互帮互助，一起谈论最佳的解决方案。最后，虽然有超时，但还是完成了代码的编写。

第四步，我们开始着手项目详细方案，我们在进行分工后，每个人对自己的任务都进行了制作并出色完成任务。

第五步，我们开始着手概要简介PPT，我们在进行分工后，每个人对自己的任务都进行了制作并出色完成任务。

四、存在的不足：在整个项目的基础上，每一个成员都完成了任务，但是在细节方面，对于我们所列出的项目计划甘特图来看，有一些成员并没有按时完成进度，导致整体的项目都向后推移。主要原因是由小组成员是第一次接触项目，对项目的流程并没有很强的意识。对于这个问题，我想在经历了这一次的项目后，所有成员都应该深刻的印象，并在以后的项目中引以为戒。

五、几点体会：在这几个星期，项目的所有组员团结合作，每一个文档，PPT，算法的分析和解决，都有每一个小组成员日夜奋斗的身影。最后项目的算法，概要文档，详细文档，项目介绍PPT都已顺利完成，小组中每一个成员都功不可没。这次参赛对我们小组的每一个成员来说意义都十分重大，小组中有些是经常参加各种比赛的，经验丰富的成员，也有一些是第一次接触团队合作完成项目的成员。对于经验丰富的成员来说，这一次项目无疑是一个增加阅历的机会。而对于第一参加这种团队项目的成员来说，这次项目的意义更为重大。首先，他们体会到了小组合作完成项目的每一个流程，对一个项目需要干什么，自己要如何融入到一个项目中都有了一定的了解，更重要的是，这一次阅历对于他们来说更为重要，这也会是他们团队合作的最重要的一步。

三、项目改进计划：

在项目功能上，在一些否定词，情感词以及双关词的判断上仍存在一定的缺陷，在接下来的工作中我们会通过一系列机器学习与NLP文本处理的方法去改进和弥补这些缺陷。

数据分析也是未来会进行重点研究的内容，在数据集逐渐完整的情况下，我们会针对一些数据特征进行数据可视化以及数据的分析，最后在web端呈现给用户。