软工实验数据分析报告

对本次实验数据进行整理，力求从中发现有价值的信息，从实验小组组员和实验本身两个不同角度出发，进行如下探索：

1 本组工作及github提交习惯分析

数据来源：github提交记录

数据可视化：如图1-1.

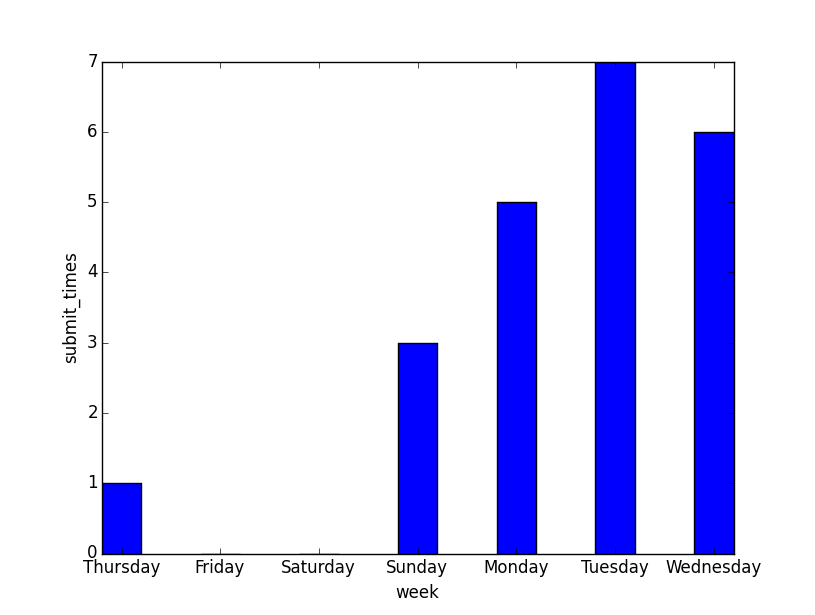


图1-1

数据解释：从github中获取了所有提交记录及对应的日期，将日期转化为一周7

天，图1-1中横轴表示一周7天，纵轴表示对应提交总次数。

数据分析：从柱状图可以容易看出本组github提交行为几乎全部发生周日到周三，

在全部提交22次记录中，有21次是在周日到周三，其中周二最高，有7次提交记录。

由于本组组员只要工作，肯定会向github提交工作内容，所以github提交记录也就可

以反映工作情况。从图中就可以直观地看出本组组员工作具有明显的拖延习惯，一周工

作只在下半周才会开始赶进度。

2 本次实验各阶段相关度分析

数据来源：github，mpp文件

数据处理：

（1）实验阶段划分。

按照mpp中阶段及github上提交物将实验划分为调研阶段、需求阶段和测试阶段。

（2）特征提取

考虑到不同阶段产出物在很大程度上说明了该阶段的主要特点，提取特征如下：产

出物是否含有文档报告、产出物是否含有规格说明、产出物是否含有代码文件，产出物

是否含有工程文件。按照课程设置提取特征：是否需要进行评审。

根据特征进行数据整理，如表2-1 。



表2-1

经过特征抽象，将各阶段用向量表示。使用余弦相似度来计算各阶段之间的相似程度。首先简单介绍余弦相似度度量，将两个阶段用对应的向量和表示，其相似度可以定义如下：

符号表示向量点积，表示向量的欧式长度，即向量自身点积的平方根。

用向量表示调研阶段、需求阶段和测试阶段，其中 = （1，0，0，1，0）， = （1，1，0，1，1）， = （1，1，1，0，1）。

计算各阶段之间的相似度：

由此分析可知，本次实验中相似度最高的两个阶段为需求阶段和测试阶段，相似度最低的是调研阶段和测试阶段。

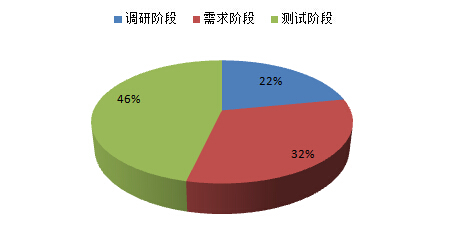
3 本次实验各阶段比重分析

数据来源：mpp文件

数据处理：将mpp中阶段合并划分为调研阶段、需求阶段和测试阶段。考察各阶段

所花时间对比。

数据可视化：如图3-1 。



数据分析：从各阶段比例图可以容易发现，本次实验测试阶段花了将近50%时间，

回顾本次实验，在测试阶段由于要开始动手实践（包括搭建测试环境，设计测试用例以

及编写测试脚本），并且在测试过程中发现需求阶段存在的一些问题，所以导致部分测

试要抛开需求阶段提供的依据，使得时间开销增大。本组认为如果需求阶段可以适当花

费更多的时间，把需求做的更加细致准确的话，测试阶段有了可靠的依据，会大大降低

测试阶段的时间开销。

4 本次实验各阶段产出物分析

数据来源：github提交历史文件

数据处理：将mpp中阶段合并划分为调研阶段、需求阶段和测试阶段。考察各阶段

提交至github的文档报告、工程文件、代码文件总数。

数据可视化：如图4-1.

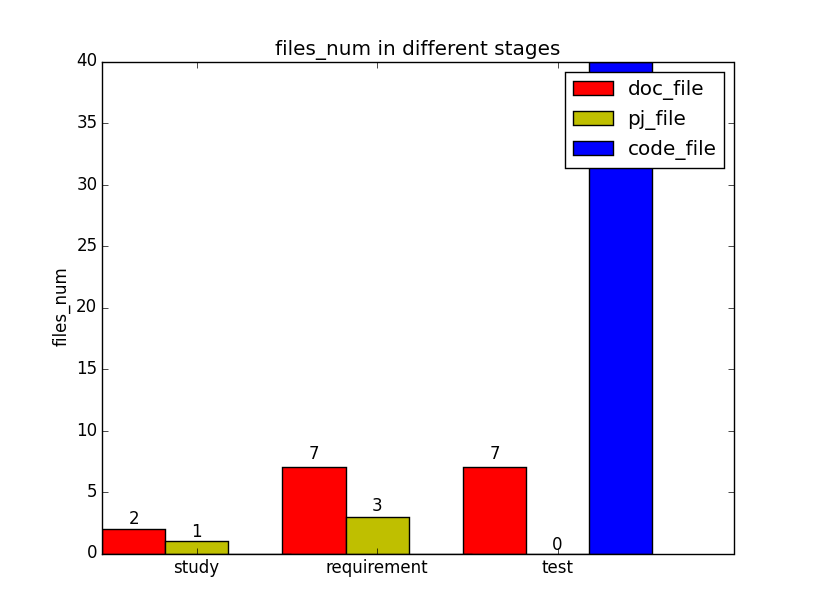


图4-1

数据分析：由图4-1可知，调研阶段产出物为文档报告及工程文件，这里的工程文件为mpp文件；需求阶段产出物为文档报告及工程文件，这里的工程文件包括模型文件和mpp文件；测试阶段产出物为文档报告和代码文件，代码文件即为测试脚本文件。由此数据可知，本次实验工作量主要集中在需求、测试阶段，其中需求阶段主要工作为撰写文档报告及模型文件，测试阶段主要工作为撰写文档报告及编写测试脚本。