**基于Lucene的分析与应用**

**统计分析**

Version 1.0

小组成员：

刘宏宇

滕延林

顾泽鹏

杨帆

周晓懿

版本变更历史

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本 | 提交日期 | 主要编制人 | 审核人 | 版本说明 |
| 1.1.1 | 20160503 | 顾泽鹏、刘宏宇 | 滕延林、杨帆、周晓懿 | 统计分析说明书v01 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# 目的

为了便于协调组内成员进行后期的工作，对项目进行跟踪和监控，对任务的进度进行安排与调控，故对后期工作进行计划。提出需求，指导后续工作。建立任务追踪链，记录项目计划于需求的变化。反映项目进展情况，明确各个成员的任务，并记录各个成员的贡献。

# 版本变更

**第一版变更：**

**变更日期：**2016-03-16至2016-03-20

**变更版本编号：**1

**变更请求：**对项目计划书的第一次修改

**变更范围：**明确项目计划书内容、修改排版问题

**评审意见：**

1、 数据的获取问题

接受

解决方案：通过爬虫爬取一定数量的非结构化数据（文本、图像或音频）构造测试数据集供测试使用，初定数据规模在1W左右

2、 上层程序的测试问题，

接受

解决方案：

衡量标准：检索时间和检索准确性（准确率、召回率）

具体实施：（1）部署原始lucene架构，从测试集中选取测试数据，在原始架构上使用测试数据作为query，将检索结果以及运行时间作为groundtruth。（2）通过运行改进后的程序，使用相同的测试数据作为query，将检索结果以及运行时间与groundtruth进行对比从而获得改进效果。

**变更的详细描述：**

* 明确编写该文档的目的
* 丰富了术语与缩略语
* 减少了部分遵循的标准
* 修改了项目分工中的用词
* 修改了交付产品的名称，使之更加标准化
* 添加了待定问题与解决方案
* 统一了文字排版，图表的标号

**因果分析：**

第一版的项目计划书的目的不够明确，使得小组成员对于后续工作不明确，所以加以修改；术语与缩略语较少，不能涵盖所有项目中所涉及的概念；许多的标准之间互相矛盾，所以减少了标准数量；将“阅读代码”改为“研读代码”；根据老师的PPT中内容作出相应修改；提出问题，使目标更加明确；统一排版。

**第二版变更**：

**变更日期：**2016-03-16至2016-03-20

**变更版本编号：**2

**变更请求：**对软件需求文档的第一次修改

**变更范围：**对软件需求文档中用例图修改，文档结构调整

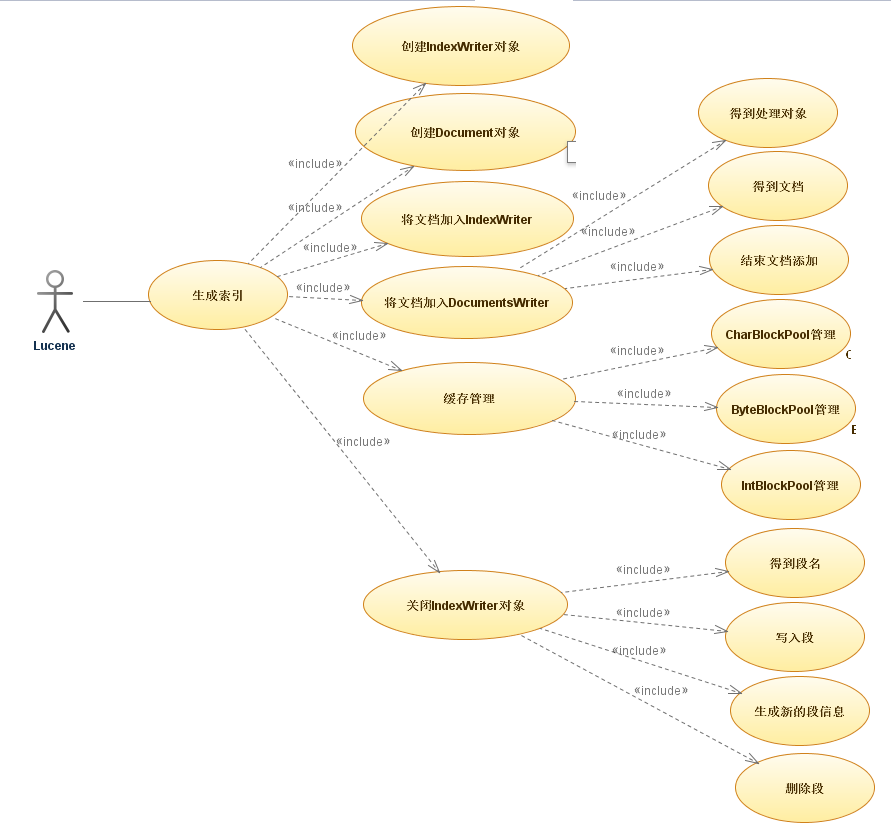
**评审意见**：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 评审对象 | 评审意见 | 处理建议 | 小组意见 |
| 1 | 《基于 Lucene 的分析与应用需求规格说明书》 | 上层应用的检索效率问题是如何考量的 | 我们已经在第一次的报告中回答了这些问题，我们会将具体的细节呈现在文档中 | 接受 |
| 2 | 《基于 Lucene 的分析与应用需求规格说明书》 | 文档中存在一些语言上的小错误 | 我们将改正错误并将改好的文档提交 | 接受 |
| 3 | 《基于 Lucene 的分析与应用需求规格说明书》 | 用例图中的Actor的概念没有理解正确 | 我们会重新修改并将改好的文档提交 | 接受 |

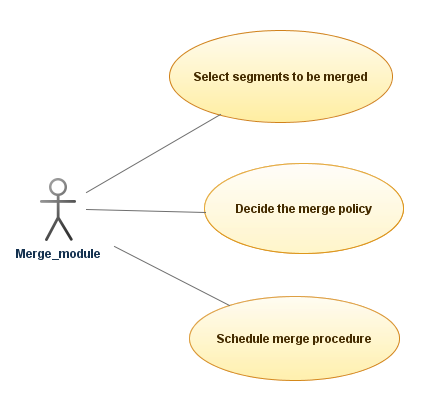
**变更的详细描述：**

* 对于功能需求的修改
* 将非功能需求单独作为一个章节
* 用例图修改

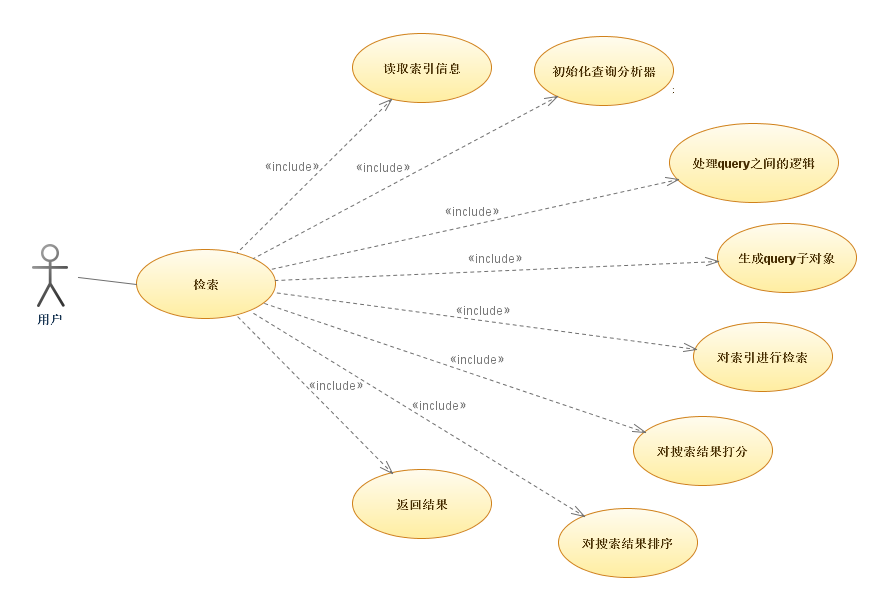
**生成索引用例图**



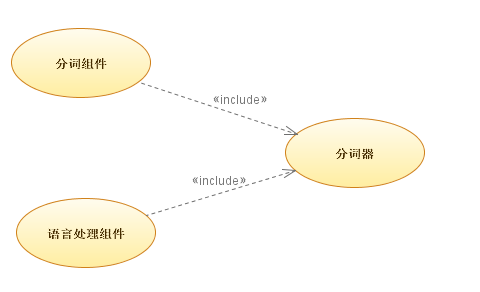
**段合并用例图**



**检索用例图**



**分词器用例图**



**因果分析：**

功能模块不明确，部分功能相互牵连；考虑到非功能需求为性能上要求，属于提高需求，并且不能以用例图的形式表示；对于部分用例图做出了修改，更加准确描述系统功能。

**第三版变更：**

**变更日期：**2016-03-20至2016-04-03

**变更版本编号：**3

**变更请求：**对软件需求文档的第二次修改

**变更范围：**修改软件需求文档的用例图，添加数据需求以及运行需求。

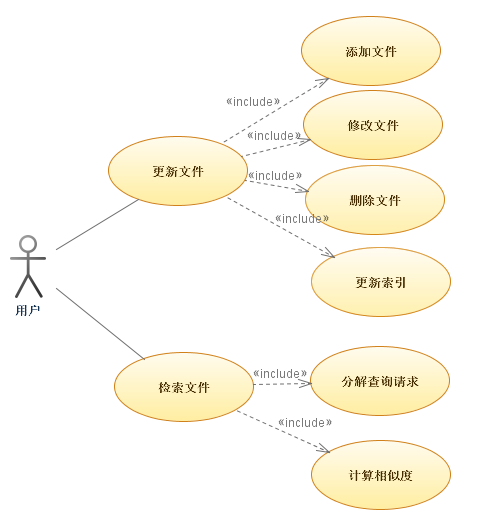
**评审意见**：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 评审对象 | 评审意见 | 处理建议 | 小组意见 |
| 1 | 《基于 Lucene 的分析与应用需求规格说明书V2.0》 | 需求的主要目的是作为设计开发的依据，因此目的部分需要调整。 | 我们将更正问题并提交。 | 接受 |
| 2 | 《基于 Lucene 的分析与应用需求规格说明书V2.0》 | 对于“将业务需求分解为功能性需求”： 需求和功能应是两个概念，要注意区分业务需求和功能需求的区别。“分解”的说法不确切，应是“转化”的关系。 | 我们将更正问题并提交。 | 接受 |
| 3 | 《基于 Lucene 的分析与应用需求规格说明书V2.0》 | 在用例图中，系统是否应该作为Actor？因为功能不是用例图应表达的事。 | 我们将对Lucene用例图中Actor的定位重新考量。 | 接受 |
| 4 | 《基于 Lucene 的分析与应用需求规格说明书V2.0》 | 用例图中的用例和RUCM的用例是有关系的。部分用例包含和模板里的包含关系对应不上，追踪性有问题。 | 我们将校验并更正用例图和RUCM中的用例，完善可追踪性。 | 接受 |
| 5 | 《基于 Lucene 的分析与应用需求规格说明书V2.0》 | 数据需求不应是测试需求，而应是开发所考虑的需求。（例如考虑文本特点、是否有图片等等） | 我们将更正问题并提交。 | 接受 |
| 6 | 《基于 Lucene 的分析与应用需求规格说明书V2.0》 | 解释自主构造一词。 | 我们将完善该部分。 | 接受 |

**变更的详细描述：**

* 大幅度调整用例图部分
* 添加数据需求
* 添加运行需求

**用例图**



**因果分析：**

原文档中的用例图有比较大的问题，例如“分词器”模块对于用户不可见，并不是用户的行为，不能以用例图来表示。另外在原用例图中出现了“系统”这样Actor，自己调用自己显然不符合逻辑，最终只保留一个用例图；该系统处理文件系统，对于数据有一定需求；该系统的运行环境的要求，包括软件环境与硬件环境。

**第四版变更：**

**变更日期：**2016-04-03至2016-04-09

**变更版本编号：**4

**变更请求：**对软件需求文档的第三次修改

**变更范围：**修改RUCM图、修改排版问题

**变更的详细描述：**

* 修改RUCM图，使之与用例图匹配
* 统一了文字排版，图表的标号

**更新文件**



**添加文件**



**修改文件**



**删除文件**



**更新索引**



**检索文件**



**分解查询请求**



**计算相似度**



**因果分析：**

RUCM图与用例图不符合，有些用例不能定位；统一排版。

# 影响因素分析

## 外在因素

#### 评审意见：包括老师和同学在格式和内容上提出的意见

评审意见贯穿了项目进行的始终，老师和同学们的评审意见开阔了我们的视野，发现了许多我们小组内部讨论时没有想到的地方，帮助我们不断完善我们的项目需求、计划并确保我们在项目进行过程中没有偏离方向。

## 内在因素

#### 对系统的了解程度：

在项目刚开始的时候，对项目的熟悉程度较低，只是大致了解Lucene开源系统是做什么的，有哪些模块，根据这些模块倒推项目的需求分析。

随着对系统了解程度的加深，为了做出实质性的改进，我们对系统具体的使用场景以及数据样例的特点进行了分析，根据不同的使用场景以及数据样例的特点，有针对性地对需求进行了改进，进而变更了相应的需求规格说明书以及任务计划。

# 相关性分析

为了研究版本变更与影响因素之间的相关性，我们通过分析版本变更追踪链，发现第一至第四版本变更主要是受到外在因素的影响，而第五和第六版本变更中，在外在因素始终发挥着重要作用的基础上，内在因素发挥着越来越重要的作用。

# 5.基础统计分析

**工作量统计**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 实验名称 | 成员 | 工作评估 | 工作量统计数据 | 工作量 |
|
| 需求测试分析 | 刘宏宇 | 确定需求 | 需求个数 | 5 |
| 确定用例 | 用例个数 | 0 |
| 撰写报告 | 报告字数 | 1312 |
| 顾泽鹏 | 确定需求 | 需求个数 | 5 |
| 确定用例 | 用例个数 | 0 |
| 撰写报告 | 报告字数 | 1256 |
| 滕延林 | 确定需求 | 需求个数 | 0 |
| 确定用例 | 用例个数 | 4 |
| 撰写报告 | 报告字数 | 1466 |
| 杨帆 | 确定需求 | 需求个数 | 0 |
| 确定用例 | 用例个数 | 4 |
| 撰写报告 | 报告字数 | 1503 |
| 周晓懿 | 确定需求 | 需求个数 |  |
| 确定用例 | 用例个数 |  |
| 撰写报告 | 报告字数 |  |

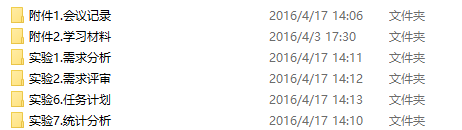
**工作难度统计**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 实验名称 | 成员 | 系数 | 阶段性数据变更 | | |
| 是否变更 | 变更原因 | 变更工作量 |
| 需求测试分析 | 刘宏宇 | 难度 | 1 | 老师课堂评价 | 2 |
| 1 | 4 |
| 1 | 672 |
| 顾泽鹏 | 难度 | 1 | 课堂评审意见 | 6 |
| 1 | 3 |
| 1 | 305 |
| 滕延林 | 难度 | 1 | 老师课堂评审 | 3 |
| 1 | 同学网评 | 5 |
| 1 |  | 218 |
| 杨帆 | 难度 |  |  |  |
|  |  |  |
|  |  |  |
| 周晓懿 | 难度 |  |  |  |
|  |  |  |
|  |  |  |

# 6.配置管理总结

为了便于对项目变更进行管控（版本控制），维护不同项目之间的版本关联，以及后期追溯软件开发过程进行总结分析，我们使用了GitHub作为基础配置管理工具。

1. 为了使项目文件具有合理的组织框架，我们通过讨论制定了文档分类组织结 构，依据各子实验进行组织分类。



1. 为了防止版本变更过程中出现命名混乱，我们通过开会讨论统一了本次软件工程综合实验文档的命名方法。
2. 为了便于追溯更新内容及更新缘由，我们制定了版本变更记录表，每次文档版本更新时，需要附上版本更新记录，记录中的字段依次为：版本号、变更时间、修改人、审核人、版本变更原因以及主要变更内容。



1. 为了确保项目文件完整性及安全性，我们采用了冗余存储的方式，除了将项目文件上传到Github上，还做好了备份工作，每次的项目文件小组成员分别存有备份。