



SHANGHAI MARITIME UNIVERSITY

**Reading Comprehension Question Answering System.**

Instructor : Liu Jin

Course Name : Software Project Manager

Student Name: Cai Shipeng 202230310226

Student Name: Shi Zhouyin 202230310210

Student Name: Liu Haoyang 202230310230

Student name: Xu Runxuan 202230310211

Date ： 2023.05.31

目录

[1.Preface 5](#_Toc136449343)

[1.1 Overview 5](#_Toc136449344)

[1.2 Project delivery products 5](#_Toc136449345)

[2. Product introduction 5](#_Toc136449346)

[2.1 Introduction to project functions 5](#_Toc136449347)

[2.1.1 Project core functionality 5](#_Toc136449348)

[2.1.2 Other features of the project 5](#_Toc136449349)

[2.2 Technical Introduction 6](#_Toc136449350)

[2.2.1 Summary extraction 6](#_Toc136449351)

[2.2.2 Machine reading comprehension 6](#_Toc136449352)

[2.2.3 Image to text function 6](#_Toc136449353)

[3. Project planning 6](#_Toc136449354)

[3.1 Project organization 6](#_Toc136449355)

[3.2 Project scheduling 7](#_Toc136449356)

[3.3 Gantt char 9](#_Toc136449357)

[4.Risk Plan 10](#_Toc136449358)

[4.1 Introduction 10](#_Toc136449359)

[4.2 Project Risk Indetification 10](#_Toc136449360)

[4.3 Project Risk assessment 10](#_Toc136449361)

[4.3.1 Quantifying risk probability 11](#_Toc136449362)

[4.3.2 Quantifying impact 11](#_Toc136449363)

[4.3.3 probability and impact matrix 12](#_Toc136449364)

[4.3.4 assessing the Risks 12](#_Toc136449365)

[4.4 Risk Budget 13](#_Toc136449366)

[5. Test Plan&&Test Case 15](#_Toc136449367)

[5.1 schedule 15](#_Toc136449368)

[1.Test 1 15](#_Toc136449369)

[2.Test 2 15](#_Toc136449370)

[3.Test 3 15](#_Toc136449371)

[4.Test 4 16](#_Toc136449372)

[5.Test 5 16](#_Toc136449373)

[6.Test 6 16](#_Toc136449374)

[5.2 design 16](#_Toc136449375)

[5.2.1 Test 1 16](#_Toc136449376)

[1. Control 17](#_Toc136449377)

[2. pass standard 17](#_Toc136449378)

[3. param explanation 17](#_Toc136449379)

[4. Input 17](#_Toc136449380)

[5. Test Result 18](#_Toc136449381)

[5.2.2 Test 2 18](#_Toc136449382)

[1. Control 18](#_Toc136449383)

[2. pass standard 18](#_Toc136449384)

[3. param explanation 18](#_Toc136449385)

[4. Input 18](#_Toc136449386)

[5.2.3 Test 3 19](#_Toc136449387)

[1. Control 19](#_Toc136449388)

[2. pass standard 19](#_Toc136449389)

[3. Input 19](#_Toc136449390)

[4. Test Result 20](#_Toc136449391)

[5.2.4 Test 4 21](#_Toc136449392)

[1. Control 21](#_Toc136449393)

[2. pass standard 21](#_Toc136449394)

[3. Input 21](#_Toc136449395)

[4. Test Result 22](#_Toc136449396)

[5.2.5 Test 5 23](#_Toc136449397)

[1. Control 23](#_Toc136449398)

[2. pass standard 23](#_Toc136449399)

[3. Input 23](#_Toc136449400)

[4. Test Result 24](#_Toc136449401)

[5.2.6 Test 6 25](#_Toc136449402)

[1. Control 25](#_Toc136449403)

[2. pass standard 25](#_Toc136449404)

[3. Input 25](#_Toc136449405)

[4. Test Result 25](#_Toc136449406)

[6.Project Execution Details 26](#_Toc136449407)

[6.1 Project Execution Details Table 26](#_Toc136449408)

[6.2 Project completion and time cost 30](#_Toc136449409)

[6.2.1 project completion 30](#_Toc136449410)

[6.2.2 time cost 30](#_Toc136449411)

# 1.Preface

## 1.1 Overview

As people's time fragmented increases, people desperately need something short and sharp to fill the gap. The popularity of short videos also confirms this. However, some people prefer to read books rather than watching videos, but do not have enough time to finish a book. This project uses machine reading comprehension to summarize the data entered by the user, so that the user can obtain the desired knowledge in the shortest time.

## 1.2 Project delivery products

Project plan, project introduction PPT, project source code, project test report

# 2. Product introduction

## 2.1 Introduction to project functions

### 2.1.1 Project core functionality

1. Realize the function of users uploading pictures and converting articles in pictures into text.

2. After entering the article, the system will understand and summarize the content in the article

3. When the user asks a question about the article, the system needs to answer the user's question according to the content of the article

### 2.1.2 Other features of the project

1. Realize the user's registration, login, modify the personal information module.

2. Diversify the mode of article upload, not limited to pictures. Users can upload the function by directly entering the content of the article.

3. The functions of article upload, system answer, user questions and other functions are realized in the front-end interface with a chat dialog box, and the user allows the system to judge what functions to perform by entering the password.

## 2.2 Technical Introduction

### 2.2.1 Summary extraction

Use textrank algorithm for summary extraction, TextRank algorithm is a graph-based sorting algorithm for keyword extraction and document summary, improved by Google's PageRank algorithm for web page importance ranking, which uses the co-occurrence information (semantics) between words in a document can extract keywords, it can extract the keywords, keyword groups of the text from a given text,and use the extractive automatic digest method to extract the key sentences of the text. The basic idea of the TextRank algorithm is to treat a document as a network of words, and the links in the network represent the semantic relationships between words.

Summary extraction is mainly divided into four parts: dividing the article into sentences, constructing a consistency matrix, calculating the importance weight score, and extracting the article in a summary according to the weight score

### 2.2.2 Machine reading comprehension

Huggingface's"Chinese\_pretrain\_mrc\_roberta\_wwm\_ext\_large" model is used for machine reading comprehension, first prepare the input data as the article and the question to be asked, secondly, encode the input, and input data into the model to get the answer, in which the model outputs the most likely answer, that is, the answer result.

The implementation of the machine reading comprehension function is mainly divided into three parts: loading the tokenizer and model, coding the input article and question awakening, and entering the data into the model to obtain the answer

### 2.2.3 Image to text function

The realization of the image to text function is divided into 9 parts: image preprocessing, binarization, tilt correction, layout analysis, character cutting, character recognition, layout recovery and post-processing and proofreading

# 3. Project planning

## 3.1 Project organization

|  |  |  |
| --- | --- | --- |
| member | role | task |
| Shi Zhouyin | Project Manager, Back-end Development Engineer,  Algorithm Engineer | Determine the task flow and ensure that the project is completed on time. Responsible for the development of back-end interfaces and the development of machine reading comprehension modules |
| Cai Shipeng | Front-end development engineer | Write graphical interfaces that make users feel simple and convenient, and ensure the correctness of data interfacing with back-end interfaces |
| Liu Haoyang | Algorithm Engineer | Responsible for developing images that can correctly recognize the user's incoming images into text and have a certain degree of fault tolerance |
| Xu Runxuan | Test Engineer | Responsible for testing front-end and back-end functional modules, and reporting test results. |

table 3.1 Project organization

## 3.2 Project scheduling

The project is planned to start from week 5 and be divided into 10 weeks, of which two milestones will be set in week 10 and week 14, and a checkpoint will be achieved each week to complete different tasks. The project is planned to participate in the project for 2 hours/week per person

3.13-3.31 (weeks 5-7).

Specify the task plan, learn the required knowledge, set up the environment, and set up the database

4.3-4.7 (week 8).

Cai Shipeng: Write the landing page

Shi Zhouyin: Writing user registration, user information proofreading function, machine reading comprehension module

Liu Haoyang: Writing of image preprocessing and binarization functions

Xu Runxuan: Test each result

4.10-4.14 (week 9).

Cai Shipeng: Write a registration page

Shi Zhouyin: Writing user registration, user information proofreading function, machine reading comprehension module

Liu Haoyang: Writing of image preprocessing and binarization functions

Xu Runxuan: Test each result

4.17-4.21 (week 10, first milestone).

Shi Zhouyin: The docking of the front-end and back-end user login and registration functions, and the docking of the algorithm-end machine reading comprehension module and the back-end

Liu Haoyang: Writing of noise removal and tilt correction functions

Cai Shipeng: The docking of the front-end and back-end user login and registration functions, and the writing of the user's modification of personal information pages

Xu Runxuan: Test the running results

4.24-4.28 (week 1 1).

Shi Zhouyin: Modify the interface of user personal information, split sentences and word segmentation functions, and build the writing of collaborative matrix functions

Liu Haoyang: Layout analysis, character cutting, writing of character recognition modules

Cai Shipeng: Writing the interface between users and back-end interactive functions

Xu Runxuan: Test each result

5.1-5.5 (weeks 12).

Labor Day, a week off

5.8-5.12 (weeks 13).

Shi Zhouyin: The writing of the calculation weight score and summary extraction module, the docking between the algorithm end and the backend, and the docking between the back-end and the front-end page.

Liu Haoyang: Character recognition, layout restoration, post-processing proofreading function writing

Cai Shipeng: The docking of the back-end and the front-end

Xu Runxuan: Test the results

5.15-5.19 (weeks 14, second milestone).

Shi Zhouyin: The algorithm ends the connection between the picture to text module, the summary extraction module, the machine reading comprehension module, and the docking of each page in the front-end.

Liu Haoyang: The algorithm connects the image to text module with the back-end

Cai Shipeng: The docking of the back-end and the front-end

Xu Runxuan: Test the results

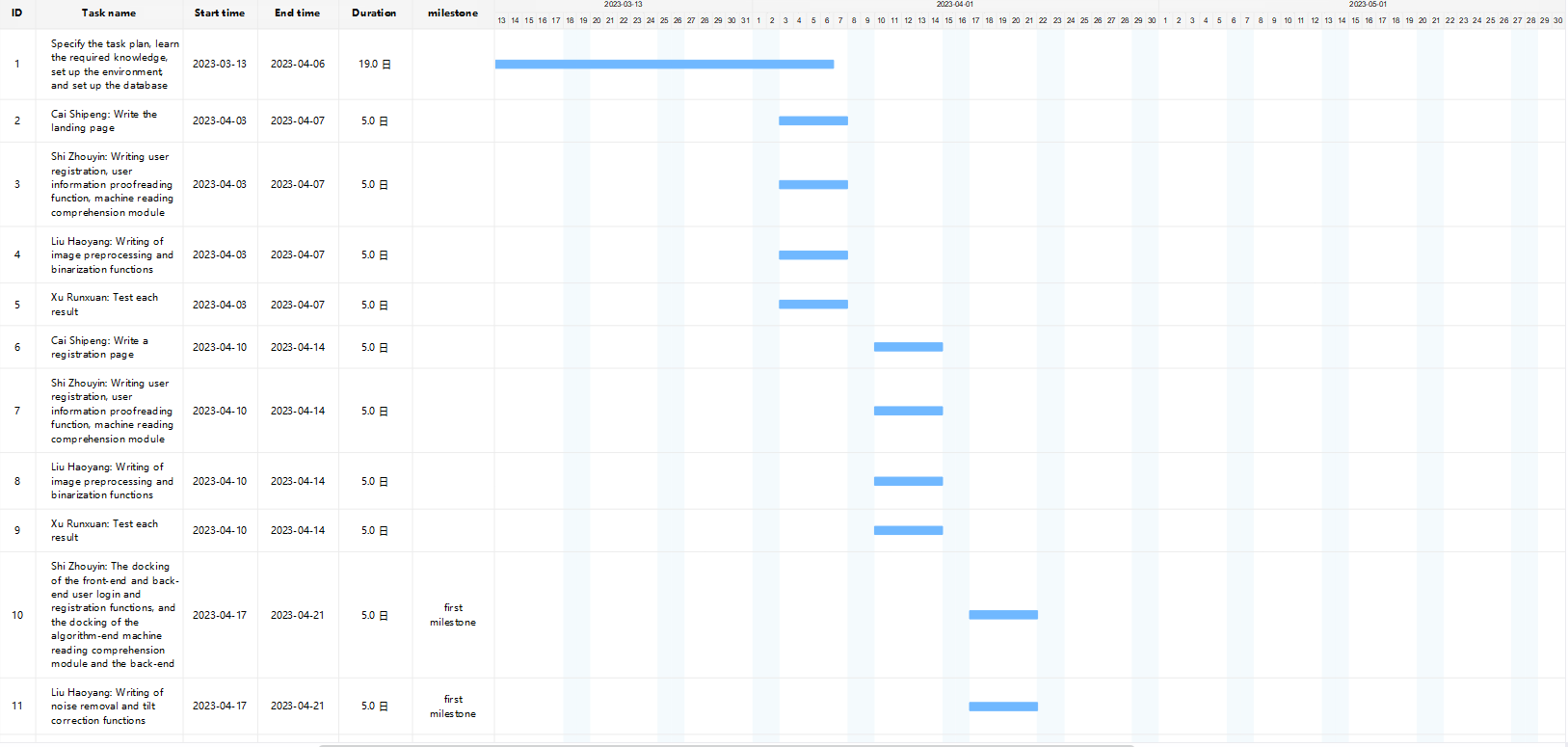
5.22-5.26 (weeks 15).

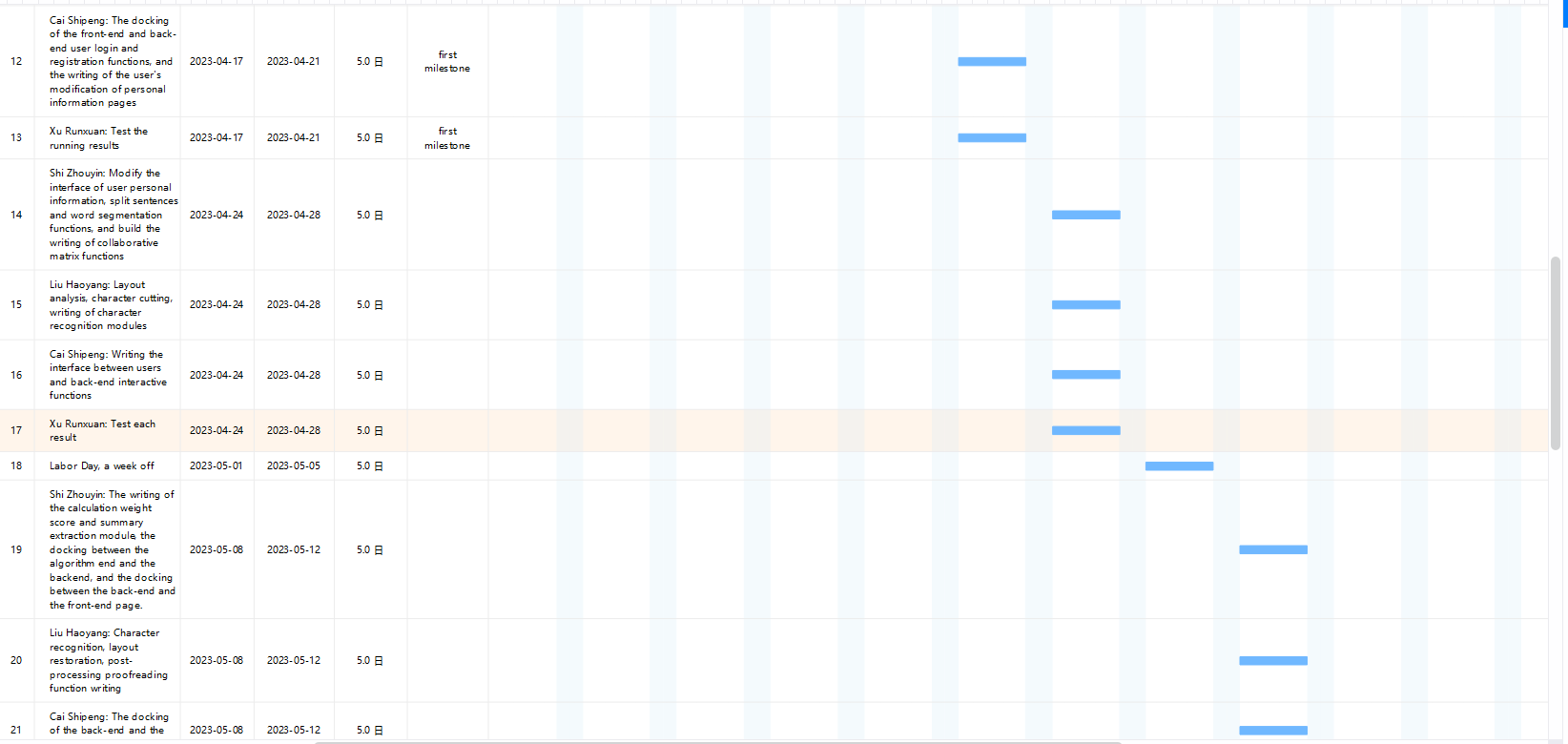
Optimize the system as a whole, test it, and view the results

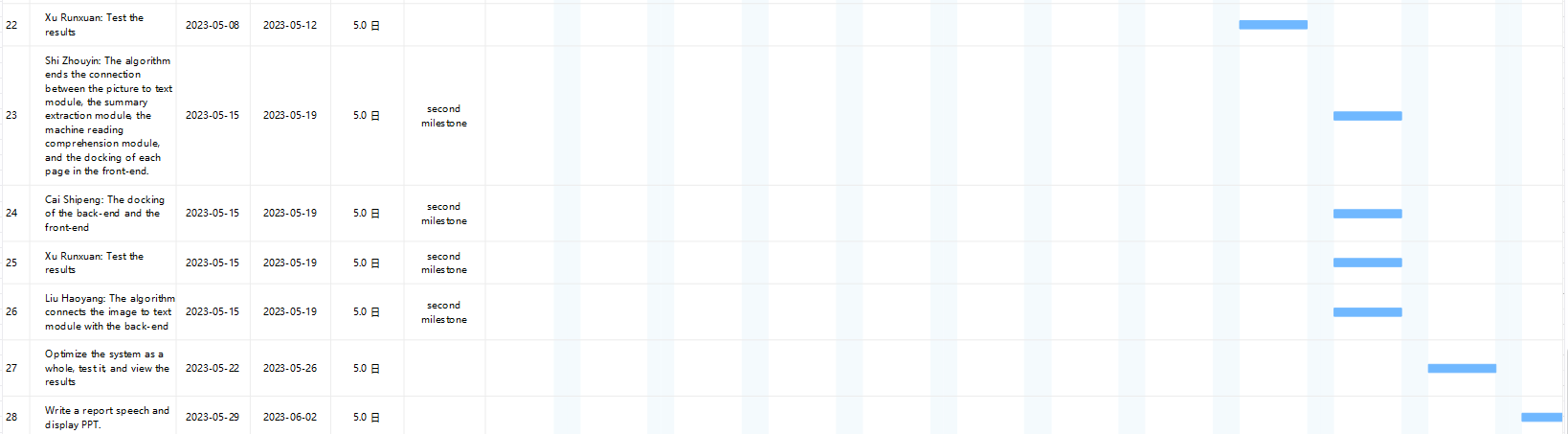
5.29-5.31 (before Thursday of week 16).

Write a report speech and display PPT.

## 3.3 Gantt char







# 4.Risk Plan

## 4.1 Introduction

A risk management plan is a document that outlines potential risks that might occur in a project or business process. It also lays out strategies and methods for reducing and controlling risks while ensuring contingency plans are in place.

## 4.2 Project Risk Indetification

We have predicted many Risk we may meet during the program, such as low proficiency in technique, sickness during the project, influence of the personal affairs and so on, we use Risk Breakdown Structure(RBS) to classify them into three levels, RBS level 0 means all the risks probably happen in project, level 1 mean the special resource of the risk and level 2 means the specific risk, Project Risk Indetification are shown in following table:

|  |  |  |
| --- | --- | --- |
| RBS level 0 | RBS level 1 | RBS level 2 |
| All risks probably happen in project | Technique Risk | low proficiency in technique |
| External Interface Risk |
| Manage Risk | Someone absent from the meeting |
| Underestimate of the workload and difficulties of a week |
| Cuurent plan is difficult to complete |
| Someone busy in their personal affairs |
| Accidental Risk | Someone get sick，recover in a week |
| The loss of current work |
| Someone get sick，recover for a long time |

table 4.2

## 4.3 Project Risk assessment

In risk assessment, risk probability and risk impact are generally considered comprehensively to assess the risk level of specific risk events, determine coping strategies, and formulate corresponding risk management plans. Therefore, accurate estimation of risk probability is one of the key factors in risk management and decision making, which usually requires analysis and calculation based on reliable data and experience.

### 4.3.1 Quantifying risk probability

Risk probability can be understood as the probability of occurrence of a certain risk event (such as accident, disaster, etc.) under certain circumstances. We qualitying risk probability into five level,the specific quantifying standards are listed as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subjective  Rating | Very low | Low | Moderate | High | Very high |
| Probabability | 0-0.10 | 0.11 -0.30 | 0.31-0.50 | 0.51-0.70 | 0.70-1.0 |

table 4.3.1

### 4.3.2 Quantifying impact

Risk impact is a measure of the impact that the occurrence of a risk event may have on an organization (or an individual) under certain circumstances. Risk impact usually involves various aspects of an organization, such as finance, operation, reputation, and security. Therefore, you need to consider the impact of different aspects in the process of estimating risk impact and calculate a comprehensive impact indicator. We qualitying risk impact into five level,the specific quantifying standards are listed as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Project  Objective | Very low/0.05 | Low/0.1 | Moderate/0.2 | High/0.4 | Very high/0.8 |
| Time cost | Insignificant  Time  increase | 1 week time  increase | 2 week time increase | 3 week time  increase | 4 week time  increase |
| Scope | Scope decrease barely  noticeable | Minor areas of scope affected | Major areas of  Scope affected | Scope reduction unacceptable to sponsor | Project end item is effectively useless |
| Quality | Quality degradation barely noticeable | Only very demanding applications are affected | Quality reduction requires sponsor approval | Quality reduction unacceptable to sponsor | Project end item effectively useless |

table 4.3.2

### 4.3.3 probability and impact matrix

We use the probability and impact matrix constructed from the risk likelihood and risk impact,which is shown as follows.In this table,blue ones means low-level risk,the white one means medium-level risk and the grey onemeans the high-level risk.The stardand of the assessment of risks are shown in table 5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Probability | Threats | | | | |
| 0.9 | 0.05 | 0.09 | 0.18 | 0.36 | 0.72 |
| 0.7 | 0.04 | 0.07 | 0.14 | 0.28 | 0.56 |
| 0.5 | 0.03 | 0.05 | 0.10 | 0.20 | 0.40 |
| 0.3 | 0.02 | 0.03 | 0.06 | 0.12 | 0.24 |
| 0.1 | 0.01 | 0.01 | 0.02 | 0.04 | 0.08 |
| Impact | 0.05 | 0.1 | 0.2 | 0.4 | 0.8 |

Table 4

|  |  |  |  |
| --- | --- | --- | --- |
| Subjective Rating | Low-level Risk | Medium-level Risk | High-level risk |
| Threats | <0.05 | >0.05 and  <0.1 | >0.1 |

table 4.3.3

### 4.3.4 assessing the Risks

We use probability score and impact score in calculate the threat score,thus assessing the risk level.

|  |  |  |  |
| --- | --- | --- | --- |
| Risk Profile | Probability | Impact | Threat |
| low proficiency in technique | 0.5 | 0.1 | 0.05 |
| External Interface Risk | 0.1 | 0.05 | 0.01 |
| Someone absent from the meeting | 0.5 | 0.1 | 0.05 |
| Underestimate of the workload and difficulties of a week | 0.3 | 0.2 | 0.06 |
| Current plan is difficult to complete | 0.1 | 0.4 | 0.04 |
| Someone busy in their personal affairs | 0.5 | 0.1 | 0.05 |
| Someone get sick，recover in a week | 0.5 | 0.1 | 0.05 |
| The loss of current work | 0.1 | 0.8 | 0.08 |
| Someone get sick，recover for a long time | 0.3 | 0.2 | 0.06 |

table 4.3.4

## 4.4 Risk Budget

For every kind of Risks,we have measures to resolve them.Here are table 7 illustrates the risk counter measures and whether we encounted the risk.

|  |  |  |  |
| --- | --- | --- | --- |
| Risk level | risk profile | Risk counter measures | Encounted? |
| Low level Risk | low proficiency in technique | Allocate time for learning and increase weekly working hours | √ |
| Someone absent from the meeting | Agree on another time to hold the meeting | √ |
| External Interface Risk | Find another useful Interface | × |
| Someone busy in their personal affairs | increase weekly working hours | √ |
| Current plan is difficult to complete | Change the project plan | × |
| Someone get sick，recover in a week | increase weekly working hours or give their work to  others | √ |
| Medium level Risk | Someone get sick，recover for a long time | Find another one to replace him | × |
| The loss of current work | Make backups for the project | × |
| Underestimate of the workload and difficulties of a week | Postpone the commit time and increase weekly working hours | × |

table 4.4

# 5. Test Plan&&Test Case

## 5.1 schedule

## 1.Test 1

|  |  |  |
| --- | --- | --- |
| **Test Content** | **Responsible Person** | **Date** |
| Test login in function | Xu Runxuan | 4.24-4.28 |

## 2.Test 2

|  |  |  |
| --- | --- | --- |
| **Test Content** | **Responsible Person** | **Date** |
| Test signup function | Xu Runxuan | 4.24-4.28 |

## 3.Test 3

|  |  |  |
| --- | --- | --- |
| **Test Content** | **Responsible Person** | **Date** |
| **Test Summary Function** | **Xu Runxuan**  **Shi ZhouYin** | **5.15-5.22** |

## 4.Test 4

|  |  |  |
| --- | --- | --- |
| **Test Content** | **Responsible Person** | **Date** |
| **Test ImageToText**  **Function** | **Xu Runxuan** | **5.24-5.28** |

## 5.Test 5

|  |  |  |
| --- | --- | --- |
| **Test Content** | **Responsible Person** | **Date** |
| **Test Machine Reading Comprehension**  **Function** | **Xu Runxuan**  **Shi ZhouYin** | **5.24-5.28** |

|  |  |  |
| --- | --- | --- |
| **Test Content** | **Responsible Person** | **Date** |
| **Test Upload**  **Function** | **Xu Runxuan**  **Shi ZhouYin** | **5.24-5.28** |

## 6.Test 6

## 5.2 design

### 5.2.1 Test 1

Interface：login

Responsible Person：Xu RunXuan

### 1. Control

Testing with Spring Integration Junit

### 2. pass standard

All Examples can input result equals to expected result

### 3. param explanation

uid:User’s identification digital

utele: User’s telephone

upwd:User’s password

uname:User’s name

### 4. Input

Example1:{uid :2,utele:13801987038,upwd:szy19991006,uname:”史周胤”}

Explanation:existed users with true login password

Expected output:return success code

Example2: {uid :2,utele:13801977038,upwd:123456,uname:”史周胤”}

Explanation:existed users with wrong login password

Expected output:return failed code

Example3：{uid：3，utele：138019656038,upwd:123434345,uname :null}

Explanation:non-existed users

Expected output:return failed code

### 5. Test Result

All examples output are equal to expected output

### 5.2.2 Test 2

Interface: adduser

Responsible Person: Xu RunXuan

### 1. Control

Testing with Spring Integration Junit

### 2. pass standard

All Examples can input the result equals to the expected result

### 3. param explanation

uid：User’s identification digital

utele: User’s telephone

upwd:User’s password

uname:User’s name

### 4. Input

Example1:{uid :null,utele:13801987038,upwd:123456,uname:”史周胤”}

Expected output:return success code

Example2: {uid :null,utele:13801964338,upwd:123756,uname:”刘浩样”}

Expected:output:return success code

Example3: {uid：13801967032,upwd:123456,uname:“蔡仕鹏”

Expected output:return success code

5. Test Result

All examples output are equal to expected output

### 5.2.3 Test 3

Interface：getSummary

Responsible Person：Xu RunXuan Shi ZhouYin

### 1. Control

Testing with Spring Integration Junit

### 2. pass standard

All Examples can successful run without exception

### 3. Input

Example 1: “该研究主持者之一、波士顿大学地球与环境科学系博士陈池（音）表示，“尽管中国和印度国土面积仅占全球陆地的9%，但两国为这一绿化过程贡献超过三分之一。考虑到人口过多的国家一般存在对土地过度利用的问题，这个发现令人吃惊。” NASA埃姆斯研究中心的科学家拉玛·内曼尼（Rama Nemani）说，“这一长期数据能让我们深入分析地表绿化背后的影响因素。我们一开始以为，植被增加是由于更多二氧化碳排放，导致气候更加温暖、潮湿，适宜生长。”“MODIS的数据让我们能在非常小的尺度上理解这一现象，我们发现人类活动也作出了贡献。”NASA文章介绍，在中国为全球绿化进程做出的贡献中，有42%来源于植树造林工程，对于减少土壤侵蚀、空气污染与气候变化发挥了作用。 据观察者网过往报道，2017年我国全国共完成造林736.2万公顷、森林抚育830.2万公顷。其中，天然林资源保护工程完成造林26万公顷，退耕还林工程完成造林91.2万公顷。京津风沙源治理工程完成造林18.5万公顷。三北及长江流域等重点防护林体系工程完成造林99.1万公顷。完成国家储备林建设任务68万公顷。“

Explanation: a common article within 500 chinese characters which is generated by chatgpt

Expected result:successfully run

Example 2: 在中国传统文化中，曲艺是一种具有广泛影响力的艺术形式。它包括评书、相声、小品、快板等多种艺术形式。曲艺的特点是幽默风趣、语言生动、表现力强，深受人们的喜爱。在中国南方的一些地区，曲艺演出还被视为传统文化的一种体现。近年来，曲艺在电视和网络等新媒体平台上得到了更广泛的传播，受到了更多年轻观众的关注。

Explanation: a short article consist of 3 sentences, which is generated by chatgpt

Expected result:successfully run

Example 3: a long article collected on Web,which has 1500 chinese characters

Expected result:successfully run

Example 4: a long article excerpted from the Report of the 20th National Congress of the Communist Party of China,which has a few part。

Expected result:successfully run

### 4. Test Result

All examples output are equal to expected output

### 5.2.4 Test 4

Interface：getartucle

Responsible Person：Xu RunXuan

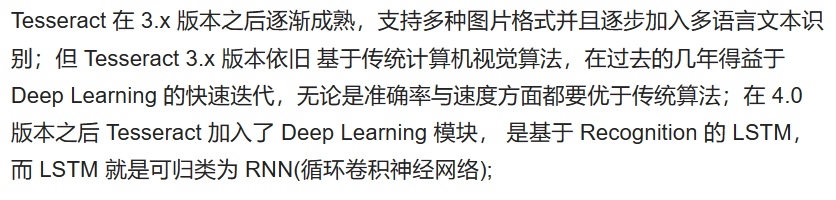
### 1. Control

Testing with Spring Integration Junit

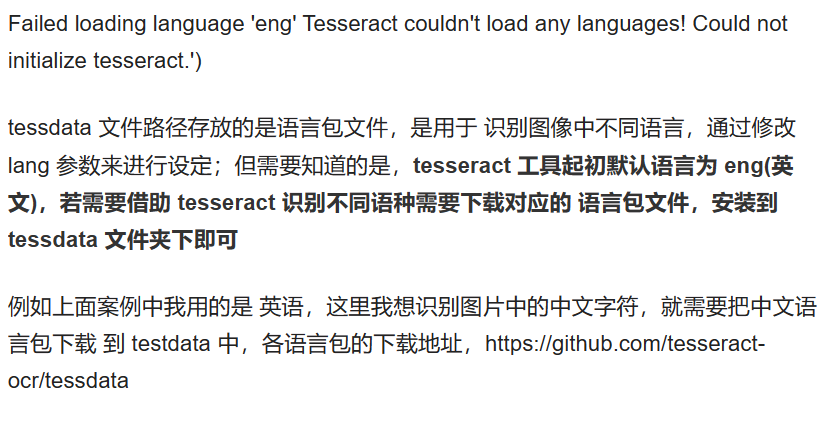
### 2. pass standard

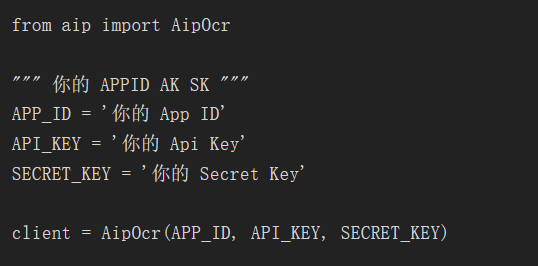
All Examples can successful run without exception

### 3. Input

Example1: 

Example2:



Example3: 

### 4. Test Result

Example1: CommonResult(code=200, message=操作成功, data=Tesseract在3.x版本之后逐渐成熟,支持多种图片格式些日逐步加入多语言文本识别;但Tesseract3.x版本依旧基于传统计算机视觉算法,在过去的几年得益DeepLearning的快速迭代,无论是准确率与速度方面都要优于传统算法;在4.0版本之后Tesseract加入了DeepLearning模块,是基于Recognition的LSTM,而LSTM就是可归类为RNN(循环卷积神经网络))

Example2：

CommonResult(code=200, message=操作成功, data=FailedloadinglanguageengTesseractcouldnttloadanylanguageslCouldnctinitializetesseract)tessdata文件路径存放的是语言包文件,是用于识别图像中不同语言,逢过修改lang参数来进行设定;但需要知道的是,tesseract工具起初默认语言为eng(英文),若需要借助tesseract识别不同语种需要下载对应的语言包文件,安装到tessdata文件夹下即可例如上面案例中我用的是英语,这里我愚识别图片中的中文字符,就需要把言包下载到testdata中,各语言包的下载地址,https:/github.comtesseract-ocrftessdata)

Example3：

CommonResult(code=200, message=操作成功, data=fromaipinportAipDer“你的APPIDMSK777APP\_ID=“你的hppID「API\_KEY=“你的ApiFey「SBCRET\_KEY=“你的SecretKey「client=Aip0cr(APP\_ID,API\_KBY,SECRET\_KEY))

### 5.2.5 Test 5

Interface：getanswer

Responsible Person：Xu RunXuan ShiZhouYin

### 1. Control

Testing with Spring Integration Junit

### 2. pass standard

All Examples can successful run without exception

### 3. Input

Example1：

Question：苹果公司成立于哪一年?

Context：苹果公司成立于1976年，总部位于美国加利福尼亚州的库比蒂诺市。这家公司最初生产的是个人电脑。如今，苹果公司生产和销售各种电子产品，如iPhone, iPad, 和Macintosh电脑。

Example2：

Question：李四喜欢看什么类型的电影?

Context：张三和李四一起去了电影院。张三看了惊悚片，但李四看的是喜剧片。电影结束后，他们一起去了餐厅吃晚饭。他们点了披萨和意大利面。

Example3：

Question: 张三和李四点了什么菜？

Context：Same to Example2

Example4：

Question: 张三和李四去了哪里?

Context：Same to Example2

Example5：

Question: 张三和李四干了什么？

Context：Same to Example2

### 4. Test Result

Example1:

Answer: 1976

Example2:

Answer:喜剧片

Example3：

Answer：披萨

Example4：

Answer：电影院

Example5：

Answer：对不起，我无法回答

### 5.2.6 Test 6

Interface：getarticle

Responsible Person：Xu RunXuan

### 1. Control

Testing with Spring Integration Junit

### 2. pass standard

All Examples can successful run without exception

### 3. Input

Example1: a txt format file

Expected result : return success code

Example2：a pptx format file

Expected result : return failed code

### 4. Test Result

All examples output are equal to expected output

# 6.Project Execution Details

## 6.1 Project Execution Details Table

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Plan | Actual | Reason |
| 4.3-4.9 | Cai Shipeng：Writing login page.  Shi Zhouyin：Writing user registration, user information verification function, and machine reading comprehension module.  Liu Haoyang：Writing image preprocessing and binarization function. | Cai Shipeng：Writing login page.  Shi Zhouyin：Writing user registration, user information verification function, and machine reading comprehension module.  Liu Haoyang：Writing image preprocessing and binarization function. |  |
| 4.10-4.14 | Cai Shipeng: Writing registration page. Shi Zhouyin: Writing user registration, user information verification function, and machine reading comprehension module. Liu Haoyang: Writing image preprocessing and binarization function. | Cai Shipeng: Developing the registration page. Shi Zhouyin: Implementing user registration and user information verification functionality. Liu Haoyang: Implementing image preprocessing and binarization functionality. | Shi Zhouyin was unable to make progress in the development of the machine reading comprehension module this week due to illness. |
| 4.17-4.21 | Shi Zhouyin: Integrating frontend and backend for user login and registration functionality, and integrating the machine reading comprehension module with the backend. Liu Haoyang: Developing noise removal and skew correction functionality. Cai Shipeng: Integrating frontend and backend for user login and registration functionality. Xu Runxuan: Testing the running results. | Shi Zhouyin: Integrating frontend and backend for user login and registration functionality, and developing the machine reading comprehension module. Liu Haoyang: Developing noise removal and skew correction functionality. Cai Shipeng: Integrating frontend and backend for user login and registration functionality. | The machine reading comprehension module has been completed, but the integration of the algorithmic part of the machine reading comprehension module with the backend and other pages has not been finished, resulting in the inability to complete testing. |
| 4.24-4.28 | Shi Zhouyin: Developing sentence segmentation and word segmentation functionality, and building collaborative matrix functionality. Liu Haoyang: Writing layout analysis, character segmentation, and character recognition modules. Cai Shipeng: Writing user interface for interaction between users and the backend. Xu Runxuan: Testing the results of each component. | Shi Zhouyin: Integrating frontend and backend for user login and registration functionality, and integrating the machine reading comprehension module with the backend.  Liu Haoyang: Performing layout analysis and character segmentation.  Cai Shipeng: Integrating frontend and backend for user login and registration functionality.  Xu Runxuan: Testing the results of each component. | There were some issues with the integration of the frontend and backend for user login and registration functionality, which have been addressed and fixed this week. Additionally, Shi Zhouyin was unable to complete the development of the sentence segmentation and word segmentation functionality, as well as the collaborative matrix functionality, due to other coursework and assignments related to the end of the semester. The character recognition module and the article upload page were also not completed this week. |
| 5.1-5.5 | Holiday | This week, Shi Zhouyin will be working on the development of the sentence segmentation and word segmentation functionality, as well as the collaborative matrix functionality. Liu Haoyang will be focusing on the character recognition module, and Cai Shipeng will be working on the article upload page. | Working overtime. |
| 5.8-5.12 | Shi Zhouyin: Developing the weight scoring and summary extraction modules, integrating the algorithmic part with the backend, and connecting the backend with the frontend pages.  Liu Haoyang: Writing the character recognition, layout restoration, and post-processing verification functionality.  Cai Shipeng: Connecting the backend with the frontend.  Xu Runxuan: Testing the results. | Shi Zhouyin: Developing the weight scoring and summary extraction modules, integrating the algorithmic part with the backend, and connecting the backend with the frontend pages.  Liu Haoyang: Writing the character recognition and layout restoration functionality.  Cai Shipeng: Connecting the backend with the frontend.  Xu Runxuan: Testing the results. |  |
| 5.15-5.19 | Shi Zhouyin: Integrating the algorithmic part with the image-to-text module, summary extraction module, and machine reading comprehension module, as well as connecting with various frontend pages.  Liu Haoyang: Integrating the algorithmic part with the image-to-text module and the backend.  Cai Shipeng: Connecting the backend with the frontend.  Xu Runxuan: Testing the results. | Shi Zhouyin: Integrating the image-to-text module, summary extraction module, and machine reading comprehension module with the algorithmic part.  Liu Haoyang: Writing post-processing and verification functionality, as well as integrating the image-to-text module with the backend.  Xu Runxuan: Testing the results. | The integration between the backend and the frontend has not been completed. |
| 5.22-5.26 | Performing overall system optimization, testing, and reviewing the results. | Integrating the backend with the frontend, conducting overall system optimization, testing, and reviewing the results. |  |

## 6.2 Project completion and time cost

### 6.2.1 project completion

The graph 6.2.1 shows the progress of the project in terms of planned completion and actual completion. The largest discrepancy occurred between April 24th and May 1st, with a 20% difference between the actual completion and the planned completion. For specific reasons, please refer to the project execution details table in section 6.1.

graph 6.2.1

### 6.2.2 time cost

The graph 6.2.2 displays the time spent by each team member during the project period.

graph6.2.2