

**University of Plymouth**

**School of Engineering,  
Computing, and Mathematics**

**PRCO304  
Final Stage Computing Project  
2019/2020**

**Q&A Comprehensive Community**

Ma Zhuang

10664074

BSc (Hons) Computing

## Acknowledgements

Firstly, I would like to thank my project director; Dr Swen Gaudl provided support and advice throughout the process and met with me to discuss the continued availability of any part of the project. In addition, I would also like to thank my parents for their firm support during college. This project and graduation experience are largely the result of their support and encouragement, and the result of my work throughout the degree period.

## Abstract

The report describes a software development project. It aims to improve the self-learning environment of students and the efficiency of self-learning. The target customers of this project are students of computer related majors. Students often encounter problems in self-study courses. It is easy to find solutions to some common problems through the Internet or books, and it is difficult or impossible to find some uncommon problems through searching for information. This situation may directly reduce the customer's self-learning efficiency. In severe cases, this may cause the customer to miss the submission date of the job and even fail to complete the job. The project aims to alleviate such problems and provide a convenient method to facilitate communication and find solutions to problems in a self-study environment.

The application created is an interactive web application that provides a platform for students. So that the problems raised by students can be best resolved as soon as possible. At the same time, it provides a way to make students' questions more likely to be discovered by customers who know how to solve them.

The development process uses an agile method with multiple incremental stages to track content posted by users and recommend content that users might be interested in through user behaviours. Identify legal, social, ethical, and professional issues that use this feature to develop applications and explain how to resolve these issues in the application.

The selected project management method was AGILE and discussed how to use this method throughout the project. The report presents the proposed requirements; how to decompose them into functional and non-functional requirements, and how to use priority user cases in the project.

Requirements and other design principles are used to determine the application architecture, which is implemented using technologies such as Java, MySQL, Neo4J, Kafka, Redis, and Elasticsearch. It introduces the testing process used on the product and discusses how to use them throughout the project.

Provide project evaluation of project deliverables to emphasize its implementation and possible improvements. Finally, an after-the-fact report will be provided

indicating what went well and what could be improved; comment on issues such as development technology and personal performance.

## Table of Contents

Acknowledgements .....	1
Abstract .....	1
1.0 Introduction .....	6
2.0 Background, objectives & deliverables .....	6
2.1 Background .....	6
2.2 Project objectives.....	7
2.3 Deliverables .....	7
3.0 Method of approach .....	8
3.1 Agile.....	8
3.2 Scrum .....	9
3.3 Test-Driven Development.....	10
4.0 Project management .....	11
4.1 Scrum .....	11
4.2 Trello.....	13
4.3 GitHub .....	13
5.0 Legal, social, ethical and professional issues.....	14
5.1 Legal issues.....	14
5.1.1 Data .....	15
5.1.2 Security .....	16
5.2 Social issues.....	17
5.3 Ethical issues.....	17
6.0 Stage 1: Further Research and initial planning .....	17
6.1 Background research.....	17
6.1.1 Research of existing systems.....	17
6.1.2 Discussion of existing systems.....	18
6.1.3 Questionnaire.....	18
6.2 System requirement.....	18
6.2.1 Functional requirements.....	18
6.2.2 Non-functional requirements .....	19
6.3.3 Students stories .....	19

6.3.4 Administrator stories .....	19
6.3.5 Extended Functionality.....	20
6.3 Technologies to be used.....	20
6.4 End of stage discussion.....	21
7.0 Stage 2: System architecture design.....	22
8.0 Stage 3: HCI design .....	22
8.1 Interface design principles .....	22
8.2 User interface .....	23
9.0 Stage 4: Design of backend system.....	23
9.1 Database .....	23
9.1.1 MySQL.....	23
9.1.2 Neo4J.....	26
9.1.3 Redis.....	28
9.1.4 Elasticsearch.....	28
9.2 Server .....	29
9.2.1 Top use case diagram.....	29
9.2.2 Class diagram .....	29
10.0 Stage 5: Implementation .....	32
10.1 Recommended function implementation.....	32
10.2 Popularity implementation .....	34
10.3 Crawler implementation .....	35
10.4 Sensitive word filtering implementation.....	35
11.0 Stage 6: Integration and test .....	36
11.1 Unit testing.....	36
11.2 Functional Testing .....	37
11.3 Usability Testing .....	37
11.4 Verification and Validation .....	37
12.0 Stage 7: Deployment.....	39
13.0 End-project report.....	39
13.1 Summary .....	39
13.2 Project Objectives .....	39
13.3 Changes to original plan .....	41
13.3.1 Requirements .....	41

13.3.2 Technologies.....	41
14.0 Project post-mortem .....	42
14.1 What went well.....	42
14.2 What could have been improved .....	42
14.3 Possible extensions to functionality .....	43
15.0 Conclusion .....	44
16.0 Reference list .....	45
17.0 Appendices .....	46
Appendix 1: User Guide.....	46
1.1 User interface.....	46
1.2 Build project .....	53
Appendix 2: Project management.....	55
2.1 Trello boards .....	55
2.2 Plans .....	73
2.3 Sprint reviews.....	75
Appendix 3: Background Search .....	79
3.1 Second-hand Research .....	79
3.2 Questionnaire.....	82
3.3 Analysis of research results .....	84
Appendix 4: Design.....	91
4.1 Register feature design .....	91
4.2 Sign in feature design .....	92
4.3 Oauth2.0 feature design.....	93
4.4 Change password feature design.....	94
4.5 Change avatar feature design .....	95
4.6 Sensitive word filtering feature design.....	96
4.7 Ask question feature design .....	97
4.8 Comment and reply feature design .....	98
4.9 Online chat feature design .....	99
4.10 Like and follow features design .....	100
4.11 Recommendation feature design .....	101
4.12 System notification feature design .....	102
4.13 Full text search feature design .....	103

4.14 Statistics feature design .....	104
Appendix 5: Test.....	105
5.1 Unit test results .....	105
5.2 Functional test results .....	119
5.3 Usability test feedback .....	128

WORD COUNT: 100291

Code submission link: <https://github.com/Final-Year-project-10664074-MaZhuang/ComprehensiveCommunityFinal>

## **1.0 Introduction**

The idea of this project was put forward due to the questions encountered by the author during the self-study process. The client targeted at this project is students in computer-related majors.

From the student's perspective, students often encounter questions when they self-learn a course. Solutions to some common questions are easy to find through the Internet or books, while some uncommon questions are difficult or impossible to find by searching for information. This situation may directly reduce the efficiency of the customer's self-study. In severe cases, it may cause the customer to miss the submission date of the assignment and even fail to complete the assignment. The project aims to alleviate such questions and provide a convenient way to facilitate communication and find solutions to questions in a self-study environment.

The application created is an interactive web application that provides a platform for students. Make the questions raised by students be solved as quickly and best as possible. At the same time, it provides a way to make the questions posted by students more likely to be discovered by customers who may know how to solve such questions.

## **2.0 Background, objectives & deliverables**

### **2.1 Background**

The main reason for creating this system is to try to improve some special problems encountered by students in self-study and improve the efficiency of self-study. For example, students encounter problems related to the underlying level of the development language during the completion of a development task. Such problems are special problems. It is difficult for students to find solutions through the Internet. They may need to discuss or ask people who have the same experience. In the self-study stage of students, the usual method is to study through books or the Internet, and often some questions are difficult or impossible to find solutions through the Internet or books. If this situation is left unresolved for a long time, it may cause the students to reduce their interest in the course and eventually lead the students to lag others. (For details, please refer to Appendix 3: Background search).

At the same time, there are other questions. For example, students in the same course who encounter the same problem may send emails to the instructor multiple times to ask for solutions. Although there are some remedial measures in academic institutions, such as a teacher-student discussion group for a course, it is difficult for students to communicate with the tutor efficiently and solve questions outside the opening hours of the discussion group. In addition, each student's communication level and understanding level are not the same. Often the same questions are asked

by different students, and the solutions are different. This may cause students to complete assignments or projects that do not meet expectations.

In addition, at the beginning of a course, the tutor will only provide some books or websites for students to search for materials. But the common questions encountered by students in the learning process do not have a unified platform for browsing. This may cause students to repeat common mistakes made by students before, which reduces learning efficiency and increases the workload of tutors.

The solution provided by the system aims to solve the above questions, provides an interactive dialogue between students and faculty or students and students, and provides a platform for students to view the common questions encountered in this course, but also for the future. Students who will study this course will make reference or preview, so as to improve students' learning efficiency and interest.

The author hopes to truly provide a solution for academic institutions to improve the efficiency of student self-study, and at the same time gain more programming experience in the completion of the project stage. The latest version of the system is aimed at users of computer-related majors, but the author hopes that the future users of the system will be all students who are not majors in the entire academic institution, and this is another motivation for completing the project.

## **2.2 Project objectives**

The main goal of the project is determined through the author's personal experience, hoping to develop a platform that can help students improve the efficiency of self-study while providing a platform for other students to share knowledge. (For details, please refer to Appendix 2: 2.2 Requirement and plans).

- a) Analyse existing Q & A systems and query literature about recommended algorithms to determine project requirements and efficient algorithms in general.
- b) Further understand user needs through questionnaires to further clarify project needs.
- c) Analysis and selection of development technologies, deployment options and implementation.
- d) Design a test plan for the final version of the system to ensure that it meets all needs and designs
- e) Preparation of academic materials related to the project, including final reports, posters and videos

## **2.3 Deliverables**

The following is a list of deliverables that should be included after the final submission of the project:

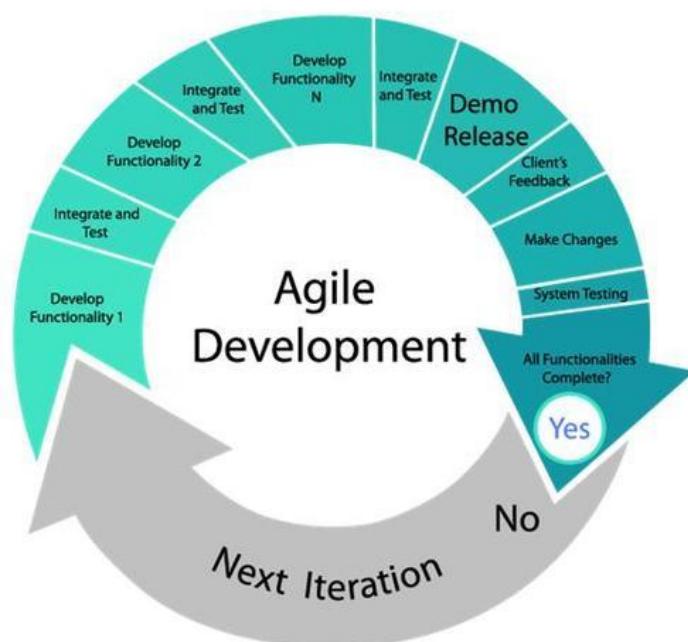
- a) A complete Web application that meets the project's expected needs and can improve the efficiency of student self-study.
- b) MySQL database used to store information such as account and Q & A details.
- c) Neo4j database, used to implement the recommendation system.
- d) Elasticsearch database, used by the system to implement full-text search.
- e) Redis database, used to store user behaviour. Such as like, follow, etc.
- f) User operation manual.
- g) Project poster.
- h) Final report.

## 3.0 Method of approach

### 3.1 Agile

For this project, it has a clear time limit, and the project has a certain uniqueness, it uses agile methods for development. In the Agile Manifesto, when advocating software development, we should focus on software interactions, customer collaboration, and response to changes in requirements to minimize unnecessary documentation or negotiation (Agile Alliance, 2017).

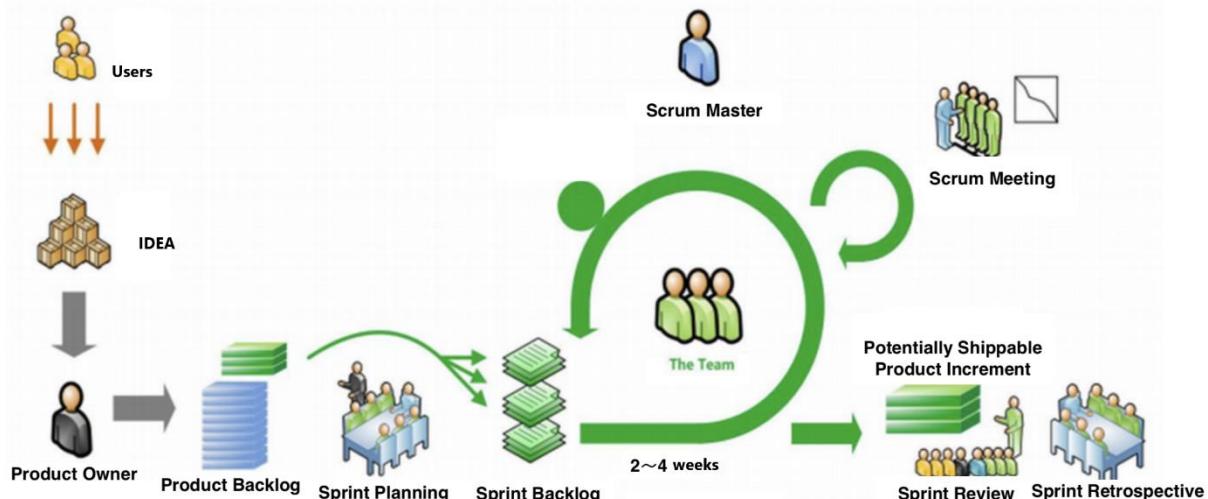
The main reason for using agile methods to develop software is that the software development process is continuous, not one-off. For the requirements determined at the beginning of the project, they may change at any time during the development phase. It is difficult to realize that the project requirements change with the changes in customer needs if traditional development methods are used. The use of agile method development can promote iterative and incremental development of the project, which greatly improves the flexibility of the project.



These features are very important for the project, because many users will participate in the development of the project to ensure that the products that users like, and use are created. If the customer is not satisfied with the product currently developed, then the flexibility or flexibility of agile development can change the requirements or design. At the same time, it also means that developers can provide a complete and runnable program after each software iteration. This is crucial for the project, because the project time is strictly limited, providing a complete and runnable software after each iteration, which can make it easier for customers to intuitively feel whether the project meets customer requirements at this stage , Reducing unimportant development and shortening the project cycle.

### 3.2 Scrum

The Scrum agile framework is used in this project. The framework focuses on the development of a series of iterative fixed-length project development schedules, which is ideal for projects with strict time constraints (Radigan, 2017).



First, the product owner obtains user requirements and generates a priority list of product requirements. At the beginning of each iteration cycle, an iteration planning meeting will be held to select the functions to be implemented in this iteration period from the product requirements list. Get a Sprint list. Next, the team will enter an iterative period of 1-4 weeks. During the iteration period, the team will carry out project development, testing, documentation, etc., need to hold a daily meeting to understand the progress of the team members' work and encountered Difficult, at the end of the iteration period, a project review meeting and a review meeting will be held. However, because this project is developed by one person, it cannot fully satisfy the Scrum framework. Therefore, the author will make appropriate changes in the project.

- **Roles**

1. **Product Owner**

In the project, the product owner is the author, who is mainly responsible for determining the project requirements, maintaining the product requirements list, and making user stories.

2. **Scrum Master**

In the project, the Scrum master is served by a mentor, who is mainly responsible for presiding over the meeting and eliminating the difficulties encountered by the team and outside interference.

### 3. Scrum Team

Due to the particularity of the project, the scrum team is also the author, who is mainly responsible for the selected development framework and actual development work.

- **Sprint Backlog**

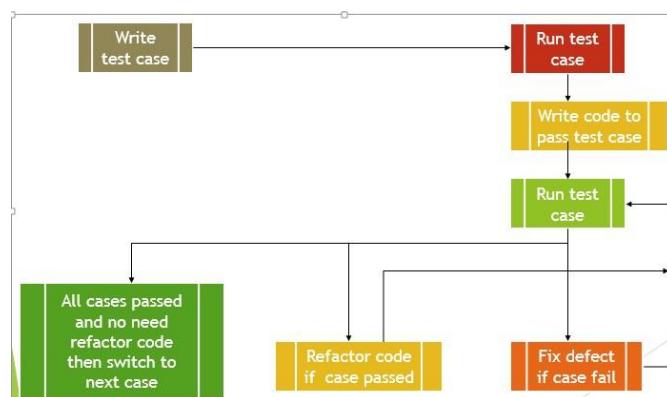
Priority list of requirements to be completed in the next sprint (Scrum Inc, 2017).

- **Meeting**

The meeting time is decided by the tutor, and the meeting is held once every week on average. The content of the meeting is mainly to report the progress of the project to the Scrum Master, what the development team completed in the previous sprint stage, and what to be completed in the next sprint plan to determine the project progress and control the direction of the project to prevent unnecessary development from causing the project failure.

## 3.3 Test-Driven Development

The use of Test-driven development (TDD) in this project can integrate the test part into every aspect of development, and then develop and test, and the development is completed and the test passes. Secondly, there may be deviations in the understanding of some requirements during the development process, so write unit tests based on test cases to contain this situation at the beginning of the work, as far as possible to avoid the situation where the developed interface does not meet the requirements. It also greatly reduces the development time.



The basic idea of test-driven development is to write test code before developing functional code, and then only write functional code that passes the test, thereby driving the entire development process with testing. The code has high strength and robustness, can quickly respond to changes, and accelerate the development process.

The basic process of test-driven development is as follows:

1. Quickly add a test
2. Run all the tests (sometimes only need to run one or a part) and find that the newly added test fails
3. Make some small changes to make the test program run as soon as possible, so the program can contain some unreasonable methods
4. Run all the tests and all pass
5. Refactor code to eliminate duplicate design and optimize design structure

## 4.0 Project management

### 4.1 Scrum

The project management method in the project is the Sprint method in Scrum. Choosing this method for project management can improve the efficiency of the development team while keeping the project in a managed environment. During the project development process, the leader will set the priority of each requirement. Then the developers develop through the sprint plan, to ensure the maximum deliverables after each Sprint phase.

Due to the particularity of the project, the scrum development framework has not yet been fully implemented. The following functions have been incorporated into the project (For details, please refer to Appendix 2):

- Sprint  
The sprint cycle, in layman's terms, is the cycle of achieving a "small goal". It usually takes 2-6 weeks.
- User story  
Users' external business needs. Take the banking system as an example, a story can be the user's deposit behaviours, or query the balance, etc. This is the so-called small goal itself.
- Task  
The specific development task split by User Story.
- Backlog  
The demand list be a list of small goals. Divided into Sprint Backlog and Product Backlog.
- Daily meeting  
A daily meeting will be used to monitor the progress of the project.  
In this project, the daily meeting is completed by the author. It is mainly to summarize the previous day's code before starting development on the second day, to quickly enter the new development link
- Sprint Review meeting (For details, please refer to Appendix 2: 2.3 Sprint reviews).  
Sprint review meeting for team members to demonstrate results.  
The Sprint Review meeting is chaired by the group tutor, and the time of the meeting is also determined by the tutor. Usually the content of the meeting is

to summarize the functions completed or uncompleted in the previous sprint stage, and the plan for the next sprint stage. Also demonstrate the deliverables that have been completed in the previous stage.

- Sprint burn down

Sprint burndown chart, to put it bluntly, is to record the demand completion of the current cycle.

- Release

The development cycle is complete, and the project releases a new available version.

- Highlight Reports

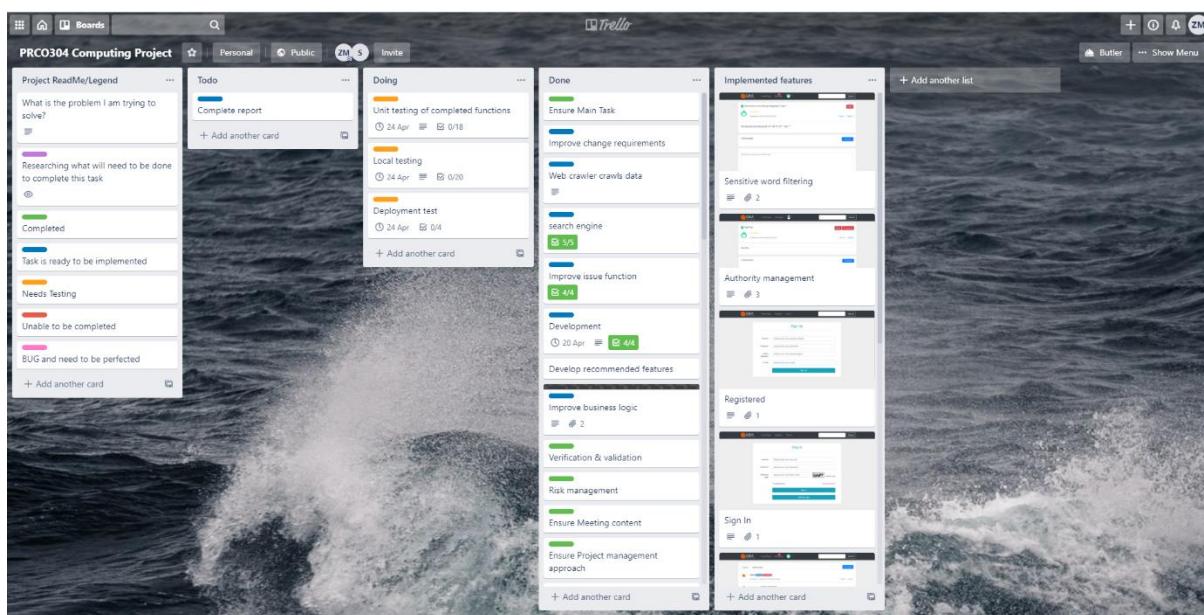
Generate key reports for weekly meetings during project development. These reports are mainly to help the development team review this week's work and set the goals to be achieved next week. In addition, this is also a means to report the progress of the project to the project manager and control the project progress.

- End Project Report

Generate final project report and discuss project performance.

## 4.2 Trello

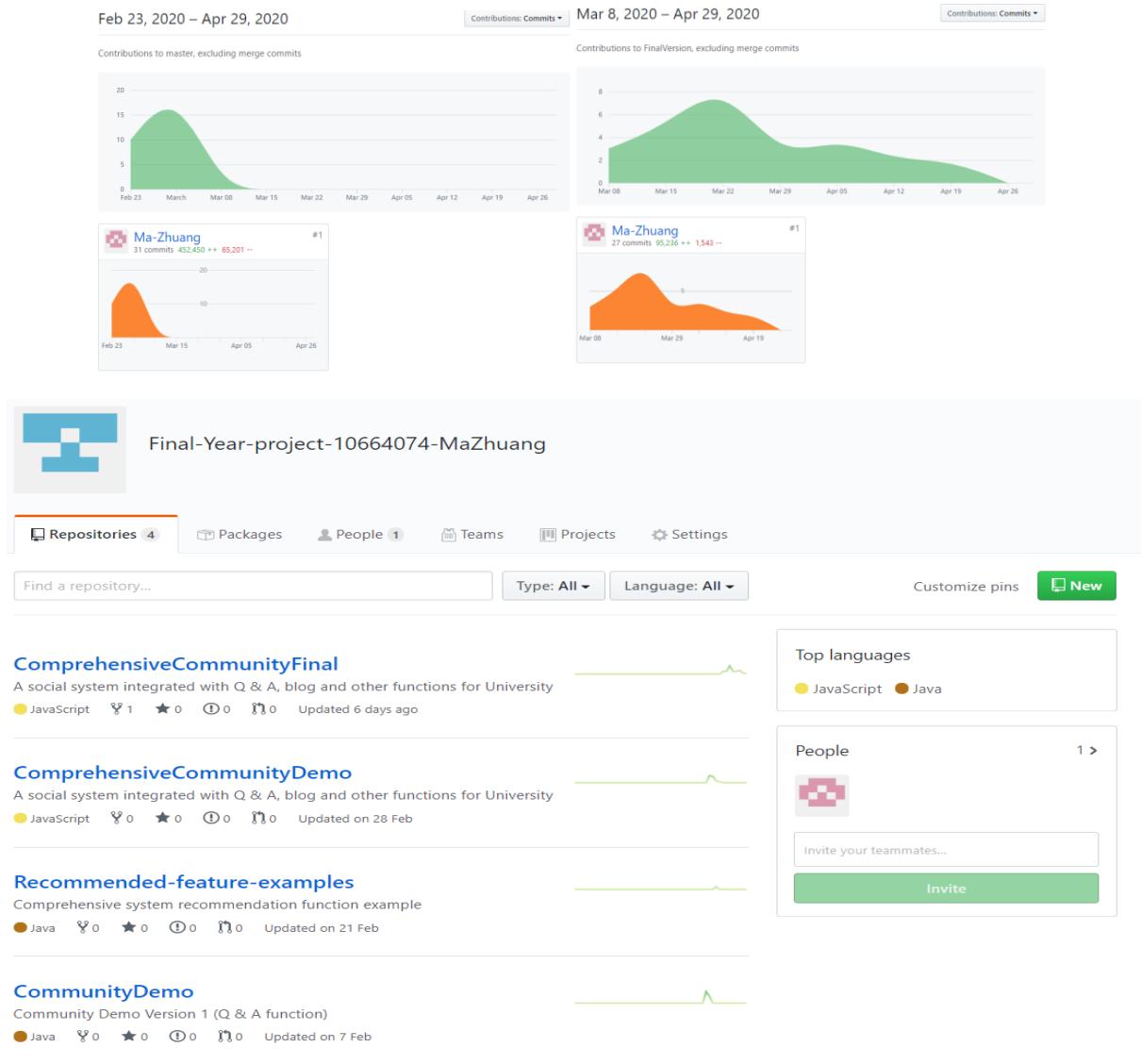
To complete the project on time and with guaranteed quality, the project management method and framework alone are not enough. In this project, the author uses Trello as a project management tool to implement the management project. Trello helps to organize a visual overview of the content being processed. Trello was used throughout the development process to help record and maintain sprint and backlog products. At the beginning of each sprint, the project manager can get the current task from the product to-do event, and each task will have detailed details to show. Therefore, the project supervisor can more easily obtain the current project progress according to the tasks in the completion list and control the project progress (For details, please refer to Appendix 2: 2.1 Trello boards).



## 4.3 GitHub

Version control at the project development stage is very necessary. Using version control can improve development efficiency, centrally manage the project code, and solve the questions of consistency and redundancy. In addition, track and record the development process of the software, assist in coordination and management of software development. On the other hand, version control can improve the disaster prevention of the project and reduce the risk of the project. For example, the developer's development equipment is damaged, and the stored files are all lost.

The version control tool used in this project is GitHub. Four repositories have been established, which respectively store the prototype code used by the project to determine requirements, the prototype code specific to the project, the prototype code used for delivery, and the prototype code finally delivered. The prototype code base for final delivery has two branches. The first branch stores the code before the project requirements are modified, and the second branch stores the code that the project is finally delivered to the user.



## 5.0 Legal, social, ethical and professional issues

### 5.1 Legal issues

Since the customers targeted by the system are students, users of the system may have a minor group. It is necessary to abide by the legal responsibility for minors. Besides, regarding the topics posted by users, try to avoid ethical and other debates.

According to the method of using personal information enacted by the "Data Protection Law" of 1998, the Act stipulates that all student-related information held by the system can only be used for specific purposes. The data related to the students mastered in this project will be listed below. The project needs to strictly abide by the principles of the data protection law to ensure that information is processed according to people's data protection rights and that security is maintained (Gov.uk, 2017).

### 5.1.1 Data

- **Users information**

The system will record the following user information:

1. Username

The user's username does not need to be a real username but may be a name made by the user himself. It fully avoids obtaining the user's real information.

2. Email

The email must be a real and active email account, because it is often used when creating new users or recommending related questions.

3. User picture

For the user's avatar, the user can set it arbitrarily, no real user photo is required.

4. Password

In order to protect the security of the user account, the user's password is encrypted by the hash algorithm and salt is added to increase the security of the password. In addition, it avoids developers who have access to the database to clearly obtain the password of an account.

5. Account type

Only the account type of the account is recorded. The account type defines the role of the account.

6. Account status

Record account status to determine whether the user is an activated account.

The system strictly complies with the user information policy stipulated in the GitHub Registered Developer Agreement. If the user uses a GitHub account to log in, the system will obtain the username of the renamed user's GitHub account, the already published Email and avatar information. Another user information will not be recorded.

In general, the data collected by the system will be used for user profiles and accounts, post questions, answer questions, like, send e-mail, online chat and other functions.

- **Crawler data**

This project strictly implements the "Content Permissions, Restrictions, and Creative Commons Licensing" regulations declared by StackOverflow in the Public Network Terms of Service. The copyright of all StackOverflow content recorded in the system belongs to StackOverflow or a third party. The system only records the public part, including the question title, summary and link address. When the user clicks on the relevant question, it will automatically jump to StackOverflow to view the details.

### 5.1.2 Security

The safe storage and protection of personal information is extremely important for a project, especially for information with minor users. Although the system will store the personal sensitive information as the minimum, it will still store the personal email address, so the security of the data stored by the system is also extremely important.

The project has adopted the following methods for safety management:

- **Authority management**

The system is divided into three types of user rights, they are ordinary users, moderators and administrator users. For ordinary users, they can make changes to their personal information, post questions, modify, like, comment and follow an account. For moderators and administrator users, in addition to having all the rights of ordinary users, they also have the operations to delete, extract and pin the published questions. In addition, administrator users can also monitor the running status of the project, view the statistics of the number of people who visit the website every day and manage the tags. The purpose of this is to control the risky operations of the data to prevent the data from being maliciously deleted to ensure the safety of the user.

- **Authentication**

The system provides two authentication methods, the first is to log in by username and password, and the second is to log in through GitHub Oauth2. No matter what kind of login method, the premise must be through email authentication to obtain user permissions.

- **Password security**

Since the system has two authentication methods, when a user logs in with a GitHub account, the user's password will not be set, and the default value is null. The password of the user will be recorded in the database. But before recording to the database, the system will encrypt the account password through hash algorithm and salting. Even people with access to the database cannot directly obtain the password of any user, ensuring the security of the account.

- **E-mail activation**

When the user registers, the system will send an activation email by email to activate the newly created account. If the user does not activate the account, no user rights will be obtained. If the customer uses the GitHub account to log in, it will be divided into two cases. The first case is that the customer's GitHub account has been set up with a public mailbox, then the system will not send an activation email. To the system account. The second case is that the customer does not set a public mailbox on GitHub, the system will automatically jump to the mailbox binding page for mailbox binding operation. After the user is bound, the system will automatically send an activation email to activate the account. This prevents some zombie accounts from being

created, which reduces the pressure on the database and reduces the risk of data theft.

## **5.2 Social issues**

All works uploaded by the user to the system only represent the user's personal views and positions and have no connection with the author. Any disputes caused by the user's behaviour or any works, information, content, etc. uploaded, stored, published, or transmitted by the user, or Any direct, indirect, accidental, and special damages caused by this will be borne by the user, and the author will not bear any responsibility.

## **5.3 Ethical issues**

The system cannot substantively examine the ownership or legality, compliance, authenticity, scientific, integrity, and validity of the works, information, and content transmitted by users; regardless of whether the system has filtered sensitive words, users Should bear legal liability for infringement or ownership disputes that may or have arisen because of the works, information, and content transmitted. At the same time, users should independently and independently judge and consider any works, information, content, etc. transmitted by other users known in the system, and measure their authenticity, effectiveness, and related risks. The risks and losses suffered by users, content, etc. are handled by users and the transmitters of works, information, and content. Other users add publicity materials or advertising information, talent recruitment needs, or otherwise display their products or services on the system in their published work information. The legal relationship between the user and the published user due to such information, content or Disputes should be resolved by themselves, the author does not bear any responsibility.

# **6.0 Stage 1: Further Research and initial planning**

## **6.1 Background research**

### **6.1.1 Research of existing systems**

Before the project begins to develop, it is extremely important to understand the solutions that already exist in the industry. Although the project has never been decided as a commercial solution, it is possible to gain a deeper understanding of the solution by looking at existing solutions in the field rather than just cloning existing products. In addition, understanding the existing solutions may be of great help in developing the project. Not all solutions provide similar functions, and each solution has its own unique features. They provide some useful inspiration for developing this project. (For details, please refer to Appendix 3: 3.1).

### 6.1.2 Discussion of existing systems

After investigation found, the comprehensive Q & A community for universities is very scarce, and some companies have directly stopped the Q & A service (e.g. Google). Other types of Q & A communities only include some basic functions, and the functions that can be achieved are relatively single, and most of them only implement the functions of asking and answering questions. On the other hand, in terms of user search results, there are also limitations to what can be searched. Of course, the most important point is that this type of system is for global users, and the user range and user quality cannot be effectively controlled. In addition, the users of some systems are more targeted users and cannot be used by anyone (such as StackOverflow). Through the above analysis, it can be shown that there is a vacancy in the Q & A community market for universities, which is feasible for developing such systems. (For details, please refer to Appendix 3: 3.3-3.3.1).

### 6.1.3 Questionnaire

Questionnaire is an effective research method. Before starting to design the system, interview users with questionnaires and record the data to be investigated. Then, experience as much as possible of all the functions of the existing question answering system. This is to promote the development of the system, avoid repeated errors, and provide better services for teachers and students.

Questionnaire survey is a first-hand survey method. Questionnaire survey is a very effective method to obtain user needs. After subsequent data analysis, it can quickly and accurately obtain the needs of many users. At this stage, it is necessary to avoid restrictions on user needs. Each questionnaire requires 35 valid data to be collected. For the final analysis stage, author first need to summarize the data of each questionnaire and analyse them one by one to determine the initial functional requirements. (For details, please refer to Appendix 3: 3.2 and 3.3-3.3.2).

## 6.2 System requirement

### 6.2.1 Functional requirements

- a) Login (or authorize login) / register account
- b) Edit personal information
- c) Post a question
- d) Answer question
- e) Like
- f) Follow
- g) Online chat
- h) Sort questions posted by users by popularity
- i) System notification
- j) Modify problem status
- k) Recommendation related questions
- l) Full text search

- m) Tags management
- n) Web Crawler

#### 6.2.2 Non-functional requirements

- a) Verification code prevents malicious login of robots
- b) The page is simple and clear
- c) Asynchronous processing
- d) Concurrent processing
- e) Sensitive word filtering
- f) E-mail verification
- g) Email reminder
- h) Cookie long-term record user credentials
- i) Highlight search keywords
- j) Monitor project operation status
- k) Authority management
- l) Unified exception handling

#### 6.3.3 Students stories

As a student I want to be able to ...

- Post questions at any time
- View questions and solutions from other users in the past
- Answer questions from other users
- Follow a certain user
- Evaluate the problem or solution (like)
- Search related questions
- Chat online with a certain user
- Edit personal information
- Change the password

#### 6.3.4 Administrator stories

As an administrator, I hope to be able to ...

- System authority control management
- Modify the user type of other users
- Manage existing tags in the system (add, delete, modify, search)
- Manage user posted content (delete, top, add refinement)
- Crawling related questions from StackOverflow
- View the status of crawled data
- View system operating status
- View the system's daily page views and page views within a certain period

### 6.3.5 Extended Functionality

From the point of view of the system, it is extremely important to improve the efficiency of problem solving after users raise the problem. Therefore, the following two methods will be used in this system to improve the efficiency of solving questions:

- Recommendation related questions  
Whenever a user publishes a new problem, the system will recommend it to every user who may be interested in the problem or up to the solution according to the characteristics of the newly posted problem.
- Questions related to users and no solutions  
Each user's expertise is different. According to the user's usage habits, the problem without a solution is displayed in the lower right corner of the page, and the user is reminded to help others to answer the question at any time.

## 6.3 Technologies to be used

Before the system is formally developed, it is particularly important to research and determine the technology used in this project. These include programming languages, development frameworks, databases, caching, and network protocols.

The core technologies of the project will be listed below:

Technology	Use
HTML/CSS/JavaScript	Front-end development language
Java	Server development language
Maven	JAR package management tool
Bootstrap	Front-end rendering
Thymeleaf	Front-end template engine
Spring MVC	System development architecture
Spring Boot	System development framework
Spring Security	System Security Framework
Mybatis	Data Persistence Layer Framework
MySQL	Database (Store users and questions details)
Neo4j	Database (Store the relationship between users and users and the relationship between users and tags)
Redis	Cache
Elasticsearch	Search engine
Kafka	Blocking queue
Selenium	Automatic test framework (Used to crawl data)

Chromedrive	Automatic test engine (Used to crawl data)
WebSocket	Full-duplex communication protocol

The technical solution for deployment is as follows:

Technology	Use
Alibaba Cloud Server (Linux)	Remote server
Tomcat	Project server
Nginx	IP reverse proxy and load balancing

In addition to the above core technologies, many libraries (JavaScript) / dependencies (Java) are also used in front-end development and server-side development, and some other secondary software tools are also provided.( The libraries and dependencies used in the project can be found in the W3C and maven repositories. It is completely open source.)

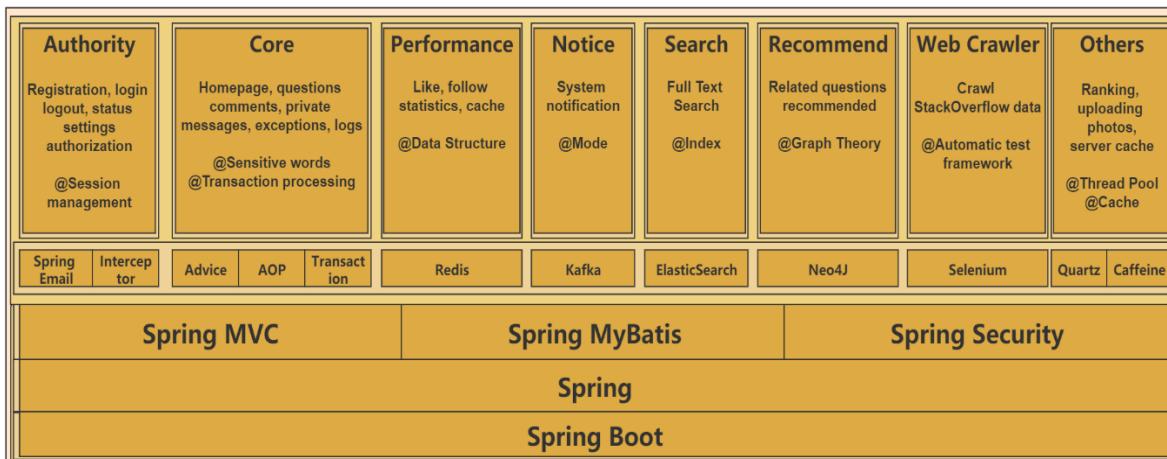
Library/Dependencies	Description
spring-boot-starter-mail	Email Dependence
kaptcha	Verification code generation tool
fastjson	JSON fast conversion format tool
okhttp	Efficient HTTP client
spring-boot-starter-quartz	job scheduling library
qiniu-java-sdk	Qiniu Cloud Service Dependence
Caffeine	High-performance cache library
spring-boot-starter-actuator	System health check library
Editormd.js	markdown format text library
Sockjs.js/stomp.js	WebSocket connection library

## 6.4 End of stage discussion

Determining project requirements and the technologies used are the main tasks at this stage. By understanding and studying the existing solutions in the industry, it helps to determine the general needs of the project and provides a detailed reference for the user interface design of the system. In addition, in order to further determine the system requirements, a questionnaire is used to investigate the target users of the project, thereby making the system requirements more precise and accurate.

## 7.0 Stage 2: System architecture design

The following figure records most of the functions and technologies in this project. In addition, the association between the used technology and functions is also recorded in the figure.



From a technical perspective. The technology used in the project all relies on the Spring Boot framework, but Spring Boot is not the core technology used by the entire project. It mainly plays an auxiliary role in the project, which can reduce the difficulty of using other technologies and reduce the cost of learning new technologies. The main core technology of the project is carried out through the Spring framework. The project uses many related modules on top of Spring technology. For example, use Spring MVC to solve the problem of front-end and back-end interaction. Spring MyBatis is used for the interaction between the system and the database. Spring Security is used for system security issues. On top of these three modules, this project also uses other technologies to develop corresponding functions.

## 8.0 Stage 3: HCI design

### 8.1 Interface design principles

Before starting this stage, the author conducted research on existing solutions in the industry and conducted some research on the design principles that a good user interface should follow. Ben Schneiderman's "Eight Golden Rules for Interface Design" (Shneiderman, 2018) is a set of key principles cited by many academic resources. The following are the principles of the interface design of the system:

- Consistency
- Versatility
- Information feedback
- Operating procedures
- Prevent mis operation

- Allow undo operation
- User control
- Reduce short-term memory load

Follow the above principles as much as possible during the design process. Due to the time limit of the project and some compromises, it may be difficult to fully comply with the above design principles. But they lay a solid foundation for interface design.

## 8.2 User interface

The overall framework of the system front-end page is as follows (for details, please see Chapter 1.1 User interface in Appendix 1):



# 9.0 Stage 4: Design of backend system

## 9.1 Database

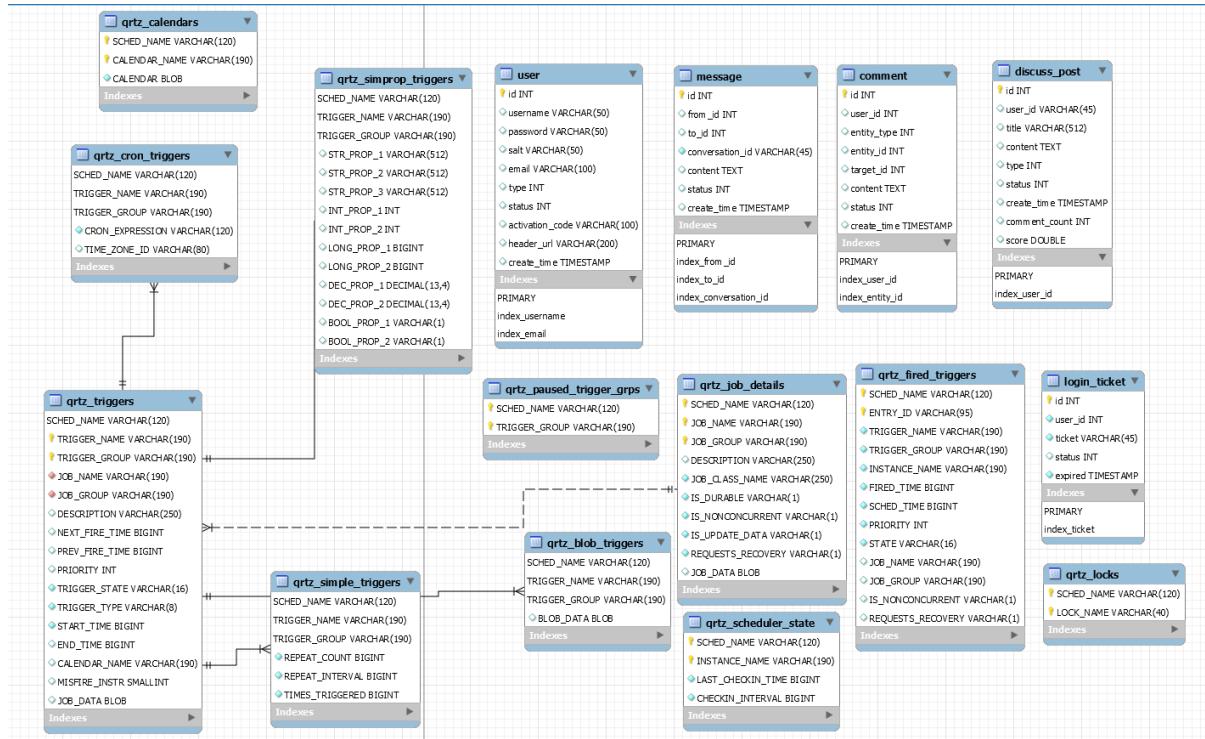
Four different types of databases are used in this project for the realization of different functions. The MySQL database is mainly used to record user information and some comments and other text. The Neo4J database is used for the system's recommendation function, which mainly stores the relationship between users or problems. The Redis database belongs to the in-memory database and has a fast read and write speed, so it acts as a cache function in this system. Elasticsearch database due to its fast and accurate search capabilities. Use it for the system's full-text search function.

### 9.1.1 MySQL

#### *9.1.1.1 ER diagram*

Considering that most of the relationships between system data are stored in the Neo4J database, only the first paradigm in database design is met when storing user

information, messages, and problems, and the database design is satisfied when recording scheduled tasks of the program the third normal form. The overall ER diagram of the database is as follows:



### 9.1.1.2 Data Dictionary

The data dictionary only discusses the data structure of the system data table. The scheduled task table is developed by the manufacturer and has no significance for fixed writing.

- **User table**

Field	Type of data	Description	Defaults
Id	Int (PRIMARY KEY)	User ID	AUTO_INCREMENT
Username	Varchar (50)	Username	NULL
Password	Varchar (50)	Account password	NULL
Salt	Varchar (50)	Increase password complexity	NULL
Email	Varchar (50)	User's email	NULL
Type	Int	0- general user 1- Super Administrator 2-Administrator	0
Status	Int	0- inactivated 1- activated	0

activation_code	Varchar (100)	Used to activate account	NULL
header_url	Varchar (200)	User avatar address	NULL
create_time	Timestamp	Account creation time	NULL

- **discuss\_post table**

Field	Type of data	Description	Defaults
Id	Int (PRIMARY KEY)	Question ID	AUTO_INCREMENT
User_id	Varchar (45)	Author ID	NULL
Title	Varchar (512)	Question title	NULL
Content	Text	Question content	NULL
Type	Int	0- ordinary 1- Sticky	0
Status	Int	0- Normal 1- Essence 2- Delete	0
Create_time	timestamp	Question posted time	NULL
Comment_count	Int	Number of comments on this question	0
Score	Double	Question score	0

- **Comment table**

Field	Type of data	Description	Defaults
Id	Int (PRIMARY KEY)	Comment ID	AUTO_INCREMENT
User_id	Varchar (45)	Author ID	NULL
Entity_type	Int	Types of comments 0- Comment 1- Reply	NULL
Entity_ID	Int	UserID/QuestionID	NULL
Target_ID	Int	UserID/QuestionID	NULL
Content	Text	Comment content	NULL
Status	Int	0- Normal 1- Delete	0
Create_time	timestamp	Comment time	NULL

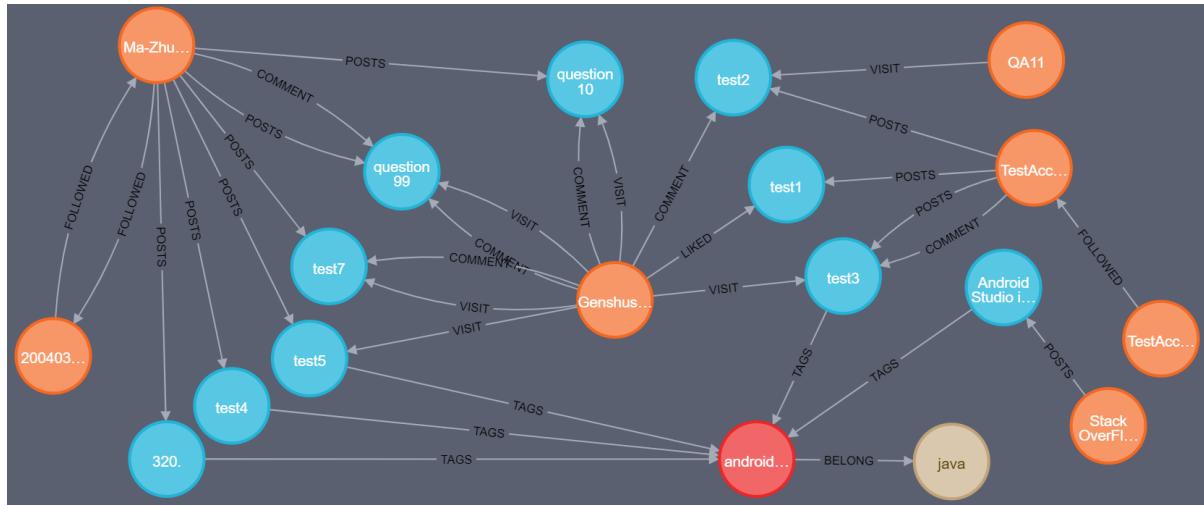
- **Message table**

Field	Type of data	Description	Defaults
Id	Int (PRIMARY KEY)	Message ID	AUTO_INCREMENT
From_id	Int	0- System 1- User	NULL
To_id	Int	UserID	NULL
conversation_id	Varchar (45)	User-to-User Conversation	NOT NULL
Content	Text	Message content	NOT NULL
Status	Int	0- Unread 1- Read 2- Delete	0
Create_time	timestamp	Date of last message	NULL

### 9.1.2 Neo4J

To improve query efficiency and avoid repeated queries affecting system performance. The system also records user information and necessary release content in Neo4J database. The details will be shown below:

#### 9.1.2.1 Node relationship diagram



**Orange:** User node

The user node has five kinds of relationships, they are FOLLOWED, POSTS, COMMENT, LIKED, VISIT. For users and users, their relationship is only FOLLOWED and is bidirectional. But for the relationship between the user and the problem, there are the remaining four, and the relationship is unidirectional, only the user node can point to the question node.

**Blue:** Question node

For the question node, it only has a single item with the tag node and points to the tag node.

**Red:** Tag node

Like the question node, the tag node has only one relationship, BELONG, and points to the category node.

**Brown:** Category node

#### 9.1.2.2 Data Dictionary

- **User node**

Field	Type of data	Description
UserId	Int	UserID
Username	String	Username
Email	String	User email
header_url	String	User avatar address
Status	Int	0- Normal 1- Essence 2- Delete
Type	Int	0- Ordinary 1- Sticky

- **Question node**

Field	Type of data	Description
PostID	Int	Question ID
Title	String	Question title
Content	String	Question content
linkUrl	String	The data obtained from StackOverflow has this attribute, the other does not
commentCount	Int	Number of comments on this question
Userid	Int	Author ID
Type	Int	2- ordinary Sticky
Status	Int	3- Normal 4- Essence Delete
Score	Double	Question score

- **Tag node**

Field	Type of data	Description
Tagname	String	Tag name

- **Category node**

Field	Type of data	Description
Name	String	Category name

### 9.1.3 Redis

In this system, the Redis database is mainly used as a cache, and it is also used to prevent duplicate data from crawling on the network. The data dictionary will be displayed below:

Field	Type of data	Description
like:entity:1(entityId)	Set	Like information
like:user:1(userid)	Set	Like information
Followee:1(userid):1(entityId)	Set	Follow information
follower:1(entityType):1(entityId)	Set	Follow information
Kaptcha:xxxx(string)	Set	Verification code
Ticket:xcff103ASF21sdf(access token)	Set	User login credentials
User:1(userid)	Set	User information cache
Post:12.32(score)	Set	Question detail cache
Crawler: http://www.sdf.com (linkUrl)	Set	Prevent repeated crawling of data
Uv:2020-01-02(data)	HyperLogLog	Record daily visits
Uv:2020-01-02(start date):2020-01-03(end date)	HyperLogLog	Record the number of visits in a certain period
DUA:2020-01-02(data)	Bit	Record the number of visits of different IP
DUA:2020-01-02(start date):2020-01-03(end date)	Bit	Record the number of visits of different IPs in a certain period

### 9.1.4 Elasticsearch

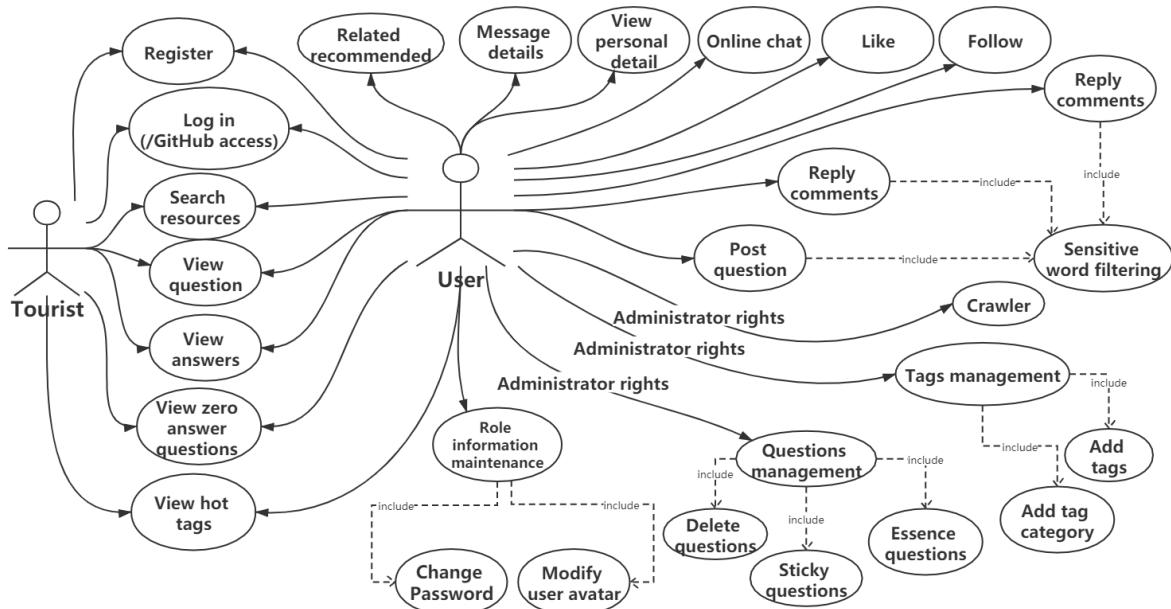
Field	Type of data	Description
PostID	Int	Question ID
Title	String	Question title
Content	String	Question content
linkUrl	String	The data obtained from StackOverflow has this attribute, the other does not
commentCount	Int	Number of comments on this question
Userid	Int	Author ID
Type	Int	0- Ordinary 1- Sticky

Status	Int	0- Normal 1- Essence 2- Delete
Score	Double	Question score

## 9.2 Server

### 9.2.1 Top use case diagram

The top-level use case diagram is used for the development of early system functions, and to determine whether the system meets the requirements during the acceptance stage.



### 9.2.2 Class diagram

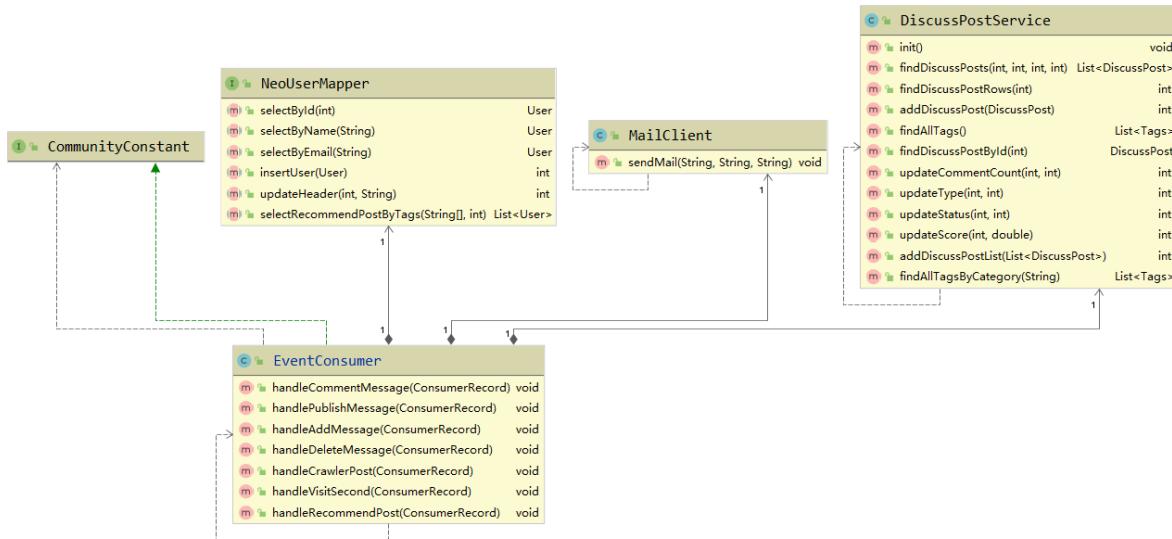
In this title, only the class diagram of the main functions of the system is shown. For the overall project diagram, please visit this address:

<https://raw.githubusercontent.com/Ma-Zhuang/Pic/master/ClassDiagram.png>

The overall structure of the project code is divided into three levels, namely the control layer, the service layer, the guide layer. The function of the DAO layer is to interact with the database to obtain the data required by the current function, and the service layer is to perform calculation or logical judgment on the obtained data to provide the required data for the controller layer. The controller layer provides the data to be displayed for the front-end page and controls the interaction of the front-end page. The call relationship between these three layers is Controller-> Service-> DAO. (For all details, please refer to Appendix 4: Design).

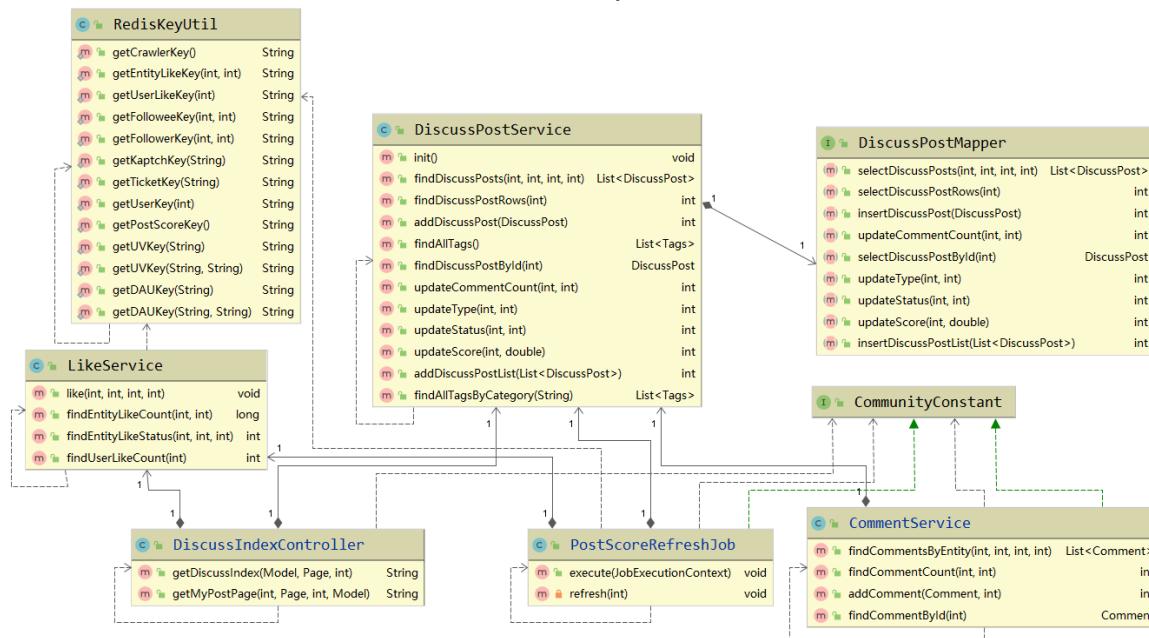
## Recommended function

The core technology used by this function is the message queue and graph database. When a student cannot post a new question, Kafka's recommended topic will be initiated to achieve an asynchronous effect and improve the performance of the system. After triggering the message queue, the system will use Neo4J's shortest path algorithm to match users who have a relationship with the question tags, thereby implementing the recommendation function.



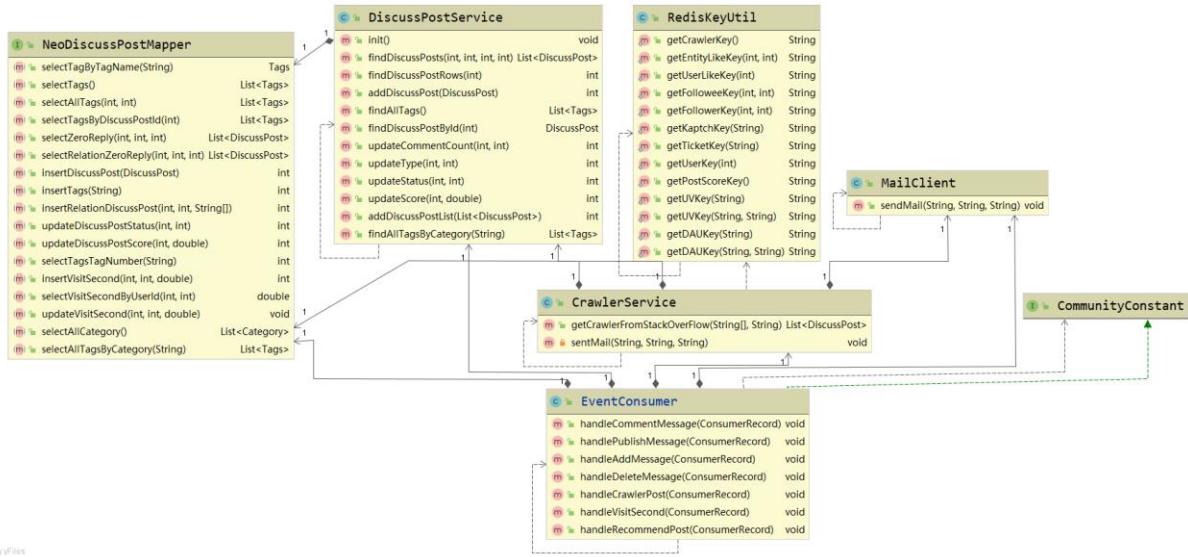
## Popularity

After a question is commented or liked, it is first recorded in Redis and then set to update the corresponding score of the changed question every 5 minutes through a scheduled task, to achieve the rank of the question.



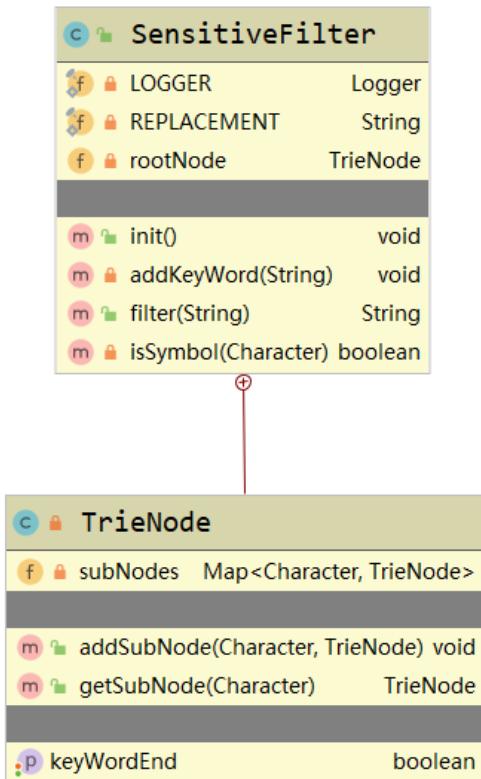
## Crawler

For the crawling function, after the administrator inserts the tag, the system will search for the related questions of the tag in StackOverflow and sort the questions in reverse order by the number of Vote, and then obtain the problems of the existing solutions. The system data is filled. The set data structure in Redis is used to prevent duplicate data from being obtained and improve system availability.



## Sensitive word filtering

All content posted by users, including question titles, question contents, comments and replies will be filtered through sensitive words and stored in the database, thereby reducing waste of human resources.



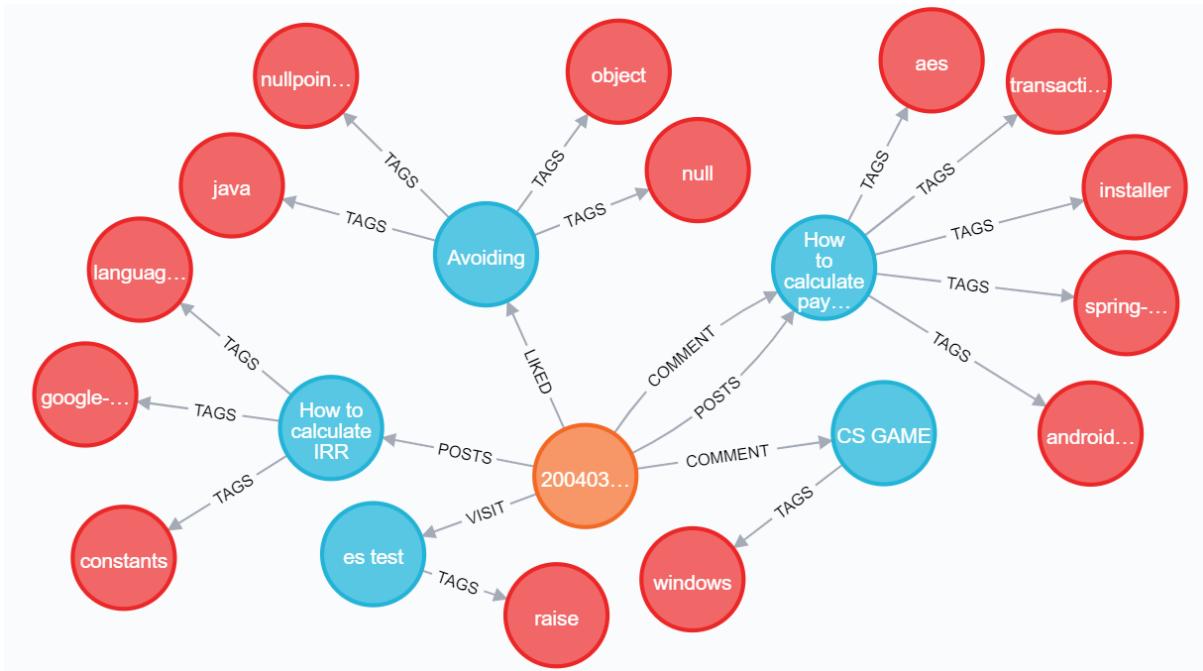
## 10.0 Stage 5: Implementation

### 10.1 Recommended function implementation

The implementation of the recommendation function of this project is mainly based on the shortest path algorithm in graph theory. Establish the connection between the question and the question tags and the users through the user behaviour recorded in the database. Since the project is public, there is no guarantee that the user's behaviour is accurately recorded. It is possible that the user only accesses the question and does nothing. Therefore, user behaviours are defined as four behaviours: visit, like, comment and post. By default, when the user access question exceeds 20 seconds, the relationship of this visit is recorded in the database.

As shown in the following figure, all the labels (node colour is red) related to the user (node colour is orange) are queried by the shortest path algorithm provided in the neo4j database, then the result is that the user may be interested or know how

Problem solved tags. **Cypher query:** " MATCH (u:User{UserId:103}) optional match (t:Tags), p=shortestpath((u)-[\*..2]->(t)) RETURN p"



After knowing how to find out which tags the user may be interested in, founded that using this method to recommend, the database will be under great pressure when the amount of data is extremely large, and there may be cases where the database service is down. . Therefore, the query method needs to be improved.

If a user issues a new question, the system only needs to obtain the label of the question, and then find out which users are related to the question's label in the database, then the system will think that the problem may be found Users are interested.

For example, if the tag of the newly posted question by the user is "raise", then use **Cypher query** "match (u: User)-[\* .. 2]-> (t: Tags) where t.tagName in [ " raise " ] and u.UserId <> 3 return distinct u.UserId as id, u.Username as username, u.email as email ", system can get all the user information related to the tag, and then notify the user by email to achieve the recommended features. The query results are as follows:

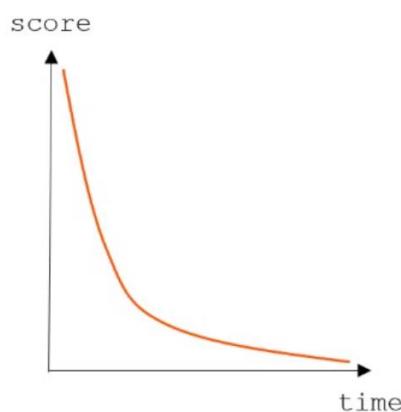
<b>id</b>	<b>username</b>	<b>email</b>
103	"20040330"	"444931946@qq.com"
105	"Genshushu"	"2686224016@qq.com"
4	"TestAccount1"	"test1@qq.com"

## 10.2 Popularity implementation

Through the investigation of the existing solutions in the industry, including the heat algorithm of "Hacker News" and "StackOverflow". After research, it is found that all the algorithm images used in the existing solutions in the industry are correspondingly reduced in time and the function curve is extremely smooth. Therefore, this project also reduces the score of the time problem evenly with the increase of time. The function image is as follows:

**The heat popularity of this system:**

$$\text{Score} = \log(\text{essential score} + \text{number of comments} \times 10 + \text{number of likes} \times 2) + (\text{post time of issue} - \text{start time of the project})$$



- **Hacker News' popularity algorithm**

$$\text{Score} = \frac{(P - 1)}{(T + 2)^G}$$

P represents the number of votes for this article

T represents the time interval from the time the article was published to the current time

G represents the activity coefficient of the article, usually 1.5 or 1.8

- **StackOverflow's popularity algorithm**

$$\text{Score} = \frac{(\log(Qviews) \times 4) + ((Qanswers \times Qscore) \div 5) + \text{sum}(Ascore)}{(QageInHours + 1) - ((QageInHours - Qupdated) \div 2)^{1.5}}$$

Qviews represents the number of views of the question

Qanswers represents the number of answers to this question

Qscore represents the difference between the number of likes and the number of dislikes on the question

Ascore represents the difference between the number of liked and disliked answers to the question

QageInHours represents the time interval from the time the question was published to the current time

Qupdate represents the answer time of the latest answer to the question

Considering that operations such as likes, and comments are all high-frequency actions. If the database is updated every time such an operation is performed, the pressure on the database will be very high, and data loss will also occur. Therefore, it is necessary to use the timed task method to update the changed problems at regular intervals. Each time user like or comment, system need to store the corresponding question in the cache. When the scheduled task is started, only the corresponding question score in the cache is updated. This greatly improves the efficiency of the algorithm, reducing unnecessary calculations and database pressure.

### **10.3 Crawler implementation**

Due to copyright issues, the data crawled by this system is only the title of the problem and the link address. In addition, with the continuous improvement of the anti-reptile system, it is now more and more difficult to crawl data, and it is difficult to obtain website data by calling APIs and downloading pages. Therefore, this system uses automatic testing framework Selenium to achieve data crawling. The data is obtained through Selenium imitating the way of artificially browsing the website and requires the use of ChromDriver.

After the administrator enters the label, the system will automatically query StackOverflow for related questions that contain the label, and at the same time, it will be arranged in reverse order according to the number of Vote and 50 questions will be displayed on one page. When crawling data, the first 15 pages of internal content are crawled, and the crawled content must contain the solution to the questions, or it will not be obtained by the system.

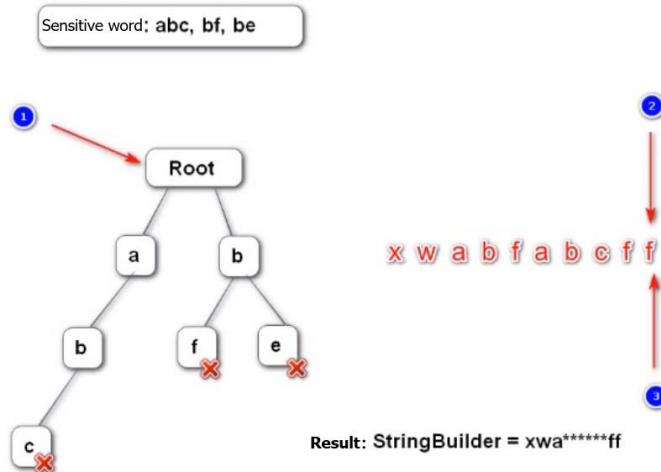
In addition, the crawler function of this system is to avoid repeated crawling of the same data through Redis. After each crawl is completed and before it is stored in the database, the link path of the problem will first be stored in the Redis set data structure. If the next time administer crawl data, Redis has the same problem link path, the system will automatically skip the problem and will not save it to the database.

All external resources in the system will be clearly marked with copyright ownership.

### **10.4 Sensitive word filtering implementation**

The first sensitive words defined are "abc", "bf" and "be". The system will automatically filter out the word drug and replace it with three asterisks.

The function of filtering sensitive words is to match the content published by the user through the defined prefix tree and determine the filtered string through three pointers. Take the first letter of each word as the first level of the prefix tree, and then create child nodes in turn, and mark a sensitive word on the last node. As shown in the picture.



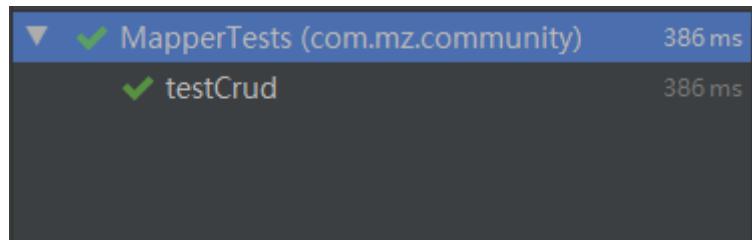
Before the string is stored in the database, three pointers are used to determine the position of the sensitive word in the string, to achieve the filtering effect. First, the character pointed by the second pointer will first be compared with the character defined in the prefix tree. Found that the character x does not exist in the prefix tree, so pointers 2 and 3 point to the next character at the same time. Similarly, the character w does not exist in the prefix tree, so the pointers 2 and 3 point to the next character at the same time. We record the characters that are not sensitive words in "StringBuilder". The character a can be matched in the prefix tree, so the pointer 1 will point to the child node a, the pointer 2 will not move, the pointer 3 will point to the next character, and the character pointed by the pointer 3 will point to the pointer 1. The next child node of the node is matched. Character b is found in the prefix tree, so pointer 1 needs to point to the next node, pointer 2 doesn't move, and pointer 3 points to the next character. However, the character f does not exist in the prefix tree, so it can be determined that the string between pointer 2 and pointer 3 is legal. So now it needs to move pointer 1 to the root node, pointer 2 to the next bit, and pointer 3 to the character pointed to by pointer 2. Then match the character with the prefix tree in the same way as before. When the pointer 3 points to the character f, it is determined that the character string between the pointer 2 and the number 3 is the defined sensitive vocabulary. So, replace this string with an asterisk and record it in "StringBuilder". The subsequent character filtering is the same method, so that the function of filtering sensitive words is completed.

## 11.0 Stage 6: Integration and test

### 11.1 Unit testing

In the process of project coding, the TDD mode is generally adopted for development, that is, after each API is written, it can be applied to the system after all tests. The tool used for unit testing is Junit. For example, all the APIs for adding,

deleting, modifying and checking the database can only be called by other classes or functions after passing the test. In general, all APIs of the system have passed the test. The test results are as follows (For all details, please refer to Appendix 5: Test):



## 11.2 Functional Testing

After the development of each function is completed, perform a functional test on each function. The test method is to run the server locally and set the expected test results in advance. Record the return result of each test during the test, and finally compare with the designed expected result to determine whether the modified function passes the test. In general, all functions of the system have passed the test. (For all details, please refer to Appendix 5: Test)

## 11.3 Usability Testing

After the project development is completed, deploy it on the remote server. Invite other students to conduct overall system testing in strict accordance with the test plan. The test plan includes all the functions of the system, the speed of execution and the performance of the system. After completing the test at this stage, a very important problem was discovered. When the data is stored in the database after the data is crawled, it cannot be inserted into the database in batches because of the huge amount of data. Eventually, the crawler function will not work properly. After research, it was decided to insert into the database in batches and multiple times. Although this efficiency is lower than that of inserting a large amount of data at once, the overall performance has not been affected, and it also ensures that the data will not be lost. (For all details, please refer to Appendix 5: Test)

## 11.4 Verification and Validation

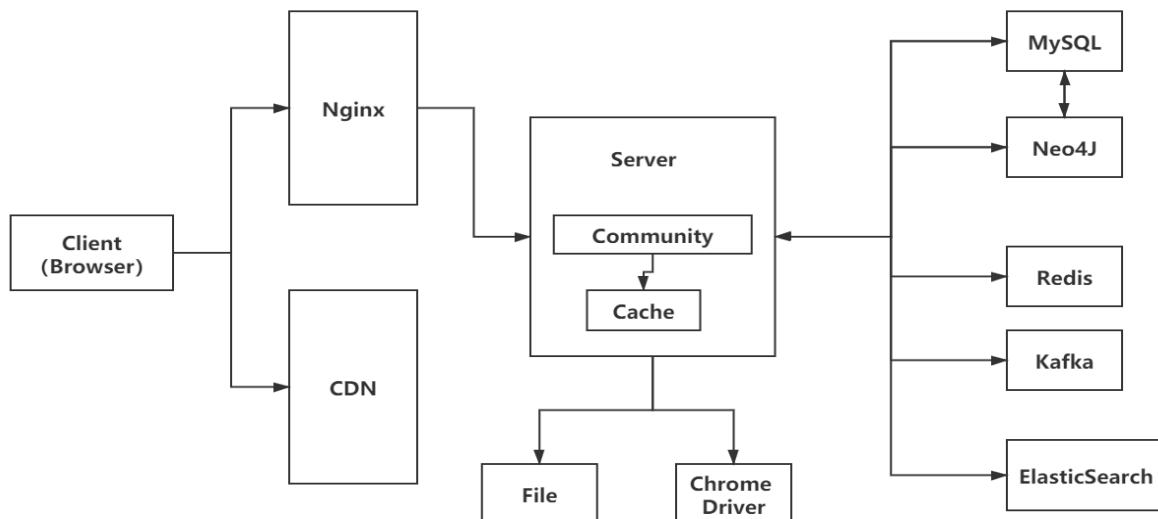
The verification and confirmation of the project is extremely important, these two points will completely determine the success of the project. The focus of Validation is to check whether the function of the system fully meets the needs of customers. The focus of verification is to monitor the quality of software to ensure that the quality of the project can meet the needs of customers (Easterbrook, 2010). Every sprint stage of the project will be verified to ensure that it can smoothly enter the next stage of development.

- Verification
  - 1. Reviewing code

2. Reviewing requirements and specification
  3. Functional testing
  4. Lots of unit test
- Validation
    1. UML following requirements elicitation
    2. Reviewing personas
    3. Functional testing against backlog
    4. Usability testing

## 12.0 Stage 7: Deployment

The deployment method used in this project is to deploy through the Alibaba Cloud server. After purchasing the domain name and server, deploy the project code to the server. The server deployment environment is a Linux system. The structure of the project in the server is as follows:



This project provides two types of services there are two types of user requests. One is for dynamic requests, such as question lists, and the other is for static requests, such as pictures or styles. When the type requested by the user is a static type, the system will access the CDN service for corresponding processing. When the type of user access is dynamic, the corresponding request to the Alibaba Cloud server is accessed.

## 13.0 End-project report

### 13.1 Summary

To summary, the purpose of the project is to create a solution to help students improve the efficiency of self-study. The system is a Web-based visualization system, so that it can get more solutions by publishing the problems encountered. At the same time, it also provides freshmen with a deep understanding of this course and the problems they may encounter. In order to determine whether the student is suitable for this course. This project is generally successful, but there are still some areas that need improvement and the function updates or additions of subsequent versions.

### 13.2 Project Objectives

This chapter will review the objectives to judge whether it has fully achieved the project objectives.

**1. Analyse existing Q & A systems and query literature about recommended algorithms to determine project requirements and efficient algorithms in general.**

At the beginning of the project, as part of the investigation and analysis phase, the goal was analysed. Through research on the existing question answering system, it has been found that they have a common working method and what convenient functions are provided for users. The systems studied include Google Answer, Quora, Techcurnch, StackOverflow and brainly. Analyse and match the interaction mode and existing functional characteristics of these systems. At the same time, through the literature research on the heat ranking algorithm and recommendation algorithm has in-depth research. Of course, legal and moral issues are also the focus of this goal, which is mainly to have a further understanding of copyright law. The latter goal is successful, after the analysis, the developers have a deeper understanding of people's needs.

**2. Further understand user needs through questionnaires to further clarify project needs.**

The main purpose of this goal is to determine project requirements. Judging from the previous goals, it is only the developers who have a rough but not detailed understanding of the system requirements. Although this is sufficient to meet the research needs of the project, there are two roles for this system, namely the role of administrator and the role of ordinary users. Through the questionnaire, author can further analyse more details of the needs of users with different roles. This goal was very successful. After 50 questionnaires were distributed, many details of the needs were obtained through analysis. This is of great help in determining system requirements, but also reduces unnecessary functional design and improves development efficiency.

**3. Analysis and selection of development technologies, deployment options and implementation.**

- Technologies

After the first two goals were completed, the development technology used in the project was determined through an analysis of needs. The analysis includes several aspects. The environmental requirements, performance, learning costs, and ecological environment of the technology finally consider the use of Java technology for development.

- Implementation

The final system basically met the originally formulated requirements and achieved overall success.

- a) Allow students to get solutions to the question by asking questions.
- b) Allows the system to recommend problems that may be of interest to users through recommendations, to achieve the effect of improving the question solving rate.
- c) Allow the system to obtain resources from the Internet to fill in system data.

- Deployment

After analysed the price, system performance, location and other factors of the remote server, the Alibaba Cloud server was finally selected as the project deployment environment. Considering that the stability of the Linux system is much better than that of the Windows system, the server system environment deployed is Linux.

Although this goal has been successful, it still has some shortcomings. For example, the performance of the system needs to be improved.

### **13.3 Changes to original plan**

#### **13.3.1 Requirements**

There is no recommended function or crawler function in the original requirements list. Consider the number of users and the system utilization rate later. Consider adding recommended functions and crawler functions. In the original version, when the recommendation function is not added, the problem posted by a certain user may be difficult to be solved quickly. Moreover, the system is a Web-based application. When users do not have access to the system, other users 'questions cannot be answered quickly. Therefore, users need to be reminded through external means. Considering the practicality, finally chose to remind relevant users by email. Secondly, considering that the system has less data, it is difficult for users to find related problems through search. This may reduce the efficiency of student self-study. So, use the crawler function to get more popular and high-ranking questions in StackOverflow. This reduces the number of repetitive and unnecessary questions posted by students.

#### **13.3.2 Technologies**

Throughout the development process, most of the main technologies have not changed. However, due to the increase and change of demand, the technology also needs to make corresponding changes.

When developing the recommended function, the original technology was to use MATLAB to process the algorithm model, but due to its complexity, it could not fully comply with the development of this system. So, replace it with Neo4J database for the basis of recommendation function. The advantage of using Neo4J database is that it is simple and easy to use, and the algorithm to achieve the function, the time consumed is basically the same.

In addition, when developing the crawler function, the IP address is easily disabled by StackOverflow due to the HttpClient technology, and the learning cost is relatively high. The Selenium automatic testing framework is now used in conjunction with the Chrome Driver to imitate human operations for data crawling. It's simple and easy-to-use features also save most of the time for development.

## **14.0 Project post-mortem**

### **14.1 What went well**

As mentioned in the previous sections, the project was generally considered successful and met most of the functional goals and all project goals set at the beginning of the project. A large part of the reason for the success of the project is due to a clear understanding of the appearance and function of the required system at the beginning of the project. A clear goal means that the design phase needs to be very specific, with well-defined challenges. These are the correct goals of the project.

The selected development technology is the correct technology in most cases. Although some are new technologies that require time for the author to master, they are generally successful. In the development stage, due to the dependency management tool Maven, it greatly saves the time cost of importing dependencies and managing dependencies. At the same time, the integration effect of the Spring Boot framework to other technologies is also very significant. MyBatis integrates with Spring Boot to use the existing API to operate the database, which significantly improves the development efficiency and reduces a lot of similar logic code.

Project management is also largely the key to project success. Set clear deadlines for each part from the beginning of the project to avoid "nothing to do". At the same time use Trello to create visual progress for all tasks. This means that a lot of energy and time can be spent on development without having to consider the next step.

### **14.2 What could have been improved**

For this project may need to improve in some respects, these issues will be crucial for future version updates. These problems are roughly divided into the following aspects:

- System performance**

Although this system uses Redis as a cache to improve system performance, there are still some functions that cannot meet the conditions of the cache and cannot improve performance. For example, the question list page, the data presented on this page comes from MySQL and Neo4J databases. In the actual testing process, it was found that the query performance of MySQL caused a long time to load the page. Even if the data query is performed by creating an index, the database query performance will be degraded when the amount of data reaches one hundred thousand.

Through inquiring information and research, it is found that the method of improving this situation can be upgraded through the way of separate reading and writing of the database. Since the author has no experience in the development and deployment of distributed databases, using this method will

increase the development time and may eventually lead to the project not being completed. The author hopes to learn in depth and improve system performance in the future.

- **WebSocket**

For the current system, only WebSocket is supported on the chat page. When the user receives the system message, he cannot receive the message in the station in time. The latest news can only be obtained by refreshing the page. The way to improve this situation can be through the WebSocket technology, so that the new data generated by the back end is automatically pushed to the front end. But this may require a message relay station, so that all users subscribe to this message relay station. Thus, the function of automatically pushing back-end data to the front-end is realized. However, considering the system performance, when the number of users increases, the message relay station must be distributed in the server. Since the author has no relevant development experience, it will increase the development cycle of the project. Eventually it may cause the project to fail, author hope it can be upgraded later.

- **Selenium**

At this stage, the system uses the Selenium automatic test framework to realize the crawler function. Because its flexibility is not high, if the StackOverflow page changes or upgrades, its data cannot be crawled. This requires reanalysing and recoding StackOverflow pages. This caused a lot of trouble for the later maintenance of the system. Since the author has not found a solution with high performance and easy to use, this technology is temporarily used for the current version.

### **14.3 Possible extensions to functionality**

The initial goal of the project is to provide services for students of all majors in a university. However, in this project, due to development time issues, its goal was narrowed to students in computer-related majors. Hope to add support for other professional students in subsequent versions. In addition, the visual monitoring interface for existing projects is not perfect, and the completed visual monitoring interface should be provided in subsequent versions. In addition to this, a user evaluation mechanism should be added to the answers provided by the user. For some high-quality answers, the system should automatically highlight the reminder or place its answer at the top of the comment area. Finally, the system should provide students with a dedicated discussion group for each course for after-class discussions to increase the number of exchanges between students.

## **15.0 Conclusion**

The project aims to create a self-learning platform for students to improve their self-learning efficiency and question-solving efficiency during the self-learning process. The project is generally successful and almost meets all functional requirements. It is now deployed and used on a remote server. This project is a challenging project for the author. Because many new technologies are used to improve system availability and system performance, such as Nginx reverse proxy, Redis cache, Message queue (Kafka), Spring Security framework, multi-threading, high concurrency and crawler systems. The purpose of using these technologies is to hope to be as close as possible to the actual enterprise development process and development technology after the project development is completed. At the same time, author hope that by extending the development time to add or expand some functions, to bring real benefits to users as much as possible.

## 16.0 Reference list

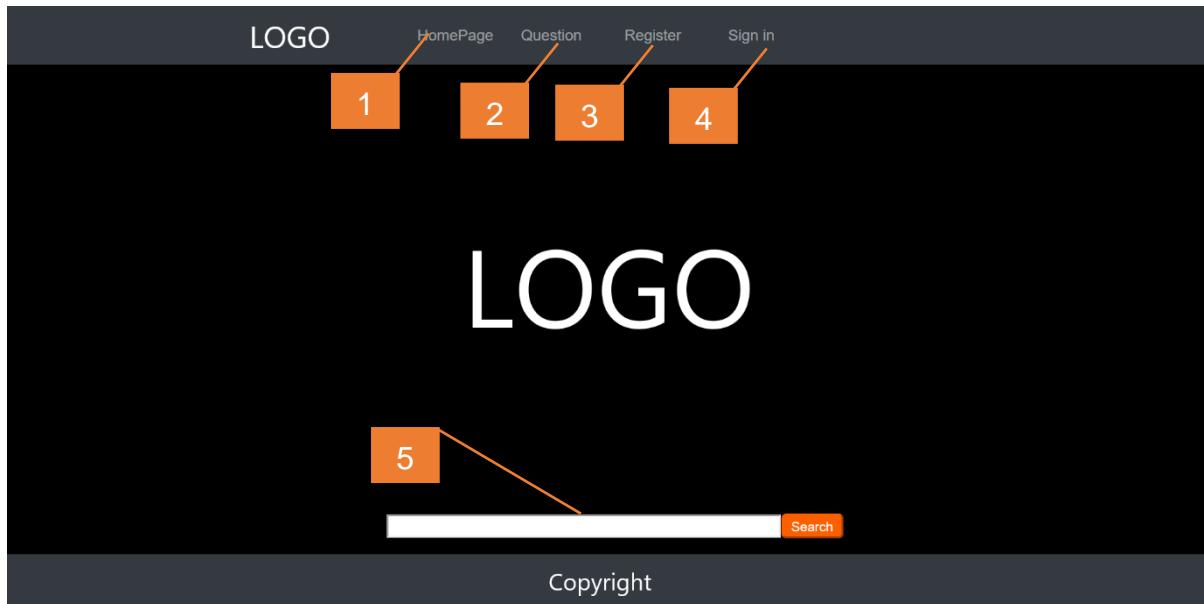
- Agile Alliance. (2017). Agile Manifesto for Software Development. [online] Available at: <https://www.agilealliance.org/agile101/the-agile-manifesto/> [Accessed 01/04/20]
- Easterbrook, S. (2010) The difference between Verification and Validation, Serendipity. [online] Available at: <http://www.easterbrook.ca/steve/2010/11/the-difference-between-verification-and-validation/> [Accessed 04/05/20]
- Kearns, M., Rankins, G. (2008). Integrating PRINCE2 and Scrum for successful new product development. Australian Institute of Project Management National Conference.
- GitHub, Inc. (2020) GitHub Registered Developer Agreement. [online] Available at: <https://help.github.com/cn/github/site-policy/github-privacy-statement> [Accessed 02/05/20]
- Gov.uk. (2017). Data Protection. [online] Available at: <https://www.gov.uk/data-protection> [Accessed 03/04/20]
- Radigan, D. (2017). Scrum | A brief introduction. [online] Available at: <https://www.atlassian.com/agile/scrum> [Accessed 01/04/20]
- Stack Exchange Inc. (2020). Public Network Terms of Service. [online] Available at: <https://stackoverflow.com/legal/terms-of-service/public> [Accessed 02/05/20]
- Shneiderman, B., 2018. The Eight Golden Rules of Interface Design. [Online] Available at: <https://www.cs.umd.edu/users/ben/goldenrules.html> [Accessed 02/05/20]
- Scrum Inc. (2017). Sprint Backlog. [online] Available at: <https://www.scruminc.com/sprintbacklog/> [Accessed 01/04/20]

## 17.0 Appendices

### Appendix 1: User Guide

#### 1.1 User interface

##### Home page



Number	Features
1	Jump to homepage
2	Jump to the question page
3	Jump to the registration page
4	Jump to login page
5	Search all resources in the system

## Question page

The screenshot shows a question page with the following features highlighted:

- 1**: Newest (Sort by latest questions)
- 2**: Most Popular (Sort by the most popular questions)
- 3**: Header picture (User avatar, click to jump to user details page)
- 4**: Header picture (User avatar, click to jump to user details page)
- 5**: Header picture (User avatar, click to jump to user details page)
- 6**: Author name and question posted data
- 7**: Hot Tags (Most used tags)
- 8**: 0 Answer (Zero reply questions related to users)
- 9**: Pagination (Navigation links: Home, Previous, 1, 2, 3, Next, Last page)

Copyright

Number	Features
1	Sort by latest questions
2	Sort by the most popular questions
3	User avatar, click to jump to user details page
4	Question related tags
5	Question title, click to jump to the question details page
6	Author name and question posted data
7	Most used tags
8	Zero reply questions related to users
9	Pagination

## Question detail page

The screenshot shows a question detail page from a Q&A website. The page has a dark header with a logo, search bar, and navigation links. The main content area displays a question and its details, followed by a related question, a comment section, and a footer.

- 1**: Points to the "Question title" which includes a "Header Picture" and a "Username".
- 2**: Points to the "Question content".
- 3**: Points to the "Related Question" section, which also includes a "Question title" and "Question subtitle".
- 4**: Points to the "Comments" section, showing a comment by a user named "Username" with "1#". It includes a "Comment" button, a reply count, and a comment input field.
- 5**: Points to a large empty text area at the bottom of the page, likely a placeholder for user comments or a feedback form.

**Copyright**

Number	Features
1	Question title
2	Question content

3	Related question
4	User comment (Username, time, content)
5	Post a comment

## Register page

The screenshot shows a 'Sign Up' form on a dark-themed website. The form fields are labeled as follows:

- Account: Please enter your account number!
- Password: please enter your password!
- confirm password: Please enter the password again!
- E-mail: Please enter your email!

Callouts numbered 1 through 4 point to these respective fields. A large teal 'Sign Up' button is at the bottom right.

Number	Features
1	Registered username
2	Password
3	Registered email address
4	Click to complete registration

## Sign in page

The screenshot shows a 'Sing In' form on a dark-themed website. The form fields and features are labeled as follows:

- Account: Please enter your account!
- Password: please enter your password!
- Verification code: please enter verification code! (with a 'refresh' button next to it)
- remember me (checkbox)
- Sign In button
- GitHub Login button
- forget password? (link)

Callouts numbered 1 through 5 point to the verification code field, the 'remember me' checkbox, the 'Sign In' button, the 'GitHub Login' button, and the 'forget password?' link respectively.

Number	Features

1	After clicking, record user information for 30 days
2	Randomly generated verification code to prevent malicious login of robots
3	Click to change password by email
4	Enter the above information and click login
5	Login with GitHub account

## Message page

### Letter

Copyright

Number	Features
1	Jump to system message
2	Send a private message to a user

## System notification

The screenshot shows a dark-themed application interface. At the top, there's a navigation bar with tabs for "HomePage", "Question", "Message", and "Header Picture". Below the navigation bar is a search bar and a "Search" button. The main content area is titled "System notification". It displays three entries, each with an orange numbered callout (1, 2, or 3) pointing to its respective icon:

- 1 Comment:** An icon of a speech bubble with a plus sign. The message says "Username Commented on your article ...". The timestamp is 25-04-2020 21:29:48. A note at the bottom right says "A total of 2 messages".
- 2 Likes:** An icon of a thumbs-up. The message says "Username Liked your article ...". The timestamp is 25-04-2020 21:30:56. A note at the bottom right says "A total of 1 messages".
- 3 Follow:** An icon of a heart. The message says "Username followed you ...". The timestamp is 07-05-2020 05:56:59. A note at the bottom right says "A total of 1 messages".

Number	Features
1	Click to view comment details
2	Click to view likes details
3	Click to view followed details

## User profile page

The screenshot shows a dark-themed application interface. At the top, there's a navigation bar with tabs for "HomePage", "Question", "Message", and "Header Picture". Below the navigation bar is a search bar and a "Search" button. The main content area shows a user profile for "Ma-Zhuang". It includes a header picture placeholder, the username "Ma-Zhuang", and the text "Posted". Below this, it says "Registered at Sun Apr 12 08:45:31 CST 2020". At the bottom of the profile section, there are three statistics with orange numbered callouts (1, 2, 3) pointing to them:

- 1 Followed 1 people**
- 2 Fans 1 people**
- 3 Got 1 likes**

At the very bottom of the screen, there's a dark footer bar with the word "Copyright".

Number	Features
1	Click to jump to the list of followed users
2	Click to jump to the fan list of this user
3	Number of likes received by this user
4	List of questions posted by this user

## Account setting page

The screenshot shows a user interface for account settings. At the top, there is a navigation bar with links for 'HomePage', 'Question', 'Message', and a profile icon. A search bar is also present. The main content area has a dark background with white text. There are two orange numbered callouts: '1' points to a 'Choose a picture' input field labeled 'Avatar:' with a 'File' button; '2' points to a 'Change Password' section containing three input fields: 'Original password:', 'New password:', and 'Confirm password:', each with placeholder text like 'Please enter the original password!'. A large teal 'Upload' button is located below the picture input, and a teal 'Save' button is at the bottom right.

Number	Features
1	Modify user's header image
2	Change user password

## Search page

The screenshot shows a search results page. At the top, there is a navigation bar with links for 'HomePage', 'Question', 'Message', and a dropdown menu. A search bar is also present. The main content area has a dark background with white text. There are four orange numbered callouts: '1' points to the 'Key word' input field in the search bar; '2' points to a 'Header Picture' thumbnail; '3' points to a 'Tag name' tag; and '4' points to a 'Likes 1 | Reply 0' link. Below these, a search result for a question is shown, including the question text 'key word XXXXXX XXXXXXXX XXXXXXXXXX XXX', the author 'Username', the publication date 'published on Wed May 06 05:41:25 CST 2020', and navigation links 'Home', 'Previous', '1', 'Next', and 'Last page'. The footer contains the text '© 2020 by Zhuang Ma(10664074)'.

Number	Features
1	Query keyword input box
2	Highlighted keywords
3	Tag name
4	Number of replies and likes on this question

## Tags management page

The screenshot shows a web application interface titled "Tag Management". At the top, there is a navigation bar with links for "HomePage", "Question", "Message", and "Header Please". A search bar and a "Search" button are also present. Below the navigation, the main content area has a header "Tag Management". There are four input fields labeled "Tag name" (numbered 1, 2, 3, 4). To the right of these fields are two blue buttons: "Add tags" (numbered 5) and "Add Category". Below the input fields is a pagination bar with buttons for "Home", "Previous", "104", "105", "106" (highlighted in blue), "Next", and "Last page". Number 5 points to the "106" button in the pagination bar.

Number	Features
1	Tag name
2	Click the button to add tags
3	Click the button to add a tag category
4	The number of times this tag is used
5	Pagination

### 1.2 Build project

Source code acquisition address: <https://github.com/Final-Year-project-10664074-MaZhuang/ComprehensiveCommunityFinal>

#### System operating environment and necessary services:

(Install and configure the service)

- a) Windows
- b) IntelliJ IDEA
- c) Java JKD 1.8
- d) Maven 3.6.3
- e) Tomcat 9.0.31
- f) MySQL 8.0
- g) Neo4J 3.5.14 community
- h) Redis 3.2.100
- i) Kafka 2.12-2.3.0
- j) ElasticSearch 6.8.6
- k) IK-Analyzer v6.8.6
- l) Chrome Browser (No version required)
- m) Chrome driver (Version corresponding to chrome browser)

## **Build project**

- 1) Clone source code from GitHub to local in Windows system environment.
- 2) Import the database initialization scripts in the root directory of the project into MySQL and Neo4J databases respectively.
- 3) Modify the database configuration (user database user name and password) file path in application-develop.properties: src / main / resources / application-develop.properties
- 4) Configure Kafka, modify the values of the dataDir variable and log.dirs variable, and change the Zookeeper data storage address and Kafka log output address (can be freely changed). The file path is: kafka\_2.12-2.3.0 / config / server.properties (kafka); kafka\_2.12-2.3.0 / config / zookeeper.properties (Zookeeper).
- 5) To configure ElasticSearch, you need to decompress the IK tokenizer to this path: elasticsearch-6.8.6 / plugins
- 6) Use the maven command to package the project into a war file. Then put it into the webapps directory of tomcat.
- 7) Start related services, including MySQL, Neo4J, Redis, kafka, ElasticSearch, Tomcat.
- 8) Open the browser and enter localhost: 8080 to use the system.

### **Note:**

1. When running locally, it is not possible to log in with Oauth2.0 due to the official GitHub API. The system will automatically jump to this project that has been successfully deployed on the remote server.
2. Due to the remote CDN object storage space, the address changes every month. If the picture cannot be loaded or the picture cannot be uploaded, please visit the address of the successfully deployed project ([www.fyp10664074.com](http://www.fyp10664074.com)). The author regularly updates the code on the remote server.

## Appendix 2: Project management

### 2.1 Trello boards

#### 2.1.1 Trello boards 1

The image shows a Trello board titled "PRCO304 Computing Project". The board has three lists: "Todo", "Doing", and "Done". The "Todo" list contains cards for "Requirement", "Development", and "Test". The "Doing" list contains cards for "Background research", "Front end", "Back end", "Databases", "Search", and "Project management tool". The "Done" list contains a single card: "Determine the technology needed". A modal window is open in the center, showing the "Done" list with the same card. The background of the board is a photograph of ocean waves.

**Made By**

Zhuang Ma  
@zhuangma2  
[Edit profile info](#)

**Description**

It's your board's time to shine! Let people know what this board is used for and what they can expect to see.

MEMBERS CAN...

[Comment on cards](#)  
[Change permissions...](#)

## **Aims**

The main purpose of this stage is to set up a system development environment to prepare for subsequent development, which includes the compilation environment, the initial database and the establishment of some external services.

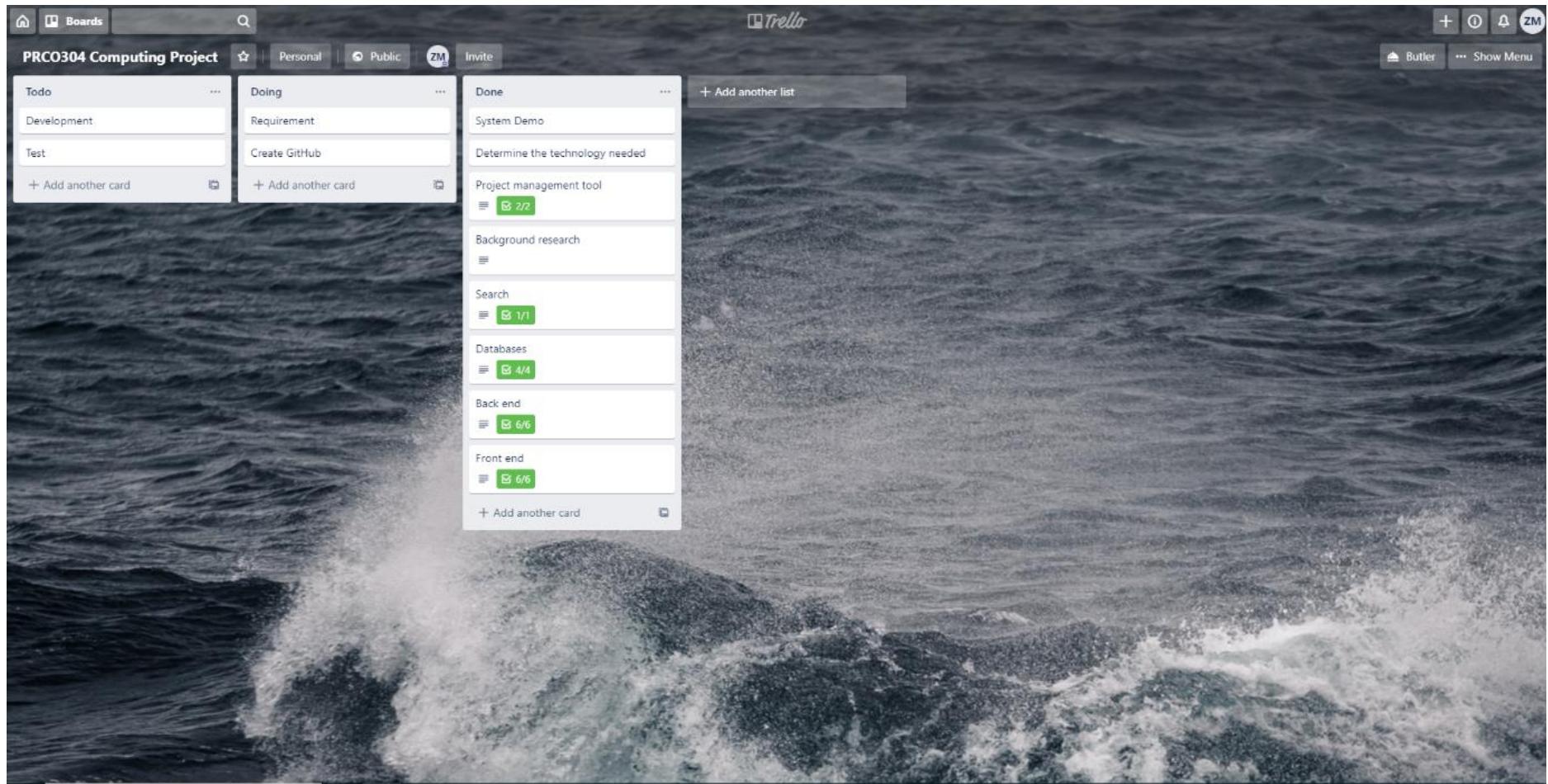
## **Achievements**

- Create a Spring Boot project
- Initialize the MySQL database
- Initialize the Neo4J database
- Connect to multiple data sources through MyBatis configuration and support CRUD operations
- Create a code repository on GitHub

## **Outcome/Discussion**

This sprint stage was very successful, and all the expected functions were completed and tested before the deadline.

## 2.1.2 Trello boards 2



## **Aims**

The main purpose of this phase is to complete background research to determine project needs. At the same time, the overall framework of the project is built, including the front-end style and back-end design (UML).

## **Achievements**

### **Outcome/Discussion**

- Simple demo of this project
- Determine the technology used
- Identify project management tools
- Complete background research

This sprint stage was very successful, and all the expected functions were completed and tested before the deadline.

### 2.1.3 Trello boards 3

The image shows a Trello board titled "PRCO304 Computing Project". The board has three main lists: "Todo", "Doing", and "Done".

- Todo:**
  - Development
  - Test
- Doing:**
  - Research development approach
  - Create Gantt chart
  - Ensure Milestone
  - Ensure Main Task
  - Create a Resource table
  - Ensure Project management approach
  - Ensure Meeting content
  - Risk management
  - Verification & validation
- Done:**
  - System Demo
  - Create GitHub
  - Requirement
  - Determine the technology needed
  - Project management tool
    - 2/2
  - Background research
    - 1/1
  - Search
    - 1/1
  - Databases
    - 4/4
  - Back end
    - 6/6
  - Front end
    - 6/6

At the top right of the board, there is a "Trello" logo and a button labeled "+ Add another list".

## **Aims**

The main purpose of this phase is to stick to the project management plan, which includes Gantt charts, milestones, resource tables and risk management, etc.

## **Achievements**

- Gantt chart
- Milestone
- Resource table
- Risk management plan

## **Outcome/Discussion**

This sprint stage was very successful, and all the expected functions were completed and tested before the deadline.

## 2.1.4 Trello boards 4

The image shows a Trello board titled "PRCO304 Computing Project". The board has three main lists: "Todo", "Doing", and "Done".

- Todo:**
  - Test
- Doing:**
  - Development
  - System overall frame design
  - Third-party account login
  - User Management
  - authority management
  - Role management
  - Blog and Issue Management
  - Comment and answer management
  - Like management
  - Blog and Issue Classification management
  - Tag management
  - Search implementation
  - Notification management
- Done:**
  - Ensure Main Task
  - Verification & validation
  - Risk management
  - Ensure Meeting content
  - Ensure Project management approach
  - Create a Resource table
  - System Demo
  - Ensure Milestone
  - Create Gantt chart
  - Research development approach
  - Create GitHub
  - Requirement
  - Determine the technology needed
  - Project management tool
    - 2/2
  - Background research
    - 1/1
  - Search
    - 1/1
  - Databases
    - 1/1

At the top right of the board area, there is a button labeled "+ Add another list".

## **Aims**

The main purpose of this stage is to develop the main functions of the project. This stage is only a simple development of system functions, and it needs to continue to be improved in the future.

## **Achievements**

- User management
- Authority management
- Question management
- Tag management
- Notification management
- Search function

## **Outcome/Discussion**

This sprint stage was generally successful, but the author encountered some problems since the author is not familiar with the way Oauth2.0 is handled. Although it extended some development time, it did not affect the entire sprint plan.

## 2.1.5 Trello boards 5

The Trello board is titled "PRCO304 Computing Project". It features three main lists: "Todo", "Doing", and "Done".

- Todo List:**
  - Test
- Doing List:**
  - Development (Due 12 Mar, 1/4 completed)
  - System overall frame design (Due 27 Feb)
  - Third-party account login (Due 28 Feb)
  - User Management (Due 29 Feb)
  - authority management (Due 1 Mar)
  - Role management (Due 2 Mar)
  - Blog and Issue Management (Due 3 Mar)
  - Comment and answer management (Due 4 Mar)
  - Like management (Due 5 Mar)
  - Blog and Issue Classification management
- Done List:**
  - System Demo
  - Ensure Milestone
  - Create Gantt chart
  - Research development approach
  - Create GitHub
  - Requirement
  - Determine the technology needed
  - Project management tool
    - Background research
  - Search (1/1 completed)
  - Databases (4/4 completed)
  - Back end (6/6 completed)
  - Front end (6/6 completed)

A legend on the left side defines colors for task status:

- Purple: Researching what will need to be done to complete this task
- Green: Completed
- Blue: Task is ready to be implemented
- Orange: Needs Testing
- Red: Unable to be completed
- Pink: BUG and need to be perfected

## **Aims**

The main purpose of this stage is to improve the already developed functions. Since the previous stage was only developed to implement functions, it is necessary to upgrade the functions at this stage to meet the system requirements.

## **Achievements**

- Improve user management
- Improve authority management
- Improve question management
- Improve tag management
- Improve notification management
- Improve search function

## **Outcome/Discussion**

Although this sprint stage was successful, there are some problems with the data exchange between the front and back ends. For example, when a user logs in with a GitHub account, the user name and mailbox cannot be obtained.

## 2.1.6 Trello boards 6

The screenshot shows a Trello board titled "PRCO304 Computing Project". The board has several lists:

- Project ReadMe/Legend**: A legend for task status:
  - Test
  - Researching what will need to be done to complete this task
  - Completed
  - Task is ready to be implemented
  - Needs Testing
  - Unable to be completed
  - BUG and need to be perfected
- Todo**: A card titled "Test".
- Doing**: A list with three cards:
  - Development (due 12 Mar, 3/4 completed)
  - Improve business logic (2 comments)
  - Improve change requirements
- Done**: A list with ten cards:
  - Ensure Main Task
  - Verification & validation
  - Risk management
  - Ensure Meeting content
  - Ensure Project management approach
  - Create a Resource table
  - System Demo
  - Ensure Milestone
  - Create Gantt chart
  - Notification management (due 9 Mar)
- Implemented features**: A list with four cards:
  - Sensitive word filtering
  - Authority management (3 comments)
  - Registered
  - Sign In
- + Add another list**: A button to add a new list.

## **Aims**

The main purpose of this stage is to improve and upgrade the problems encountered by the existing functions of the system. Including the verification method of user identity, front and back end data transmission and other aspects.

## **Achievements**

- Solve the problem of front-end and back-end data transmission
- Solve the problem of multiple users verifying their identity at the same time in a multi-threaded environment
- Solve the problem of data synchronization between multiple databases

## **Outcome/Discussion**

This sprint stage was very successful, and all the expected functions were completed and tested before the deadline.

## 2.1.7 Trello boards 7

The Trello board is titled "PRCO304 Computing Project". It features three main lists: "Todo", "Doing", and "Done".

- Todo:**
  - Simple unit testing of completed functions
  - Improve issue function
  - Web crawler crawls data
  - search engine
- Doing:**
  - Development (20 Apr, 3/4)
  - Improve business logic (2)
  - Improve change requirements
  - Develop recommended features
  - Deployment test
- Done:**
  - Create Gantt chart
  - Notification management (9 Mar)
  - Search implementation (8 Mar)
  - Tag management (7 Mar)
  - Blog and Issue Classification management (6 Mar)
  - Like management (5 Mar)
  - Comment and answer management (4 Mar)
  - Role management (2 Mar)
  - Blog and Issue Management (3 Mar)
  - authority management (1 Mar)

A legend on the left side defines colors for task status:

- Purple: Researching what will need to be done to complete this task
- Green: Completed
- Blue: Task is ready to be implemented
- Orange: Needs Testing
- Red: Unable to be completed
- Pink: BUG and need to be perfected

## **Aims**

The main purpose of this stage is to develop new user requirements, including the recommendation function and crawler function.

## **Achievements**

- Recommendation function
- Crawler

## **Outcome/Discussion**

This sprint stage was very successful, and all the expected functions were completed and tested before the deadline.

## 2.1.8 Trello boards 8

The screenshot shows a Trello board titled "PRCO304 Computing Project". The board has four main lists:

- Todo**: Contains one card: "Complete report".
- Doing**: Contains three cards: "Unit testing of completed functions" (due 24 Apr), "Local testing" (due 24 Apr), and "Deployment test" (due 24 Apr).
- Done**: Contains several cards:
  - "Ensure Main Task"
  - "Improve change requirements"
  - "Web crawler crawls data"
  - "search engine" (status 5/5)
  - "Improve issue function" (status 4/4)
  - "Development" (due 20 Apr, status 4/4)
  - "Develop recommended features"
  - "Improve business logic" (status 2/2)
  - "Verification & validation"
  - "Risk management"
  - "Ensure Meeting content"
  - "Ensure Project management approach"
- Implemented features**: Contains cards for "Sensitive word filtering", "Authority management", "Registered", and "Sign In", each accompanied by a screenshot of a user interface.

A legend on the left side of the board defines colors for task status:

- Green: Completed
- Blue: Task is ready to be implemented
- Orange: Needs Testing
- Red: Unable to be completed
- Pink: BUG and need to be perfected

## **Aims**

The main purpose of this stage is to test all functions in the system. The testing methods include three types, unit testing, functional testing, and deployment environment testing.

## **Achievements**

- Unit testing
- Functional testing

## **Outcome/Discussion**

Although this sprint stage cannot be considered a success. Because the system cannot run in the deployment environment, the deployment environment test cannot be completed. This problem will be left to the next sprint stage.

## 2.1.9 Trello boards 9

The Trello board is titled "PRCO304 Computing Project". It features three main columns: "Todo", "Doing", and "Done".

- Todo:** Contains cards for "Researching what will need to be done to complete this task", "Completed", "Task is ready to be implemented", "Needs Testing", "Unable to be completed", and "BUG and need to be perfected".
- Doing:** Contains a card for "Complete report" (due 12 May, 9/19 completed).
- Done:** Contains cards for "Q&A Comprehensive Community", "Complete poster", "Record video", "Deployment test" (due 24 Apr, 4/4 completed), "Local testing" (due 24 Apr, 20/20 completed), "Unit testing of completed functions" (due 24 Apr, 18/18 completed), "Ensure Main Task", "Improve change requirements", "Web crawler crawls data", and "search engine".

On the right side of the board, there is a sidebar with a "+ Add another list" button and other Trello navigation links.

## **Aims**

The main purpose of this stage is to deal with the problems left over from the previous sprint stage, and to complete the deployment environment test. Finally, after all the system functions pass the test, report, video and poster production.

## **Achievements**

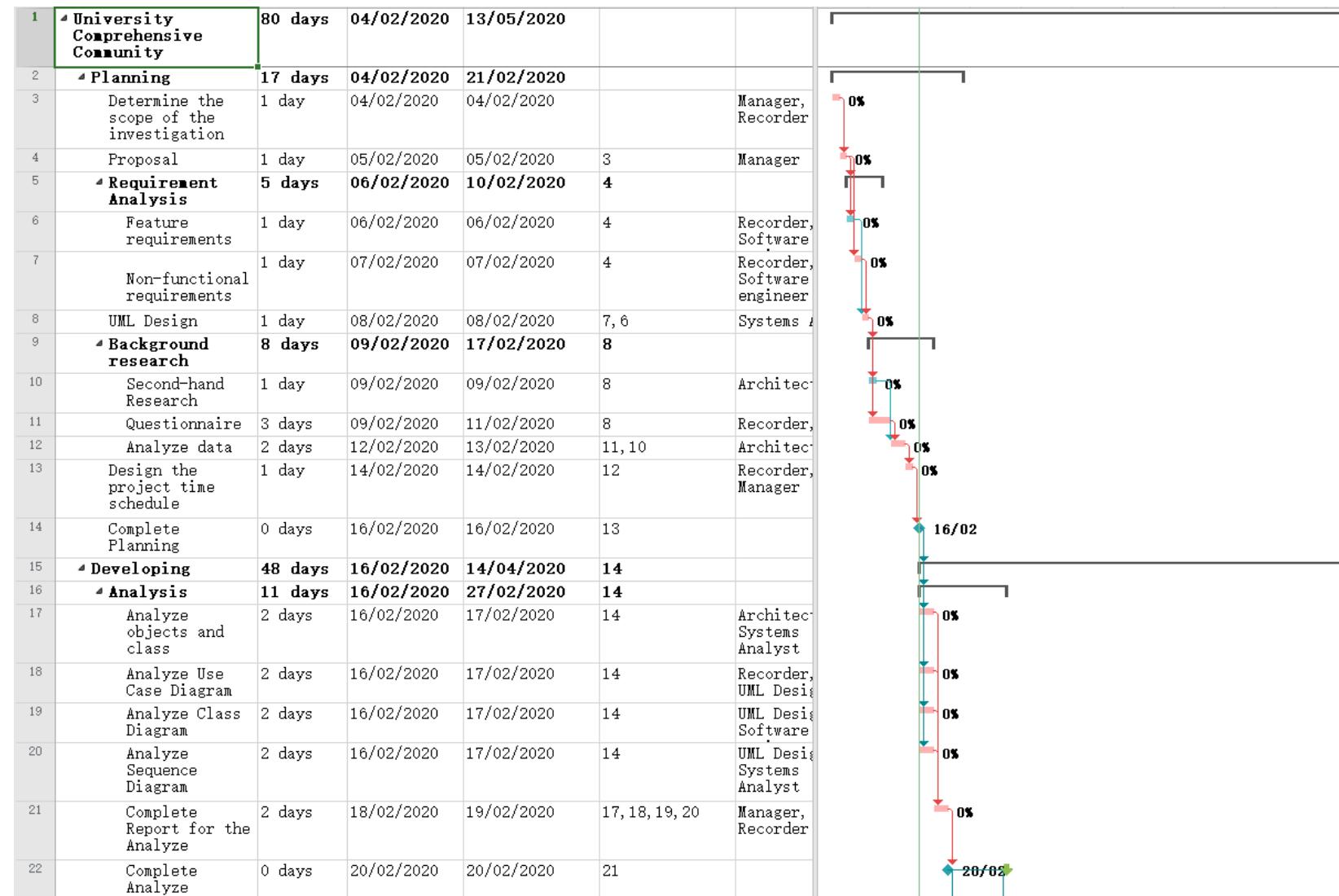
- Deployment environment testing
- Record project video
- Make a project poster

## **Outcome/Discussion**

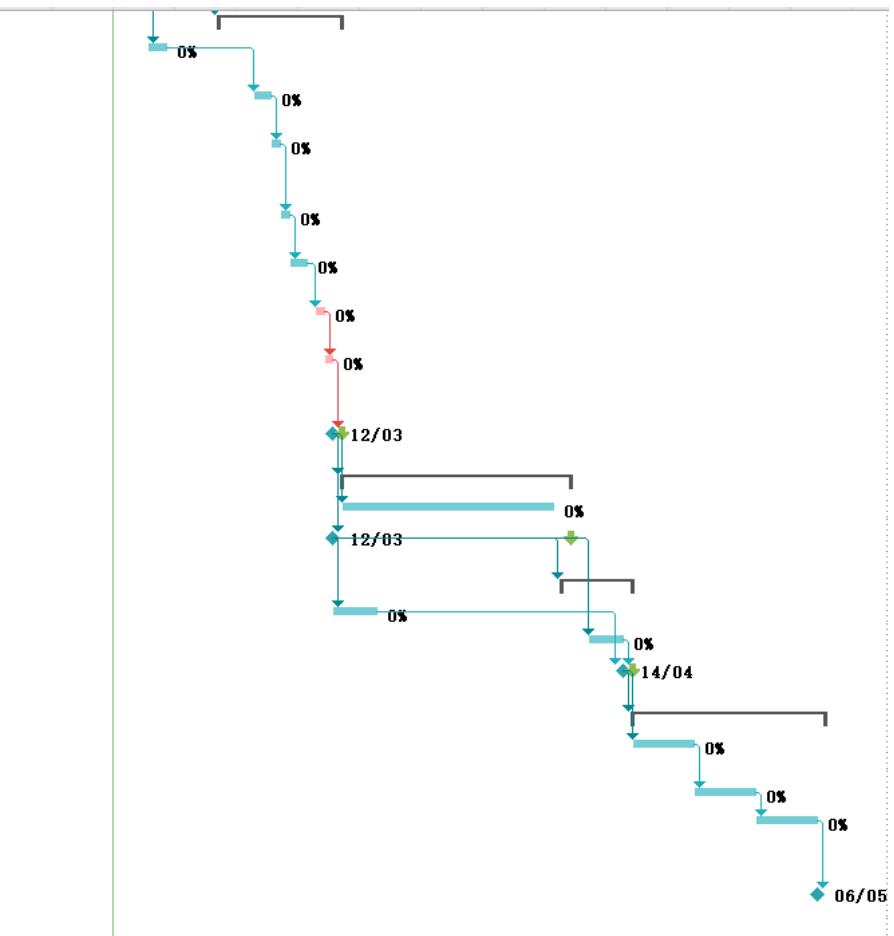
This sprint stage was very successful, and all the expected functions were completed and tested before the deadline.

## 2.2 Plans

### 2.2.1 Gantt chart



23	▪ <b>Design</b>	<b>12 days</b>	<b>28/02/2020</b>	<b>12/03/2020</b>	22	
24	Design Use Case Diagram	2 days	20/02/2020	21/02/2020	22	UML Design Architect
25	Design Class Diagram	2 days	03/03/2020	04/03/2020	24	UML Design Architect
26	Design Sequence Diagram	1 day	05/03/2020	05/03/2020	25	Database administrator
27	Design Function	1 day	06/03/2020	06/03/2020	26	Hardware architect
28	Design User Interface	2 days	07/03/2020	08/03/2020	27	Manager, Media page
29	Design Testing report	1 day	10/03/2020	10/03/2020	28	Architect Database
30	Complete Report for Design	1 day	11/03/2020	11/03/2020	29	Manager, Recorder
31	Complete Design	0 days	12/03/2020	12/03/2020	30	
32	▪ <b>Program</b>	<b>20 days</b>	<b>13/03/2020</b>	<b>07/04/2020</b>	<b>31</b>	
33	Codeing	18 days	13/03/2020	05/04/2020	31	Database
34	Complete Coding	0 days	12/03/2020	12/03/2020	31	Manager, Recorder
35	▪ <b>Testing</b>	<b>6 days</b>	<b>07/04/2020</b>	<b>14/04/2020</b>	<b>34</b>	
36	Black box test	3 days	12/03/2020	16/03/2020	34	Architect
37	White box test	2 days	10/04/2020	13/04/2020	34	Manager, Interface
38	Complete Testing	0 days	14/04/2020	14/04/2020	37, 36	
39	▪ <b>Evaluating</b>	<b>16 days</b>	<b>15/04/2020</b>	<b>06/05/2020</b>	<b>38</b>	
40	Project evaluation	5 days	15/04/2020	21/04/2020	38	Manager
41	Self-evaluation	5 days	22/04/2020	28/04/2020	40	Architect
42	Complete Evaluation Report	5 days	29/04/2020	05/05/2020	41	Manager, Recorder
43	Complete evaluation	0 days	06/05/2020	06/05/2020	42	



### **2.2.2 Risk**

According to the project's Gantt chart, its critical path is focused on design and analysis. This shows that the risk of this project lies in changes in user needs.

There are certain risks in determining user requirements and background research in the planning stage. In addition, there are risks in designing system models during the development stage. To reduce risk:

1. Find existing similar systems for functional analysis. Whenever possible, look for features with good user feedback. In short, take the essence and discard the dross.
2. When the user needs are obtained by issuing a questionnaire, some dirty data may be collected, which will affect the accuracy of the needs. Therefore, as many questionnaires as possible should be distributed when issuing the questionnaire to expand the scope of the collected data and reduce the impact of dirty data on the results.
3. When analysing and designing a system, there may be situations where developers do not understand the system requirements. Therefore, it is necessary to find similar functions in existing similar systems as much as possible, to increase development efficiency, reduce development direction errors, and ultimately perfectly realize user requirements.

### **2.2.3 Milestone**

Milestone Name	Deadline	Complete Content
Complete basic system functions	March 16, 2020	System prototype
Add features based on user feedback	April 1, 2020	System prototype
Added notification function	April 20, 2020	System prototype
Notify by mail	April 28, 2020	System prototype
Complete all functions of the system	May 1, 2020	Integrating all prototype
Complete testing and deployment	May 8, 2020	Deployed

## **2.3 Sprint reviews**

PRCO304: Sprint reviews Report
<b>Name: Ma Zhuang</b>
<b>Date: 04/02/2020</b>
<b>Review of work undertaken</b> <b>Propose project idea and intended technology</b> <b>Problems to be solved:</b> <ul style="list-style-type: none"><li>• Special problems in the learning process.</li><li>• Tutors answers the same question repeatedly.</li></ul>

- A lot of time to find a particular problem.
- Low personal learning efficiency.
- Less communication between industries within the college.

**Intended technology:**

- JAVA
- Spring framework (Boot, MVC, Data, Security, etc.)
- JavaScript and D3.js
- CSS and HTML
- Bootstrap and Thymeleaf
- MySQL
- H2 Database
- Neo4J
- MongoDB
- Hibernate
- Elasticsearch
- Maven and Gradle
- GitHub OAuth API

**Plan of work for the next week**

Create a demonstrable prototype

**Meeting summary**

**PRCO304: Sprint reviews Report**

**Name: Ma Zhuang**

**Date: 11/02/2020**

**Review of work undertaken**

1. Background research
2. Questionnaire investigation
3. Requirement analysis
4. Simple prototyping
5. Making class diagrams
6. Make a top-level use case diagram

**Plan of work for the next week**

1. Make Gantt chart
2. Set milestone
3. Statistical resources
4. Risk analysis
5. Design database
6. Making class diagrams
7. Make a top-level use case diagram

**Meeting summary**

<b>PRCO304: Sprint reviews Report</b>
<b>Name: Ma Zhuang</b>
<b>Date: 25/02/2020</b>
<b>Review of work undertaken</b>
<ol style="list-style-type: none"> <li>1. Gantt chart</li> <li>2. Risk</li> <li>3. MVP</li> <li>4. Internal Test and User test</li> <li>5. Mysql store user information, etc.</li> <li>6. MongoDB storage articles and questions, etc.</li> <li>7. neo4j is used to recommend articles and users</li> </ol>
<b>Plan of work for the next week</b>
<ol style="list-style-type: none"> <li>1. Design and implement system framework</li> <li>2. Complete basic system functions</li> <li>3. Rewrite MongoDB as a system service to run alone</li> <li>4. Complete an overall system prototype, which will be used for secondary development in subsequent development</li> </ol>
<b>Meeting summary</b>
Supervisor's suggestion: Try to use one database to store all the data in the system. Multiple databases may cause data redundancy or affect performance

<b>PRCO304: Sprint reviews Report</b>
<b>Name: Ma Zhuang</b>
<b>Date: 10/03/2020</b>
<b>Review of work undertaken</b>
<ol style="list-style-type: none"> <li>1. Email activation function</li> <li>2. Sign up, login and Github account login</li> <li>3. Generate login verification code</li> <li>4. Account setting function</li> <li>5. Sensitive word filtering</li> <li>6. Post articles, likes, comments and more</li> <li>7. follow feature</li> <li>8. Private message function (Redis)</li> <li>9. system notification (Kafka,zookeeper)</li> <li>10. Full Text Search (Elastic Search)</li> <li>11. Authority control (Spring Security)</li> </ol>
<b>Plan of work for the next week</b>
<ol style="list-style-type: none"> <li>1. Complete the ask question function</li> <li>2. Complete recommended functions</li> <li>3. Optimize private messaging functions, such as using the websocket protocol</li> <li>4. Consolidate database to avoid data redundancy</li> </ol>

- |   |
|---|
| <ol style="list-style-type: none"> <li>5. If possible, the Neo4J database will be used in place of other databases for core functionality. In terms of optimization, the redis in-memory database will be used for cache processing.</li> <li>6. In the case of all the above, a crawling function will be added for the article module.</li> </ol> |
|---|

<b>Meeting summary</b>
------------------------

<b>PRCO304: Sprint reviews Report</b>
<b>Name: Ma Zhuang</b>
<b>Date: 27/03/2020</b>
<b>Review of work undertaken</b>
<ol style="list-style-type: none"> <li>1. Rebuild code</li> <li>2. Added Neo4j database</li> <li>3. Completed online chat (WebSocket)</li> <li>4. Tag management of published content</li> </ol>
<b>Failed:</b>
The deployment on the local tomcat server was successful, but the deployment on the remote server failed.
<b>Plan of work for the next week</b>
<ol style="list-style-type: none"> <li>1. Troubleshooting unsuccessful remote deployments</li> <li>2. Improve post questions Module</li> <li>3. Complete recommended functions</li> <li>4. Crawl top questions and answers from stack overflow</li> <li>5. Full-text search of all resources of the system</li> </ol>
<b>Meeting summary</b>

<b>PRCO304: Sprint reviews Report</b>
<b>Name: Ma Zhuang</b>
<b>Date: 24/04/2020</b>
<b>Review of work undertaken</b>
<ol style="list-style-type: none"> <li>1. Recommended function</li> <li>2. Crawl top questions and answers from stack overflow</li> <li>3. Full-text search of all resources of the system</li> <li>4. Online deployment</li> </ol>
<b>Plan of work for the next week</b>
<ol style="list-style-type: none"> <li>1. System test</li> <li>2. Record video</li> <li>3. Complete the report</li> </ol>
<b>Meeting summary</b>

## Appendix 3: Background Search

### 3.1 Second-hand Research

Second-hand research can help me identify the most promising markets by excluding unsatisfactory markets and lay the groundwork for further field investigations.

**Second-hand research can help me identify the most promising markets by excluding unsatisfactory markets and lay the groundwork for further field investigations.**

In international marketing, the role of second-hand data research is mainly manifested in the following two aspects:

Second-hand research is an important source of information, for the next step in the development of research direction to lay the foundation.

Second-hand research can provide the necessary background information for field research, so that the goal of field research clearer, so as to save time and research costs, for field research to lay the foundation.

**Here are screenshots and sources of some data collected through second-hand research.**

Sources: <http://answers.google.com/answers/>



**Google Answers is no longer accepting questions.**

We're sorry, but Google Answers has been retired, and is no longer accepting new questions. Search or browse the existing Google Answers index by using the search box above or the category links below.

<a href="#">Arts and Entertainment</a>	<a href="#">Reference, Education and News</a>
<a href="#">Business and Money</a>	<a href="#">Relationships and Society</a>
<a href="#">Computers</a>	<a href="#">Science</a>
<a href="#">Family and Home</a>	<a href="#">Sports and Recreation</a>
<a href="#">Health</a>	<a href="#">Miscellaneous</a>

[Google Home](#) - [Answers FAQ](#) - [Terms of Service](#) - [Privacy Policy](#)

Sources: <https://www.quora.com/>

Quora Home Answer Spaces Notifications 2 Search Quora Add Question

Feed Technological Ideas Interesting Information Interesting Virtual Reality (VR) Future of Technology Future Scenarios The Future Technology Trends Fashion and Style Cooking Music Science Technology

About · Careers · Terms · Privacy · Acceptable Use · Businesses

## Virtual Reality (VR)

Follow · 111.6k Bookmark

Read Answer Most Viewed Writers

Answer · Virtual Reality (VR)  
**What is the future of virtual reality?**

Daniel Dan, knows English  
Answered Sep 9

Greetings to all who care about this topic! The trick is that the future of virtual reality has already begun. People and statistics talk about it. Let's see. According to Statista, now there are over... (more)

61 0 1

Sponsored by GrooveHire  
**How broke people can live the high life.**

GrooveHire Sponsored

Related Topics

- Augmented Reality 88.8k Followers
- Virtual Worlds 54.9k Followers
- Oculus VR 4.1k Followers
- Oculus Rift (product) 33.8k Followers
- Augmented Reality (AR) Companies 511 Followers
- Headsets (audio) 29.7k Followers
- Reality 310.6k Followers
- Virtual Reality Gaming 779 Followers
- VR Headsets 8k Followers
- Virtualization 39k Followers
- HTC Vive 7.9k Followers
- Samsung Gear VR 6.8k Followers
- Google Cardboard 6k Followers
- Technology Trends 9m Followers
- Future of Technology 1.8m Followers

Sources: <https://techcrunch.com/>

The screenshot shows the TechCrunch Videos section. At the top, there's a banner for Ford with text about lease options. Below it, a sidebar on the left lists various categories like Login, Search Q, Startups, Apps, Gadgets, Videos, Audio, Newsletters, Extra Crunch, Advertise, Events, More, Transportation, Apple, Tesla, and Security. The main content area features a large title 'Videos'. Below the title, a navigation bar includes links for All, News, Gadgets, Features, Reviews, Interviews, Apps, Disrupt, Battlefield, Sessions, Crunch Report, and More. A video thumbnail for 'TC Top Picks Berlin 2019: Cyanite' is displayed, showing a dark interface with a waveform and the Cyanite logo.

Sources: <https://stackoverflow.com>

The screenshot shows the Stack Overflow homepage. On the left, there's a sidebar with links for Home, PUBLIC (Stack Overflow, Tags, Users, Jobs), TEAMS (What's this?, Free 30 Day Trial), and a search bar. The main content area is titled 'Top Questions' and displays a list of ten questions. Each question card includes details like votes, answers, views, tags, and the asker's name. To the right of the main content are several sidebar panels: 'Blog' (with a link to 'How Shapeways' software enables 3D printing at scale'), 'Featured on Meta' (with a link to 'The Overflow #10: The 500-mile email'), 'TLS 1.0 and TLS 1.1 removal for Stack Exchange services', 'Did Stack Exchange cut the number of negative comments nearly in half between...', 'An account of my meeting with the Stack Overflow management team', 'Custom Filters' (with a link to 'Create a custom filter'), 'Watched Tags' (with a link to 'edit' and a 'java' tag), 'Ignored Tags' (with a link to 'Add an ignored tag'), and a Microsoft Azure advertisement.

Sources: <https://brainly.com/>

The screenshot shows the homepage of Brainly. At the top left is the "BRAINLY" logo. To the right are "LOG IN", "JOIN NOW", and "ASK QUESTION" buttons. The main title "Go from questioning to understanding" is displayed prominently in large, bold, black font. Below the title is a subtitle: "Brainly is the knowledge-sharing community where 150 million students and experts put their heads together to crack their toughest homework questions." A search bar with the placeholder "What is your question?" and a magnifying glass icon is located below the subtitle. To the right of the search bar is a colorful illustration of two students, one with red hair and one with blue hair, looking through a microscope. Below the search bar are category links: "All", "Mathematics", "History", "English", and "Biology", each with a small icon. Navigation arrows for "All" categories are on either side of the main content area.

### **Ruled by students, supported by parents.**

#### **3.2 Questionnaire**

Field research is an effective research method. Before you start designing your system, you need to interview users and record the data you are investigating. Then you need to experience as much as possible of the full functionality of the existing question answering system. This is to facilitate the development of the system, avoid repetitive errors, and provide better services for schoolteachers and students.

#### **Questionnaire**

The questionnaire survey is a first-hand survey method. The questionnaire survey is a very effective way to obtain user needs. After subsequent data analysis, you can quickly and accurately obtain the needs of a large number of users. At this stage, it is necessary to avoid restrictions on user needs. Each questionnaire requires collection of 35 valid data. For the final analysis phase, I first need to summarize the data of each questionnaire and analyse them one by one to determine the initial functional requirements.

- **Questionnaire question**

University Comprehensive Community User Needs Survey

Hello there. I am a computer student at Plymouth University. The purpose of this questionnaire is to identify user needs in the university comprehensive community. Your answer will provide an important reference for my software development. The questionnaire data and personal information are used for academic research only. Please feel free to reply.

1. The type of major you are studying (Single choice)

- Engineering
- Science
- Business
- Art
- Medicine
- Other \_\_\_\_\_

2. When you encounter a professional-related problem, you first find a solution (Single choice)

- Q & A website
- Search engine
- Ask classmates
- Ask the teacher
- Other \_\_\_\_\_

3. Do you feel that the answers you get online match your question? (Single choice)

- absolutely okay
- Most can
- Some can
- Rarely can

4. Do you want a professional teacher to help solve the problem? (Single choice)

- Yes of course
- Unwilling

5. Do you have a problem during the break or on the weekend that needs to be addressed by the teacher? Can I find the teacher in real time? (Single choice)

- Basically, all right
- Sometimes can

- Most cannot

6. As a questioner, what kind of questions do you prefer to ask through the platform?

- Professional Question Answering (Multiple choice)
- Analysis of knowledge points
- Professional Details
- Learn how to progress
- Other \_\_\_\_\_

7. What causes you to abandon a question and answer system (Multiple choice)

- Charge
- Fewer resources
- The page is not beautiful
- Other \_\_\_\_\_

8. If there is a system for campus Q & A, do you think that as a user, what do you want it to bring you? (Multiple choice)

- Guidance in learning direction
- Confused question answer
- Recommended learning style
- other \_\_\_\_\_

9. For the user experience of the existing Q & A apps and Q & A platforms, what are you most satisfied with / are you least satisfied with?

---

10. What features do you want in a complete Q & A system

---

### 3.3 Analysis of research results

#### *3.3.1 Analysis for Field research*

After in-depth use of the Quora system and understanding of all its functions, the following conclusions were reached.

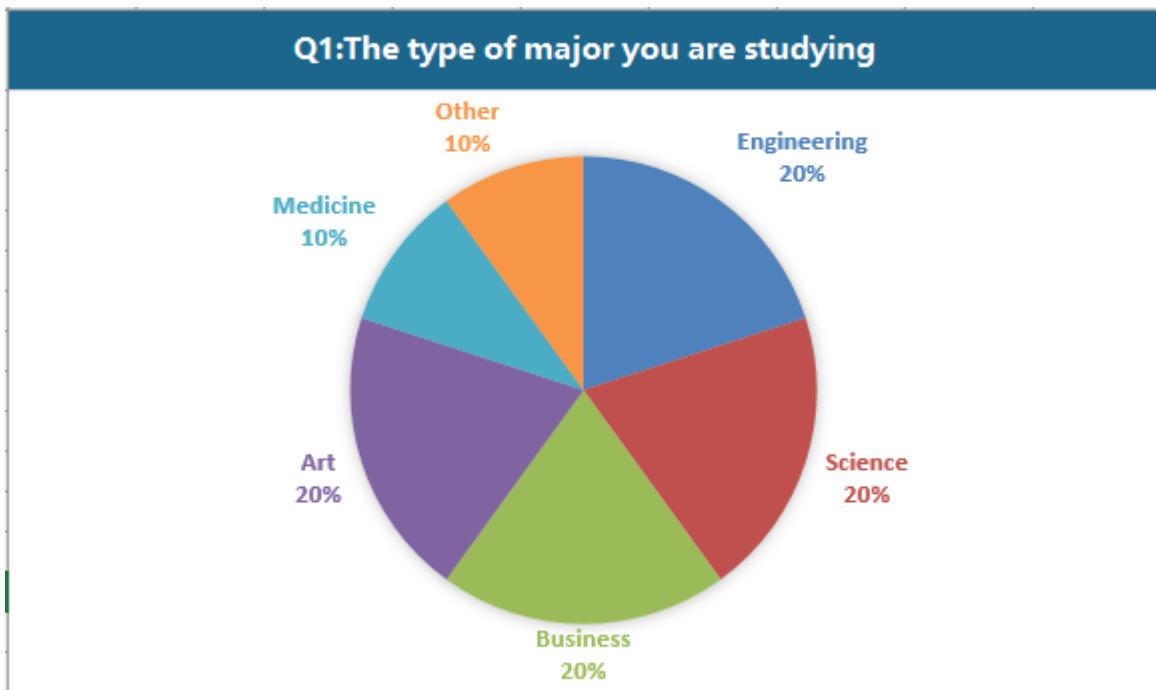
- **Functional aspects**
  - i. The features of the Quora system are relatively comprehensive, especially in terms of asking questions and answering questions. When users ask questions, the system standardizes the way users ask questions and sets tags for questions posted by users. The advantage of this is that users can accurately search for relevant solutions when querying problems.
  - ii. Quora has also established many discussion groups to allow users with the same preferences to communicate in the same discussion area. This approach will undoubtedly increase user usability and loyalty to the system.
  - iii. At the beginning of new user registration, the system supports third-party account logins, reducing the hassle of users creating multiple accounts. Users only need to have one account to log in to other platforms. Secondly, when a new user logs in for the first time, the system will ask the user to choose their preferences and the discussion group they want to join. These practices fully demonstrate the powerful functions of the system for user division and management.
- **User experience**

Quora has a very clear display page, user self-management module and issue preview function. These indicate that Quora's human-computer interaction is well done, and it is worth quoting in project development.
- **Content aspect**
  - i. Because Quora has many groups, the content displayed on the homepage is too cluttered, and it cannot quickly locate the problems that users are interested in or the problems to be solved.
  - ii. There is no specific retrieval method for the content of the questions answered by the user, which results in that the solution searched by the user is not 100% accurate, and the user needs to judge the solution by himself.

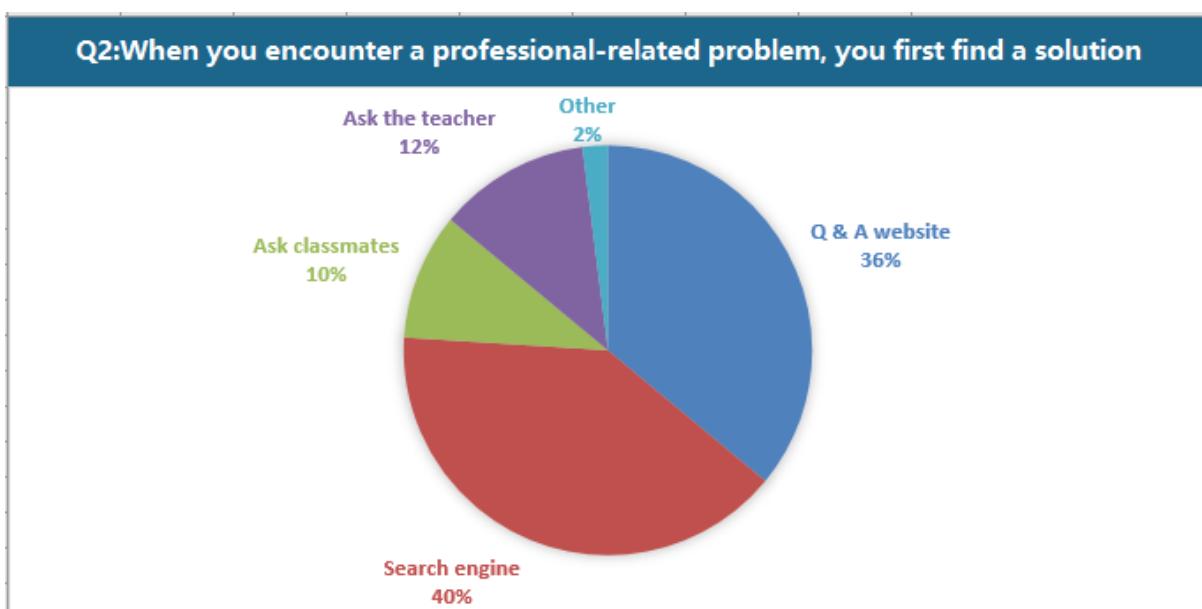
### 3.3.2 Analysis for Questionnaire

A total of 50 questionnaires were distributed and 50 were recovered, of which 47 were valid data.

**Q1:** The data obtained from the survey shows that the users participating in the survey are evenly distributed, so the survey data can be used as a reference



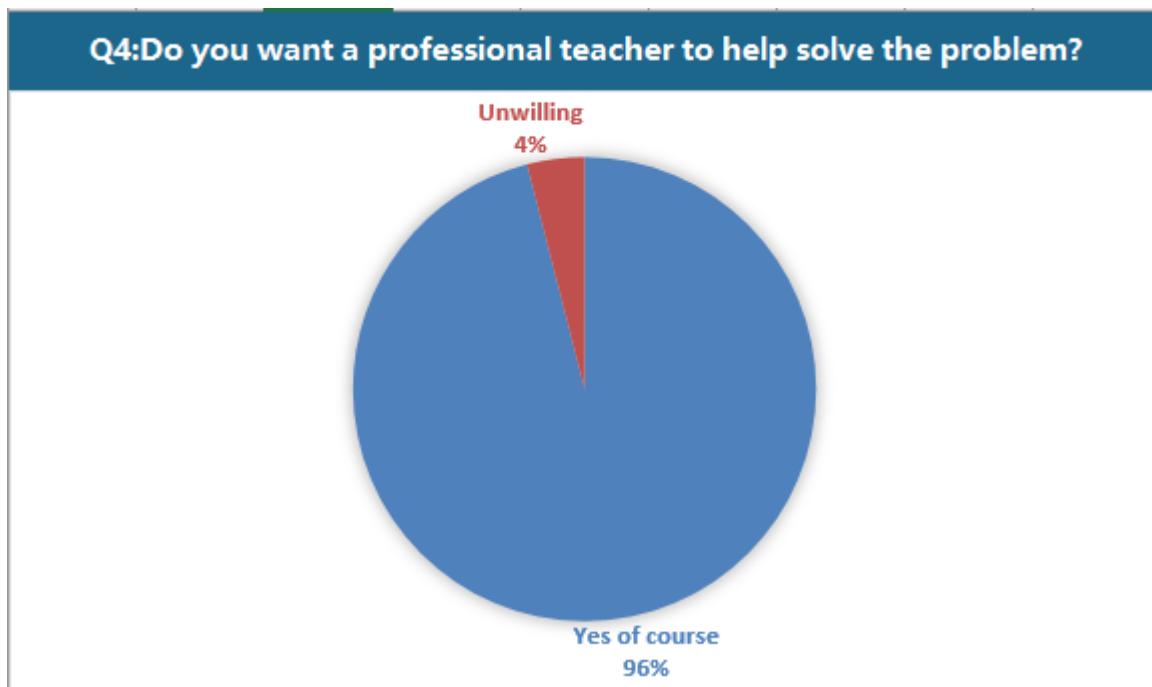
**Q2:** Through data analysis, when users encounter learning difficulties, the first solution they use is a search engine (40% of the total), followed by a Q & A system (36%). Ask classmates or teachers to find solutions that account for 10% and 12%, respectively. Most users still prefer to search for solutions to problems through the Internet.



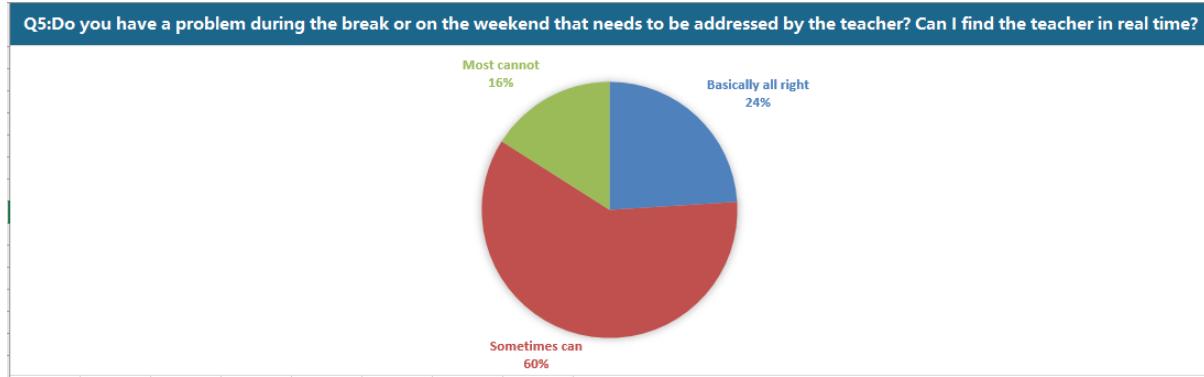
**Q3:** When users search for answers through the Internet, only a few of them can find useful information (forty-eight percent), and only a small number of users can query the Internet for accurate solutions.



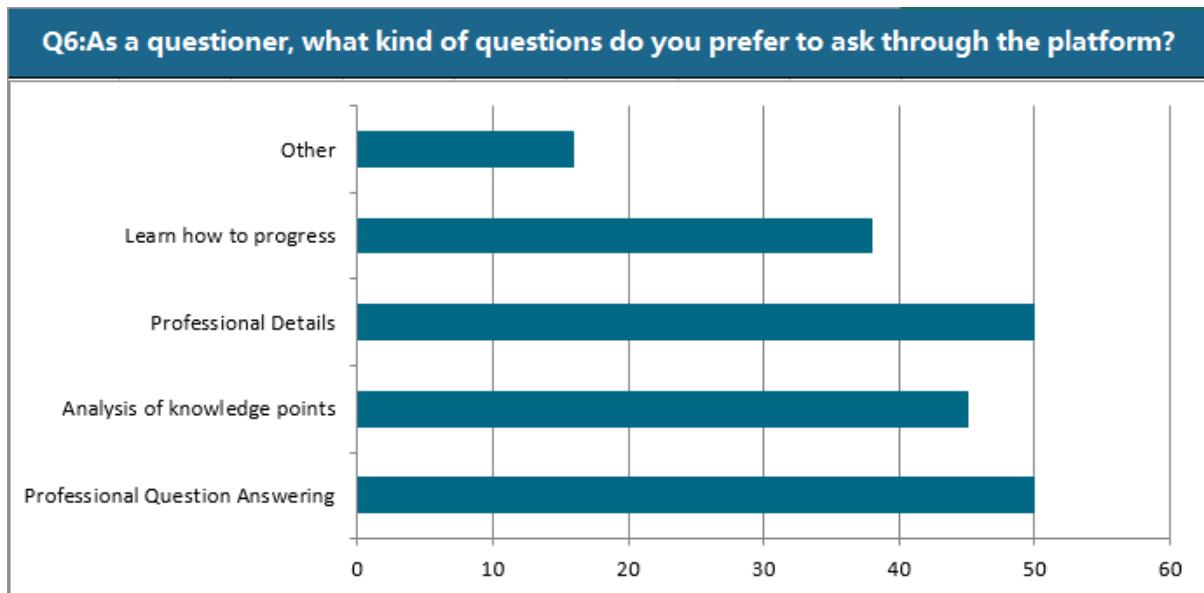
**Q4:** When it comes to whether a professional mentor is required to solve a professional problem, the intentions of the users appear to be particularly consistent. Among them, 96% of the users want professional mentors to coach professional problems, and almost no one is unwilling.



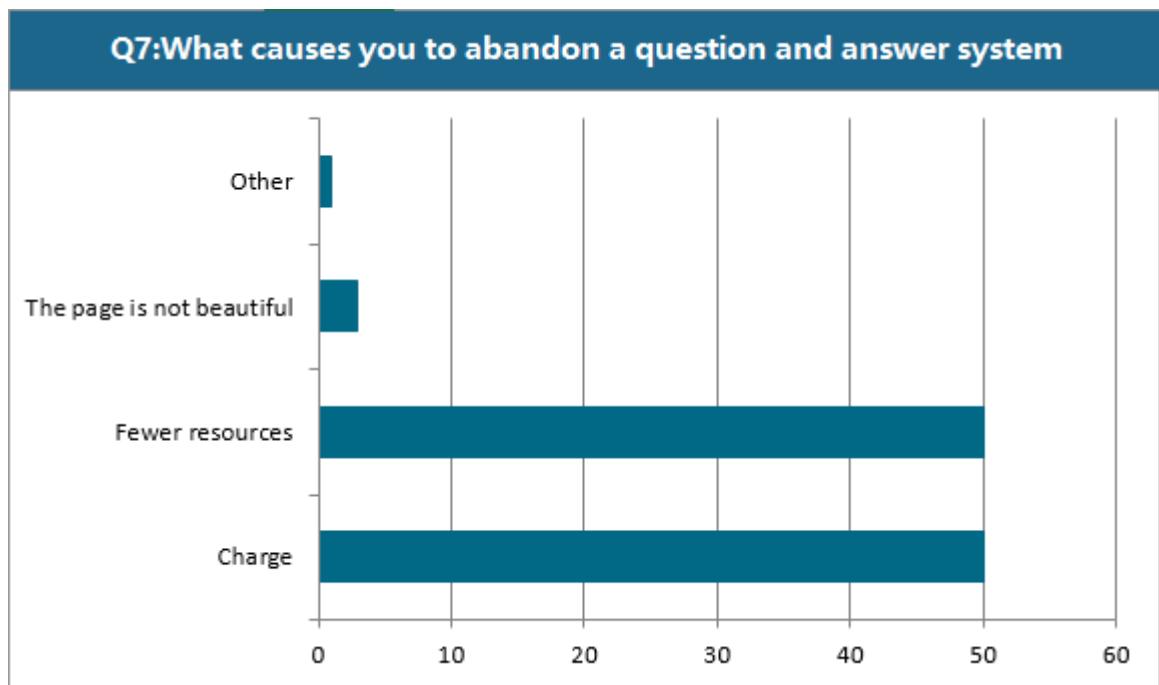
**Q5:** Among the user groups who ask teachers, most users can't get in touch with their mentors quickly and easily when they contact the teachers on weekends or breaks to answer questions. (60%) Even a small percentage of users with 16% cannot contact the mentor, and only 24% of users can contact the teacher at any time.



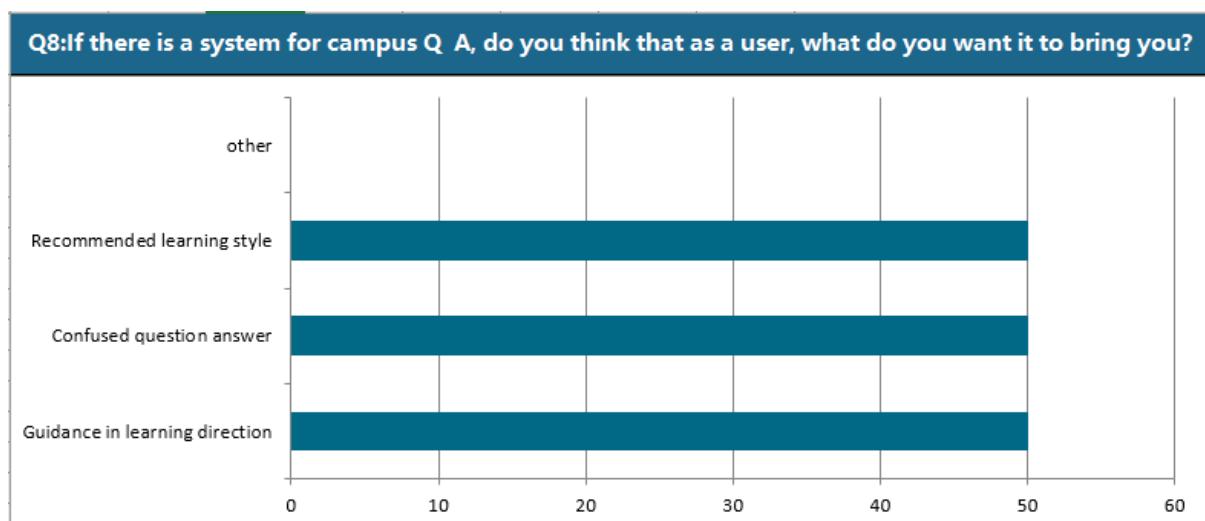
**Q6:** The data collected through the sixth question shows that the questions that the user most wants to ask as a questioner are mainly professional-related questions and solutions corresponding to professional questions. Secondly, users are keen to ask questions about knowledge points, learning routes, and employment.



**Q7:** Data on what caused the user to abandon the use of the question and answer system. Almost all users will abandon the system due to factors such as fees and the quality of the solution to the problem.



**Q8:** Through data analysis, it is known that users most want to obtain useful information on the Q & A system, such as professional answers to professional questions, learning routes, recommended learning methods, and the latest related industry consulting.



**Q9: For the user experience of the existing Q & A apps and Q & A platforms, what are you most satisfied with / are you least satisfied with?**

- **Positive**
  - i. Has many distributions, providing users with more than the relevant knowledge of the profession.
  - ii. Provide third-party account login to reduce the hassle of remembering account and passwords.
  - iii. Categorize questions for easy query
- **Negative**
  - i. Search results have a lot of duplicates
  - ii. The content of the solution is called clutter, and you cannot find an accurate solution quickly.

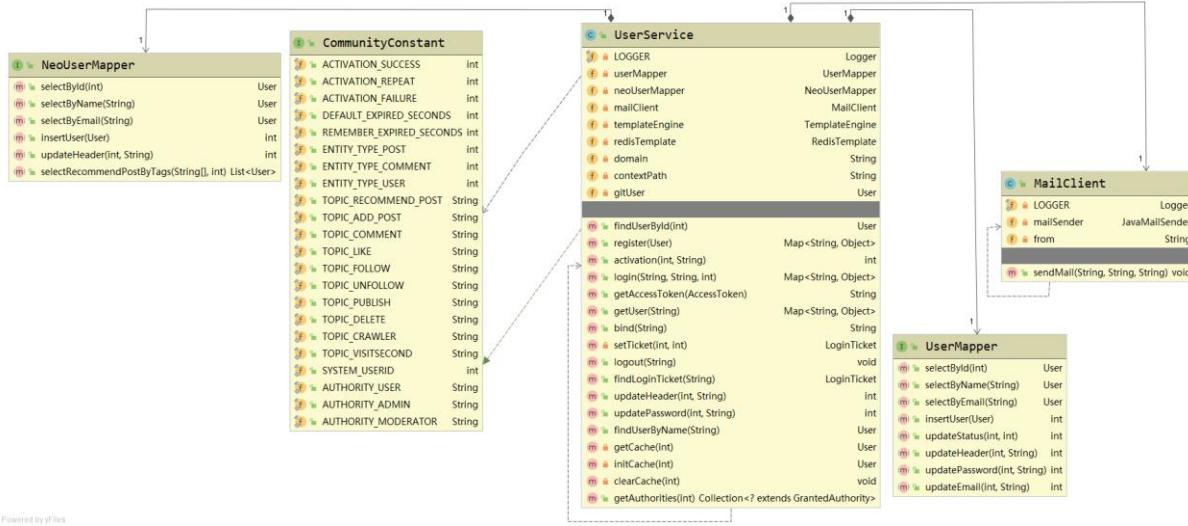
**Q10: What features do you want in a complete Q A system**

1. Blog
2. Online communication
3. Industry Information
4. Tutors participate in Q & A
5. Recommend articles or questions based on user interests

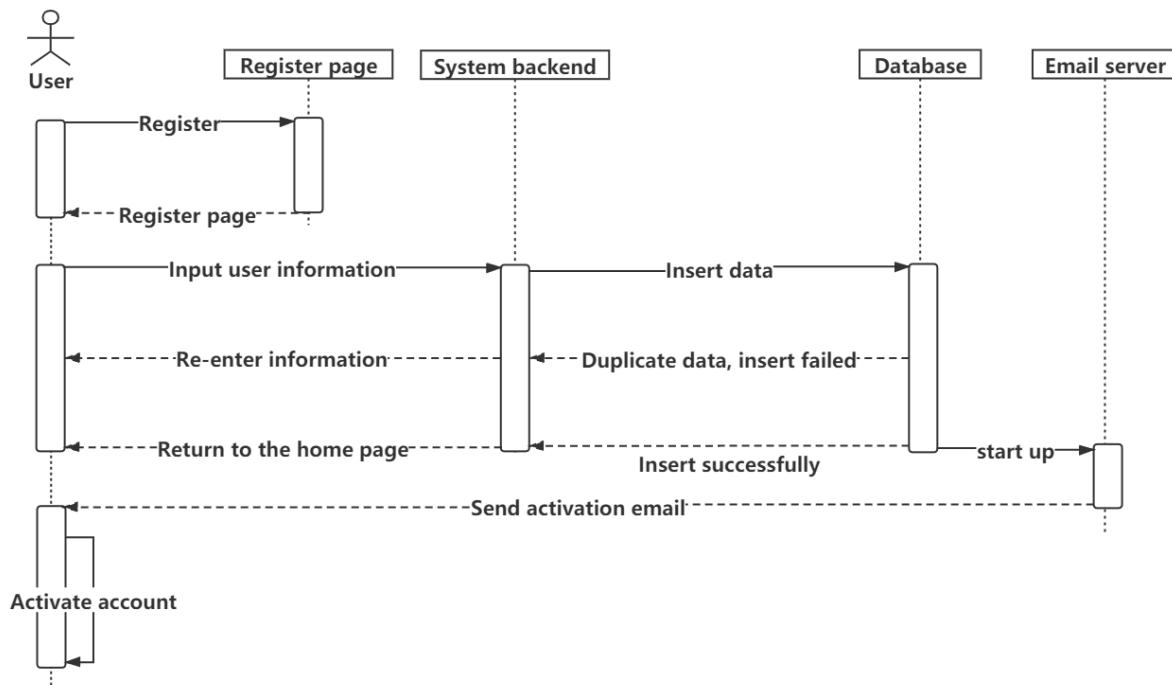
## Appendix 4: Design

### 4.1 Register feature design

#### 4.1.1 Class diagram

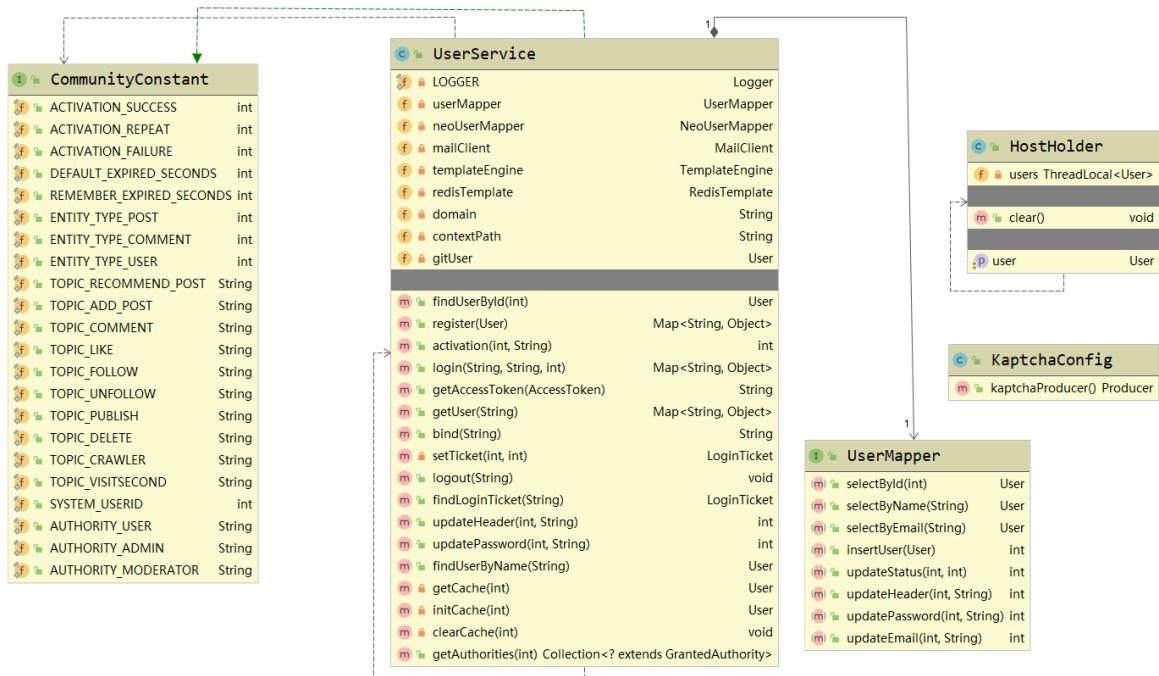


#### 4.1.2 Sequence diagram

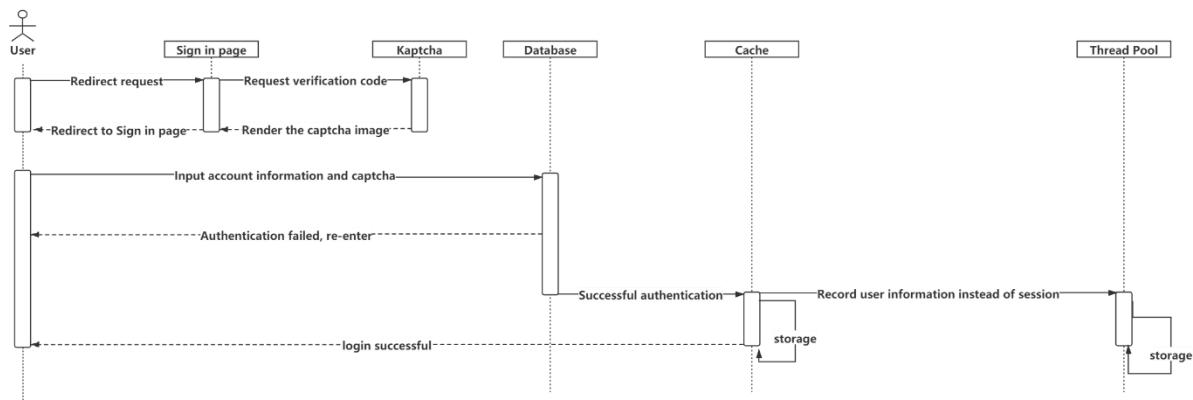


## 4.2 Sign in feature design

### 4.2.1 Class diagram

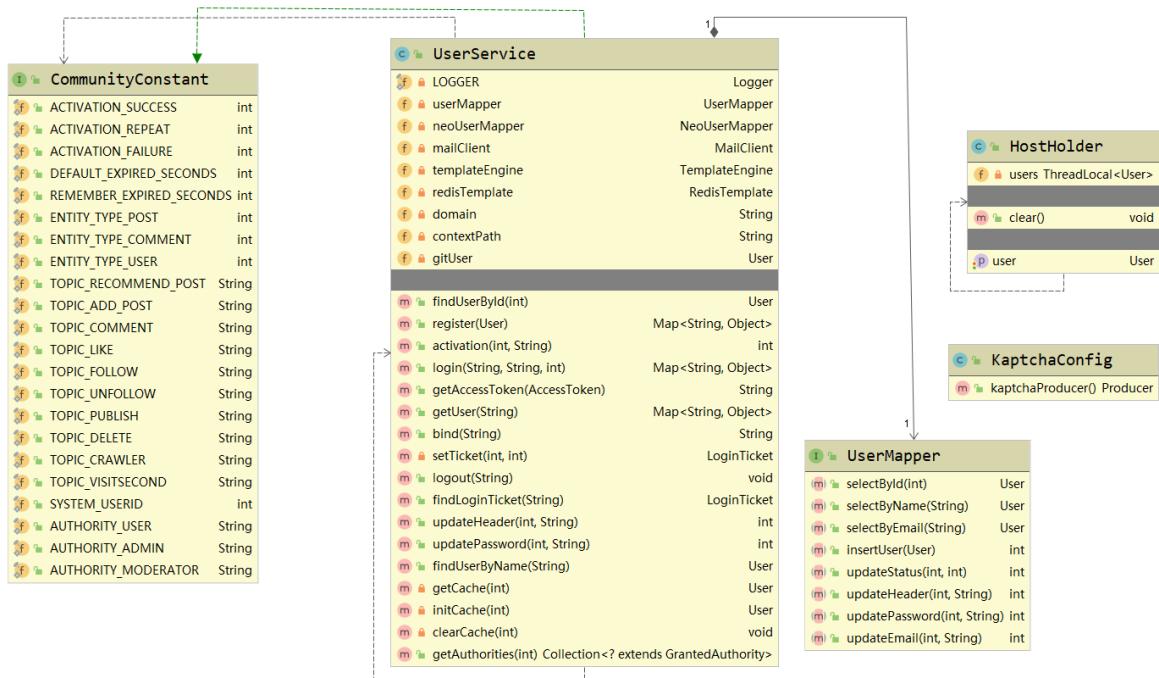


### 4.2.2 Sequence diagram

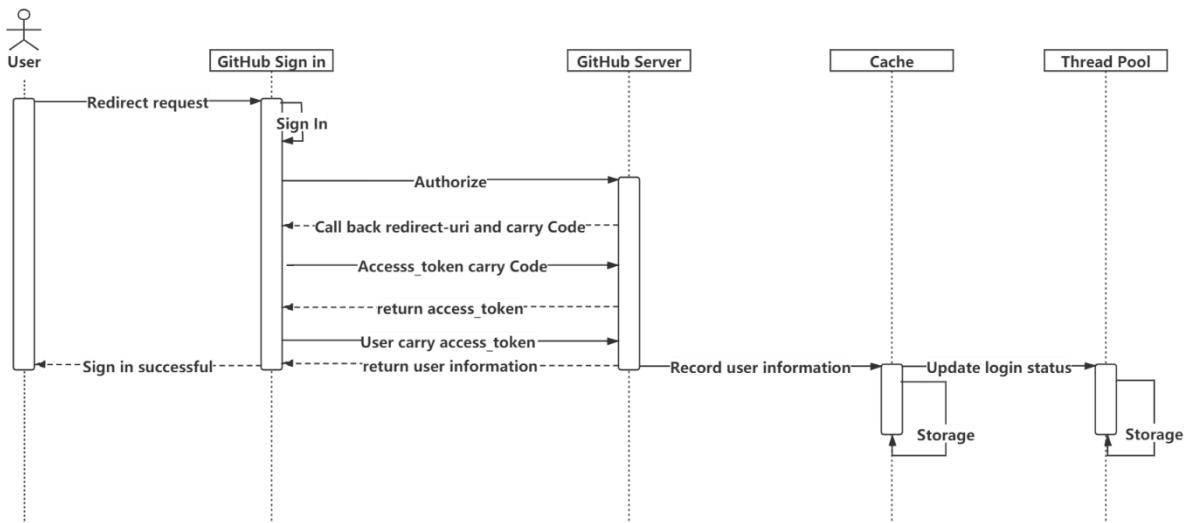


## 4.3 OAuth2.0 feature design

### 4.3.1 Class diagram

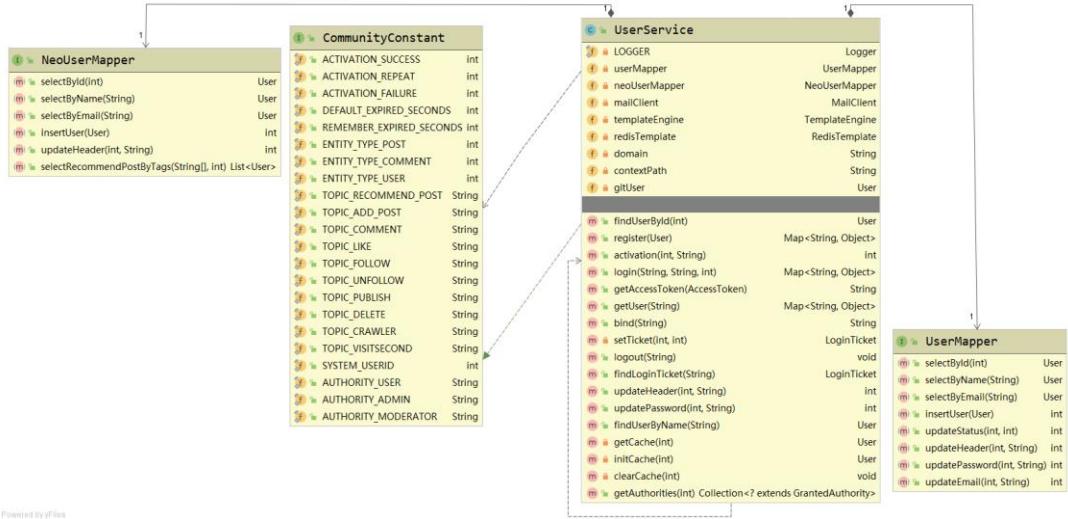


### 4.3.2 Sequence diagram

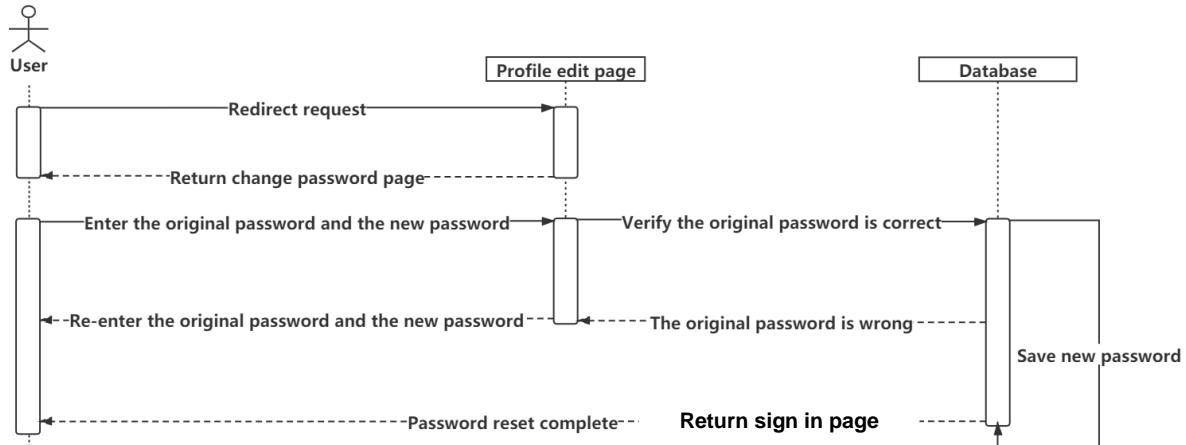


## 4.4 Change password feature design

### 4.4.1 Class diagram

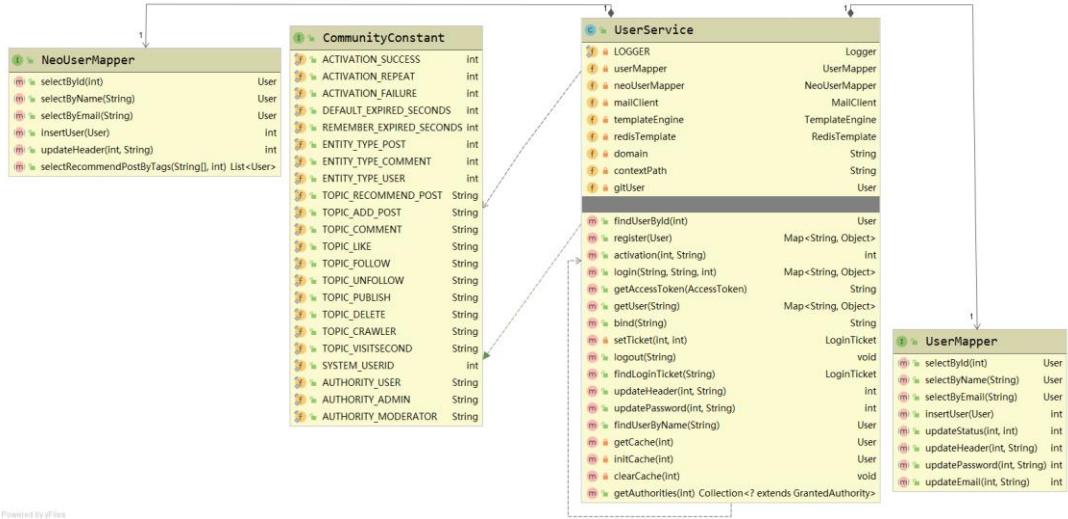


### 4.4.2 Sequence diagram

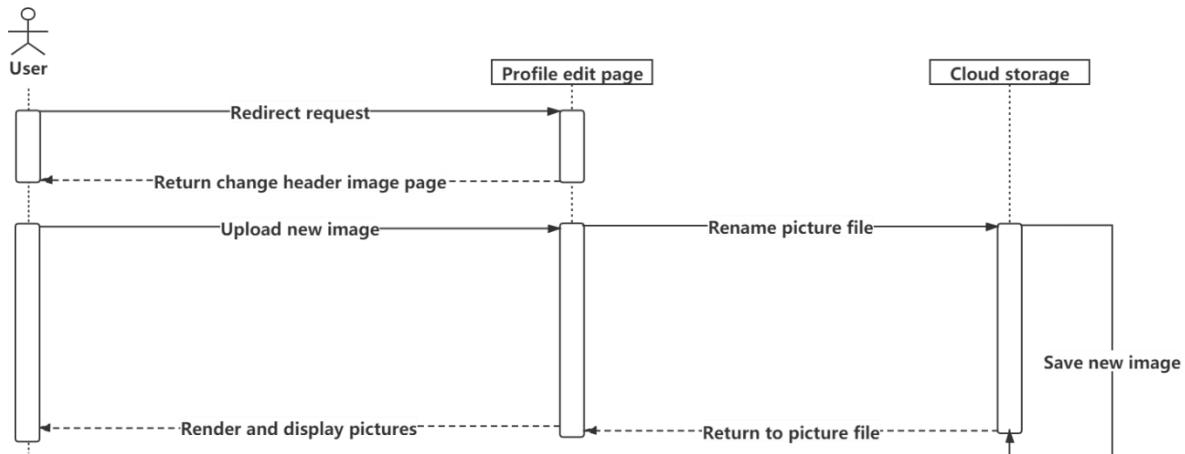


## 4.5 Change avatar feature design

### 4.5.1 Class diagram

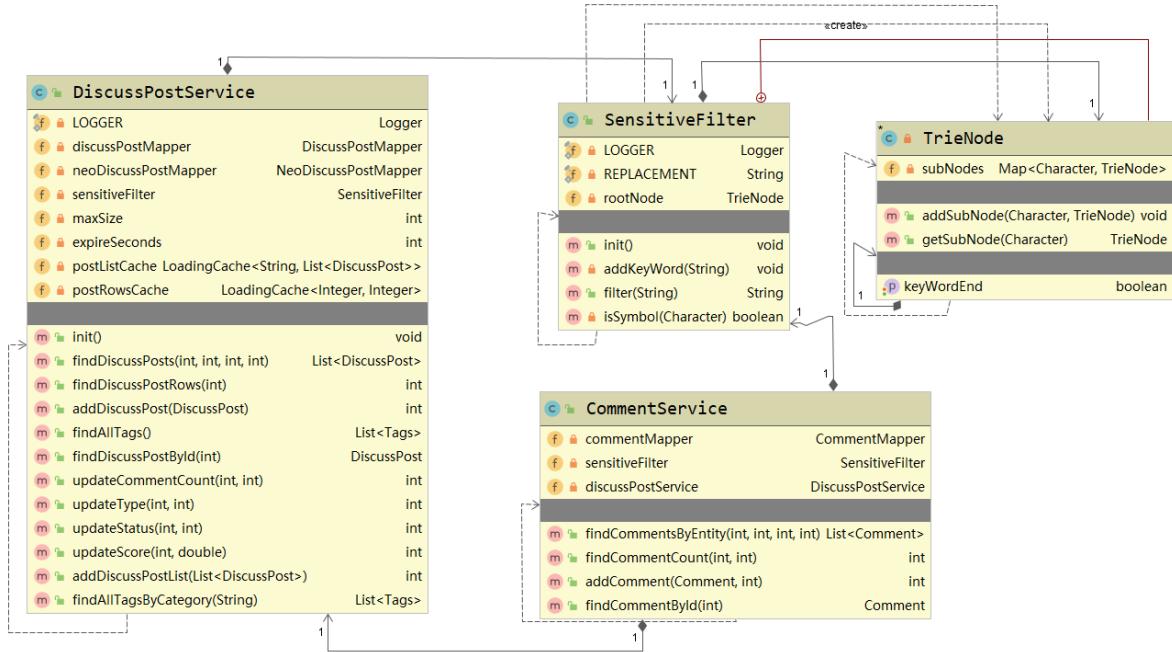


### 4.5.2 Sequence diagram

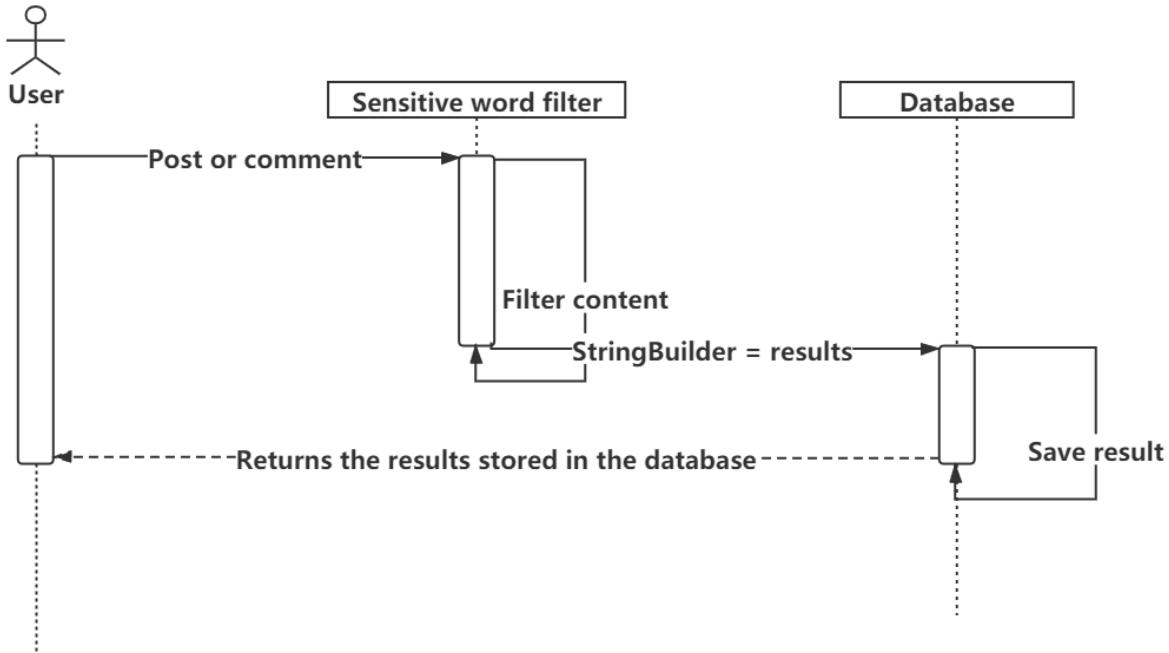


## 4.6 Sensitive word filtering feature design

### 4.6.1 Class diagram

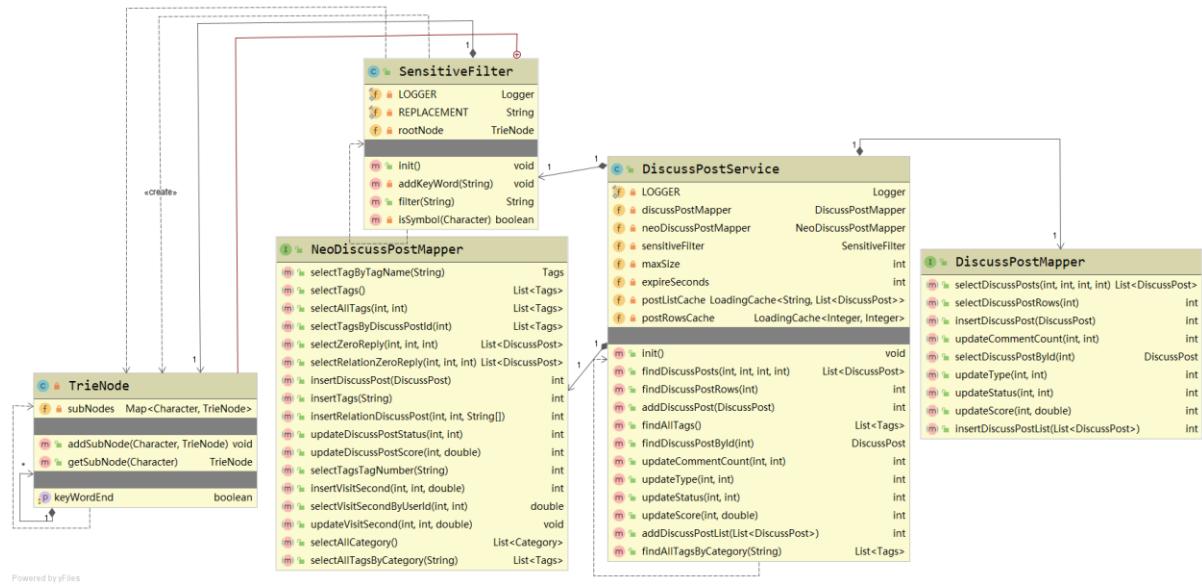


### 4.6.2 Sequence diagram

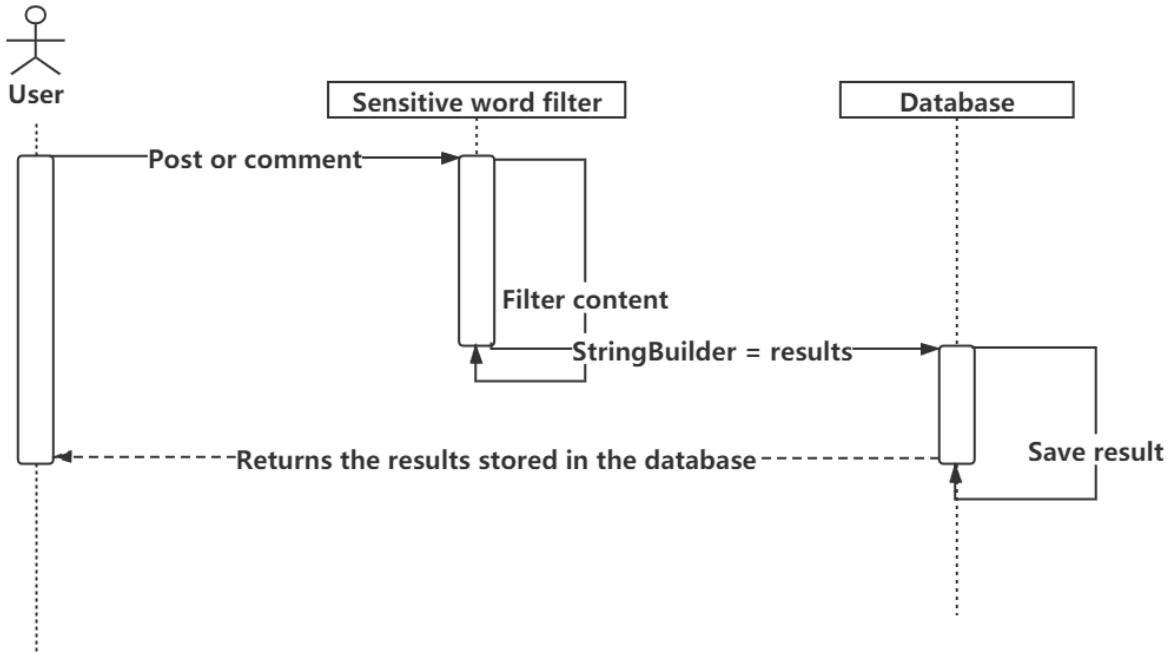


## 4.7 Ask question feature design

### 4.7.1 Class diagram

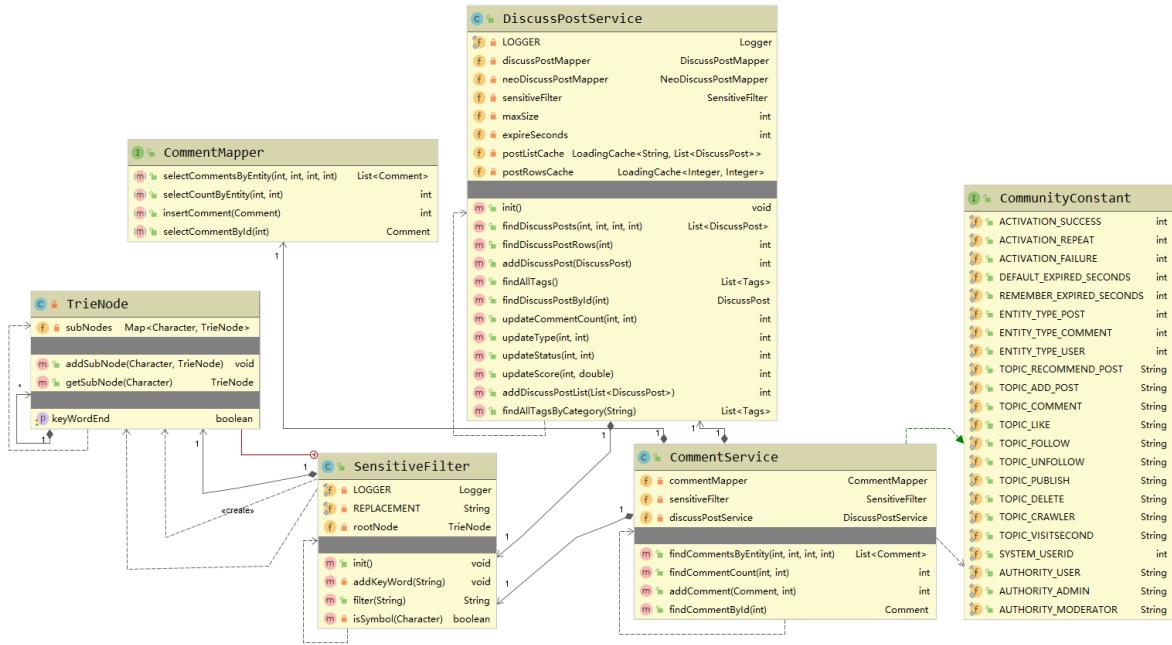


### 4.7.2 Sequence diagram

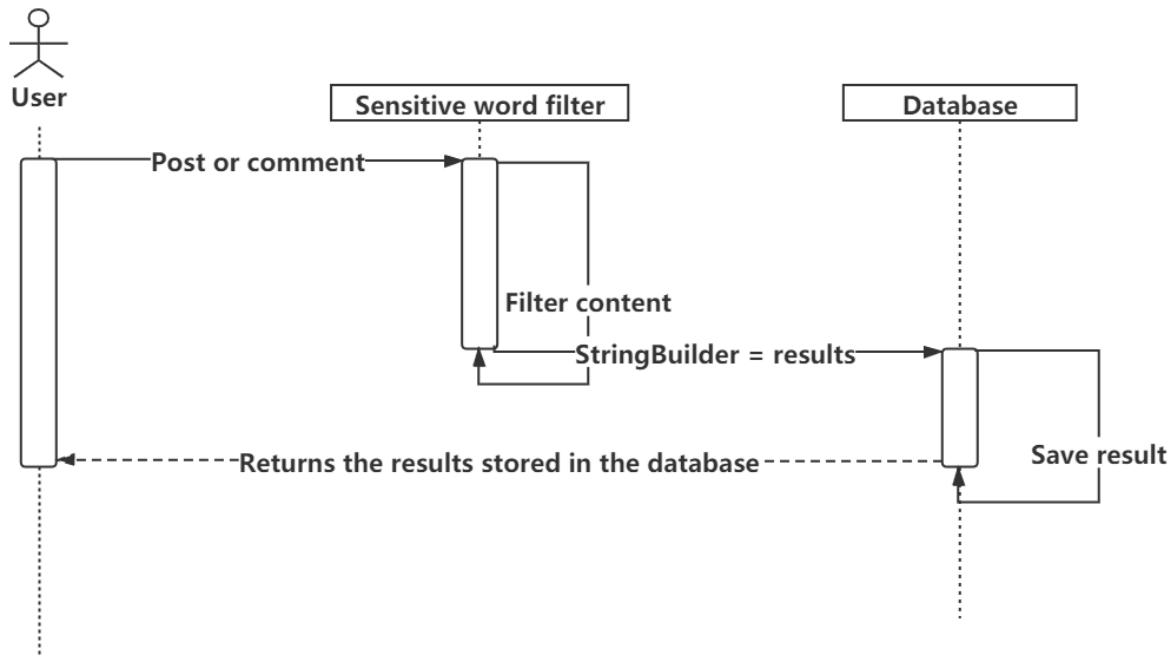


## 4.8 Comment and reply feature design

### 4.8.1 Class diagram

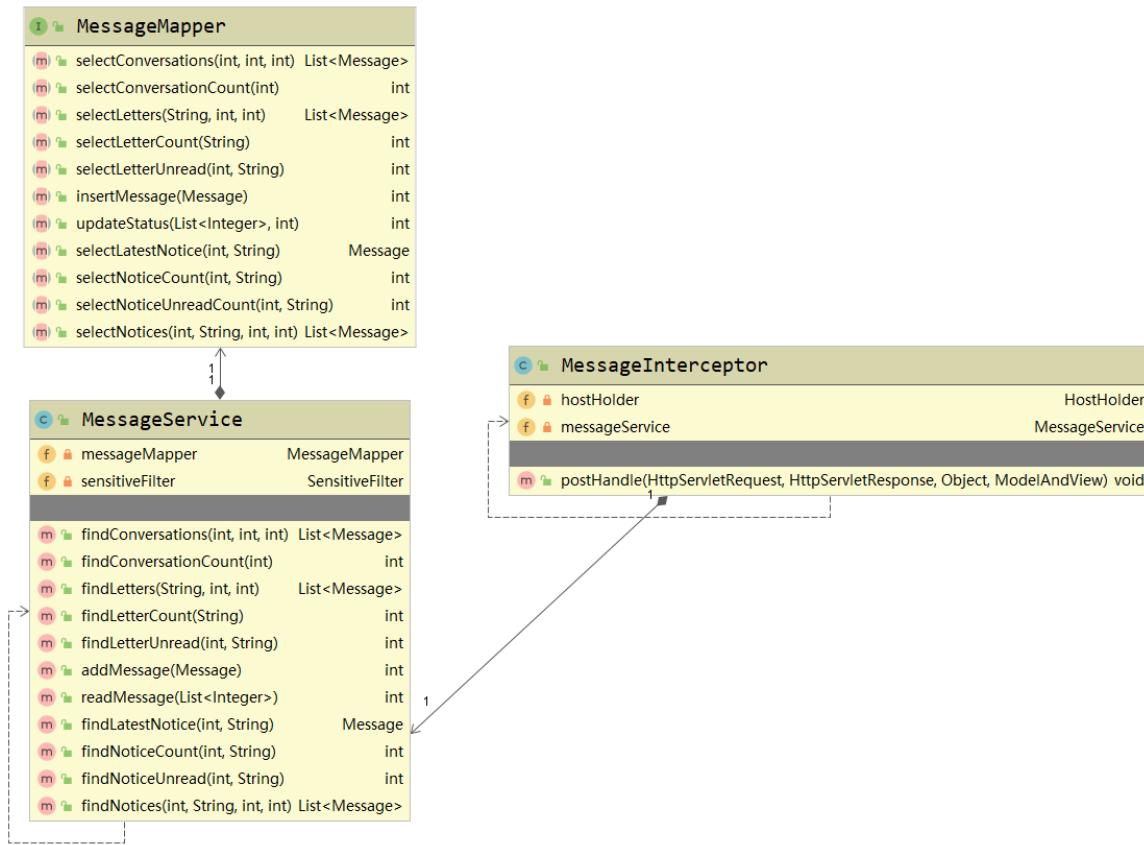


### 4.8.2 Sequence diagram

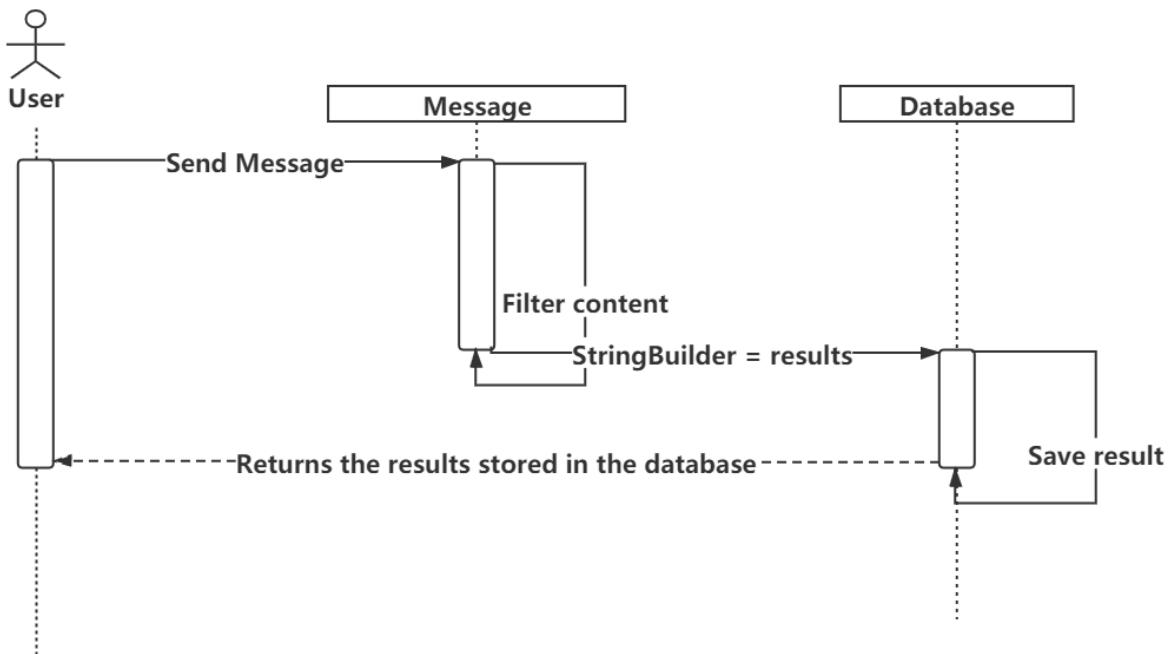


## 4.9 Online chat feature design

### 4.9.1 Class diagram

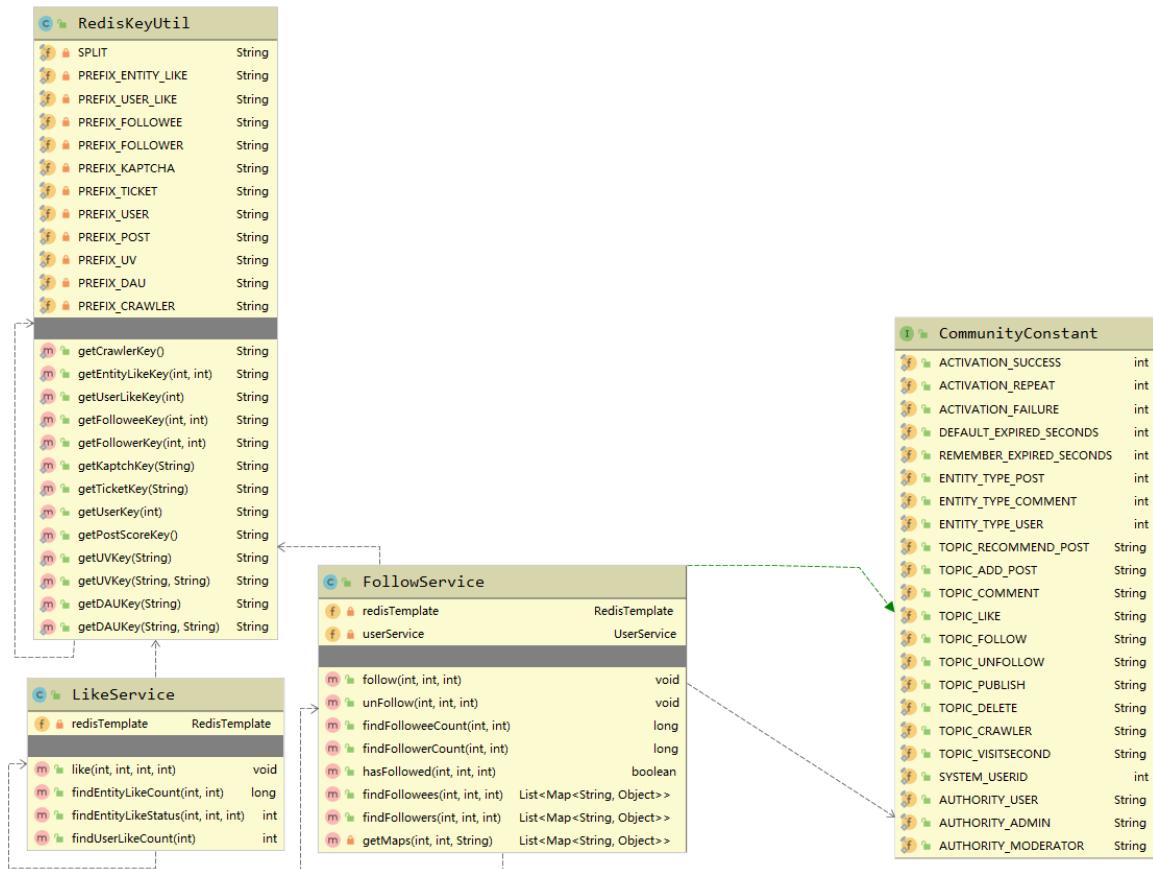


### 4.9.2 Sequence diagram

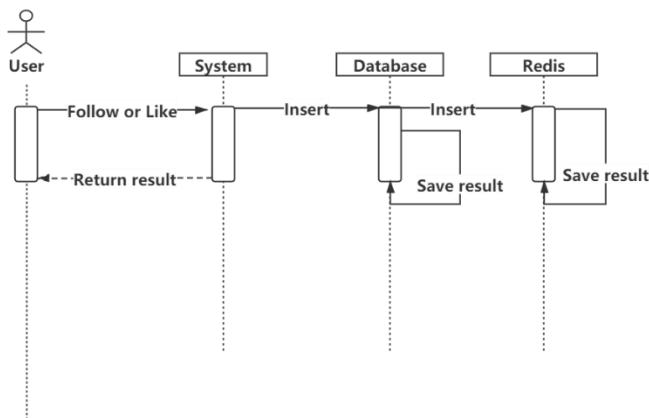


## 4.10 Like and follow features design

### 4.10.1 Class diagram

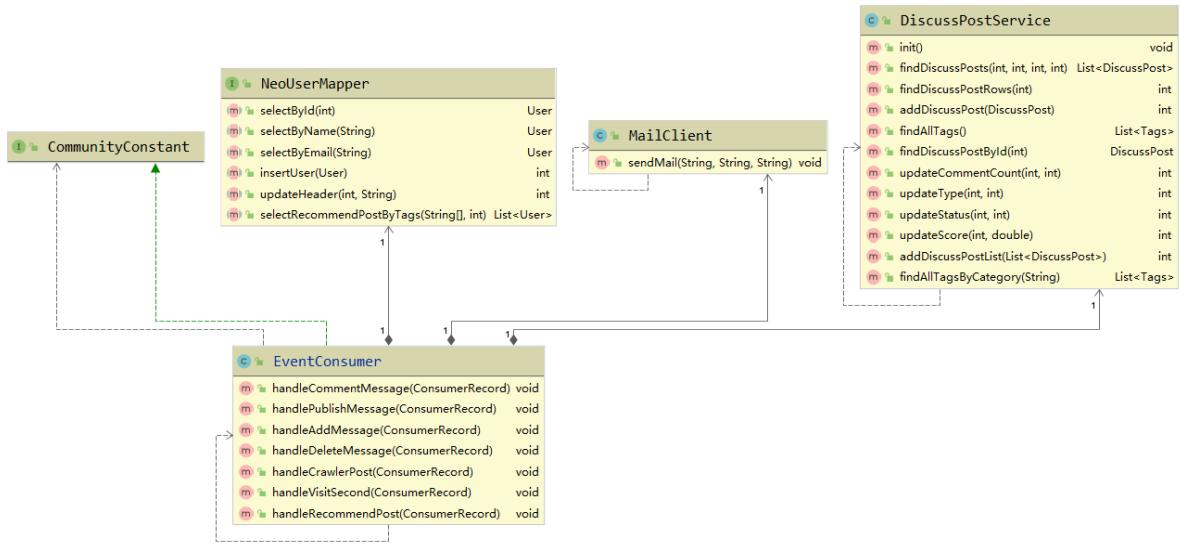


### 4.10.2 Sequence diagram



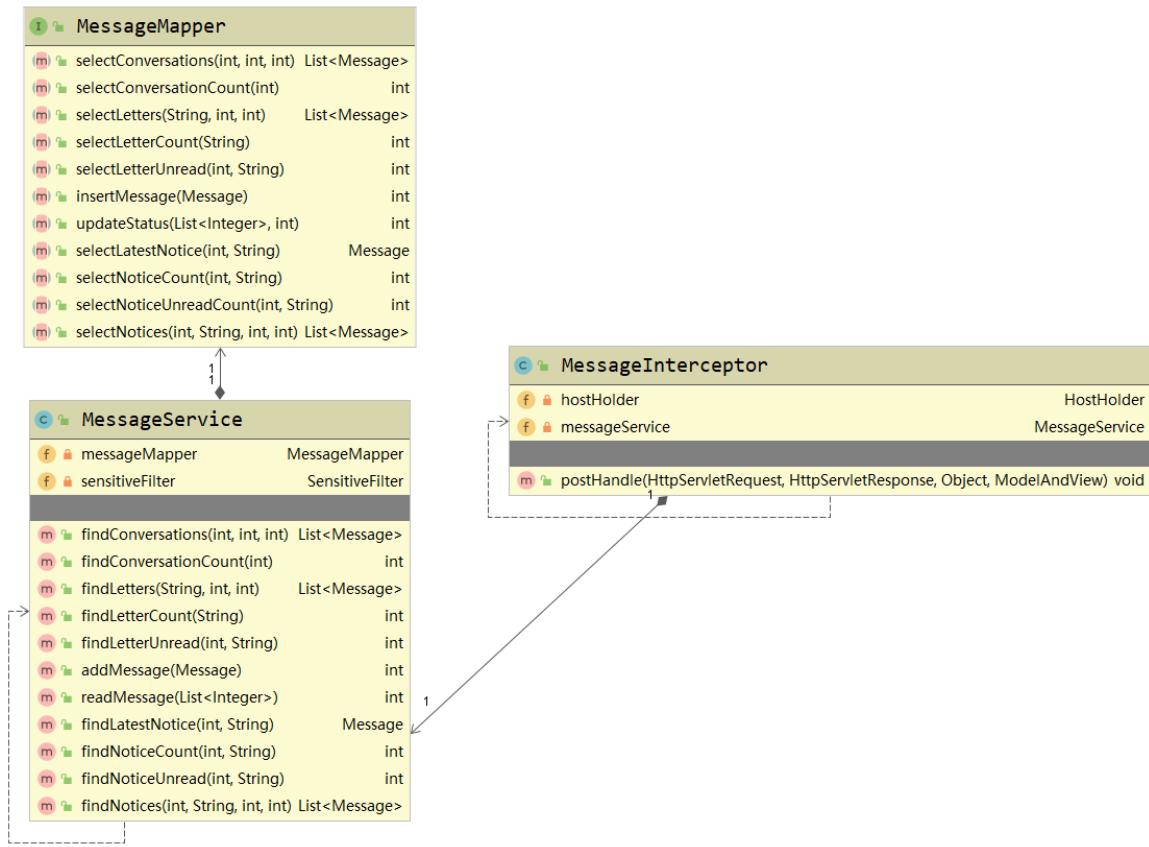
## 4.11 Recommendation feature design

### 4.11.1 Class diagram

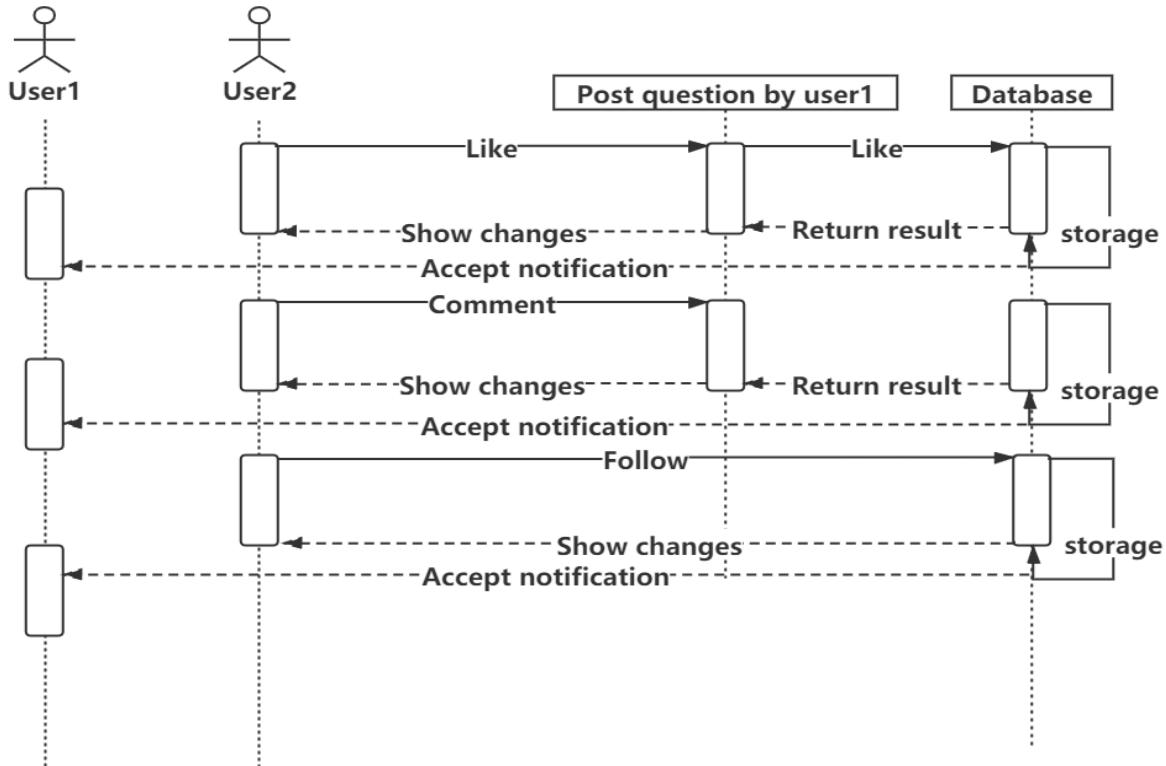


## 4.12 System notification feature design

### 4.12.1 Class diagram

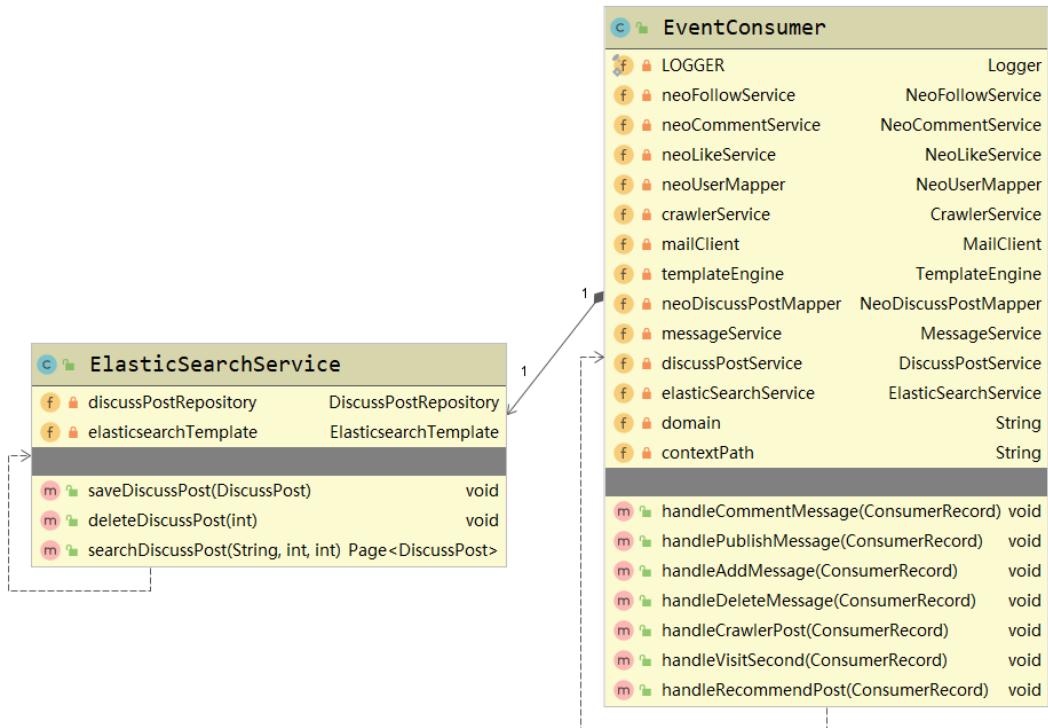


### 4.12.2 Sequence diagram

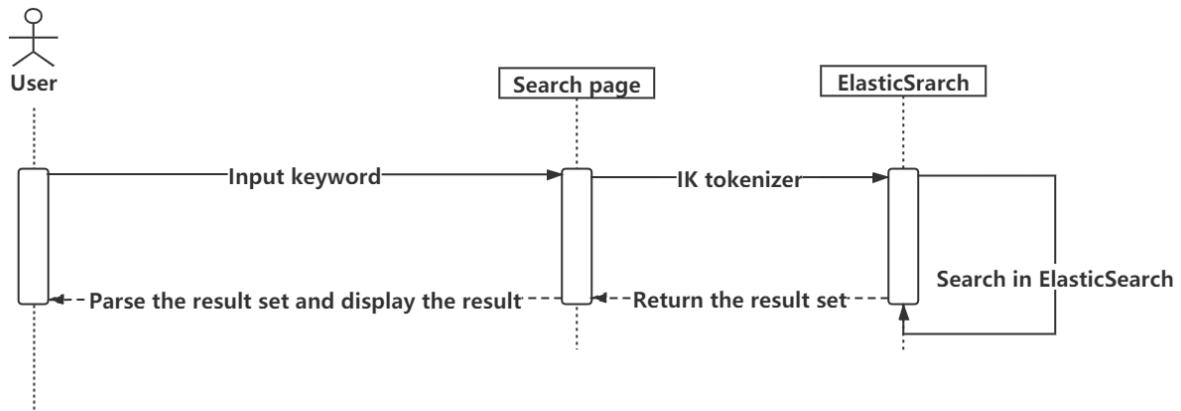


## 4.13 Full text search feature design

### 4.13.1 Class diagram

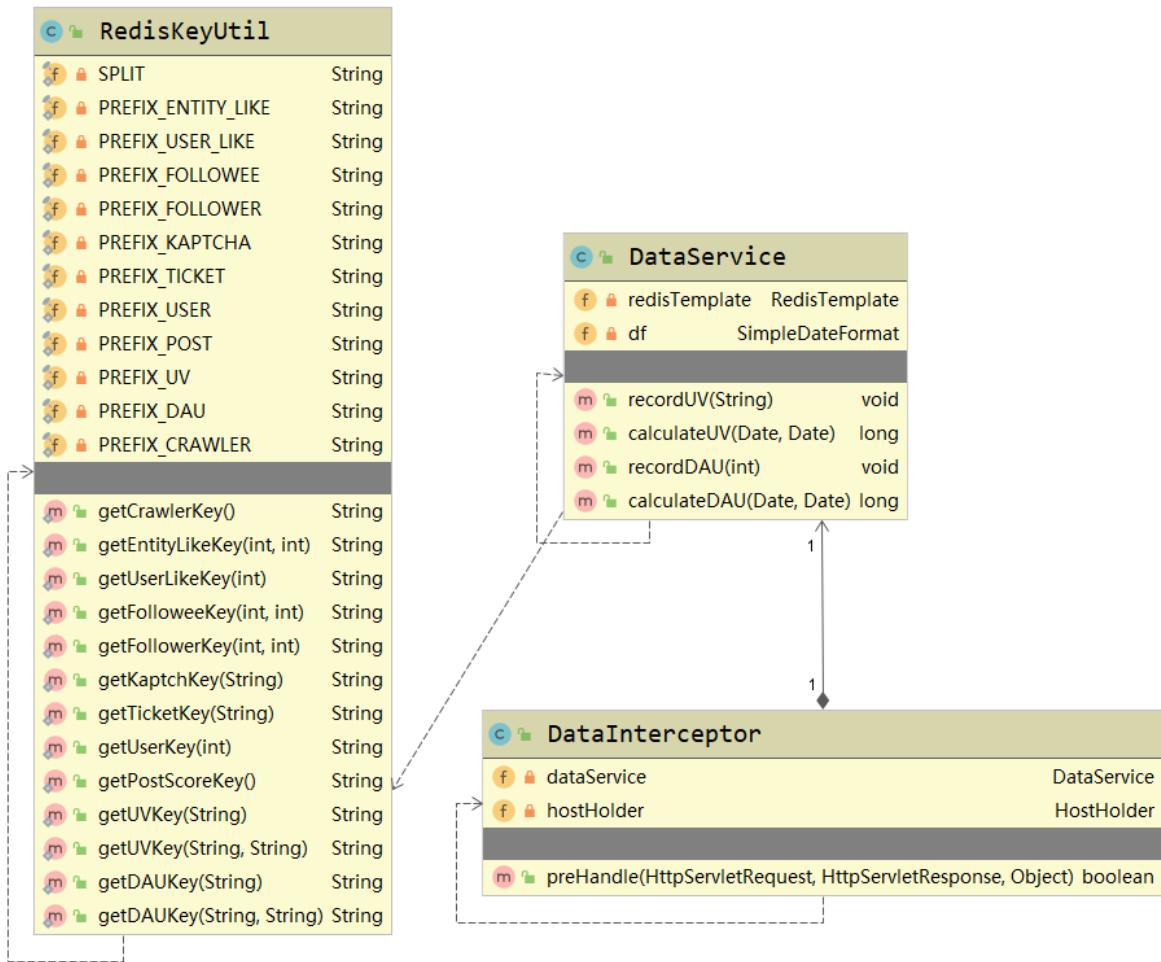


### 4.13.2 Sequence diagram

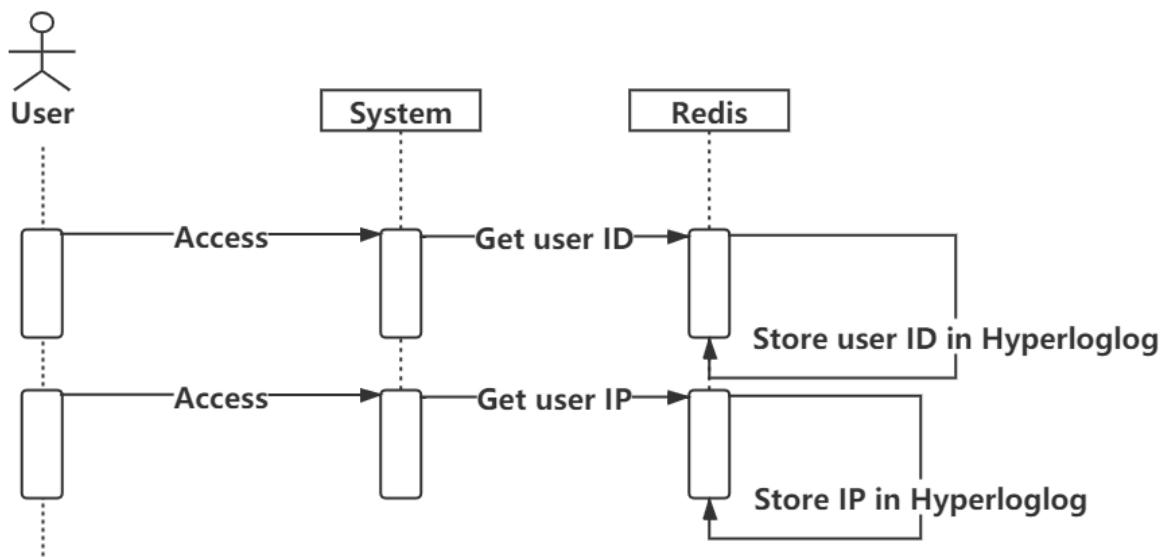


## 4.14 Statistics feature design

### 4.14.1 Class diagram



### 4.14.2 Sequence diagram



## Appendix 5: Test

### 5.1 Unit test results

#### 5.1.1 Mapper test

Function name	Test Data	Expected Outcome	Actual Result	Result Type
<b>selectCommentsByEntity</b>	Status:0; entity_type:2; entity_id:19; offset:4; limit:1	Id:26; User_id:101; Entity_type:2; Entity_id:19; Target_id:0; Content:xwww Status:0 Create_time: 2020-05-06 03:40:19;	Id:26; User_id:101; Entity_type:2; Entity_id:19; Target_id:0; Content:xwww Status:0 Create_time: 2020-05-06 03:40:19;	PASS
<b>selectCountByEntity</b>	Status:0; entity_type:2; entity_id:19;	Count = 6	Count = 6	PASS
<b>insertComment</b>	User_id=101; Entity_type=2; Entity_id=19; targetId=0; content:test; status=0; create_time=now time	1	1	PASS
<b>selectCommentById</b>	Id=26	Id:26; User_id:101;	Id:26; User_id:101;	PASS

		Eneity_type:2; Entity_id:19; Target_id:0; Content:xwww Status:0 Create_time: 2020-05-06 03:40:19;	Eneity_type:2; Entity_id:19; Target_id:0; Content:xwww Status:0 Create_time: 2020-05- 06 03:40:19;	
<b>selectDiscussPosts</b>	Offset:0; Limit:10	All data from the top ten questions in the database	All data from the top ten questions in the database	<b>PASS</b>
<b>selectDiscussPostRows</b>	Status≠2; type≠4; user_id=101	23	23	<b>PASS</b>
<b>insertDiscussPost</b>	User_id=4; Title=test1; Content=test1; Type=0; Status=0; Create_time=2020-04-24 17:09:33; Comment_count=0; Score=0;	1	1	<b>PASS</b>
<b>updateCommentCount</b>	Id=1329; Comment_count=2	1	1	<b>PASS</b>
<b>selectDiscussPostById</b>	Id=1329	Id=1329; User_id=4; Title=test1; Content=test1;	Id=1329; User_id=4; Title=test1; Content=test1;	<b>PASS</b>

		Type=0; Status=0; Create_time=2020-04-24 17:09:33; Comment_coount=2; Score=84;	Type=0; Status=0; Create_time=2020-04- 24 17:09:33; Comment_coount=2; Score=84;	
<b>updateType</b>	Id=1329; type=1;	1	1	<b>PASS</b>
<b>updateStatus</b>	Id=1329; Status=1;	1	1	<b>PASS</b>
<b>updateScore</b>	Id=1329; score=100;	1	1	<b>PASS</b>
<b>insertDiscussPostList</b>	One hundred sets of question data are inserted at the same time	100	100	<b>PASS</b>
<b>selectConversations</b>	User_id=105; Offset=0; Limit=10	Id=82; From_id=105; To_id=107; conversation_id=105_107; content=hhh; status=0; create_time=2020-05-15 00:38:06;	Id=82; From_id=105; To_id=107; conversation_id=105_1 07; content=hhh; status=0; create_time=2020-05- 15 00:38:06;	<b>PASS</b>
<b>selectConversationCount</b>	User_id=105;	1	1	<b>PASS</b>
<b>selectLetters</b>	conversationId=105_107; offset=0; limit=10;	Id=82; From_id=105; To_id=107;	Id=82; From_id=105; To_id=107;	<b>PASS</b>

		conversation_id=105_107; content=hhh; status=0; create_time=2020-05-15 00:38:06;	conversation_id=105_1 07; content=hhh; status=0; create_time=2020-05- 15 00:38:06;	
<b>selectLetterCount</b>	conversationId=105_107;	1	1	PASS
<b>selectLetterCount</b>	conversationId=105_107;	3	3	PASS
<b>insertMessage</b>	From_id=105; To_id=107; conversationId=105_107; content=hhh; status=0; create_time=2020-05-15 00:38:06	1	1	PASS
<b>updateStatus</b>	List of id=79,81,82; Status=1;	3	3	PASS
<b>selectLatestNotice</b>	User_id=105; Topic=comment;	Id=73; From_id=1; To_id=105; conversationId=commnet; content=xxxx; status=1; create_time= 2020-05-06 03:40:19;	Id=73; From_id=1; To_id=105; conversationId=commn et; content=xxxx; status=1; create_time= 2020-05- 06 03:40:19;	PASS
<b>selectNoticeCount</b>	User_id=105;	2	2	PASS

	Topic=comment;			
<b>selectNoticeUnreadCount</b>	User_id=105; Topic=comment;	0	0	<b>PASS</b>
<b>selectNotices</b>	User_id=105; Topic=comment; Offset=0; Limit=10;	Id=73; From_id=1; To_id=105; conversationId=commnet; content=xxxx; status=1; create_time= 2020-05-06 03:40:19;	Id=73; From_id=1; To_id=105; conversationId=commn et; content=xxxx; status=1; create_time= 2020-05- 06 03:40:19;	<b>PASS</b>
<b>selectById</b>	User_id=105;	Id=105; Username= Genshushu; Password= 573e385e b7b2e8f8d10d 96a798cb3a40; Salt= 66428; Email=2686224016 @qq.com; Type=0; Status=1; activation_code= 80adbc fdfccc47fea34b 92c9bd21de2e; header_url= <a href="http://images">http://images</a> .	Id=105; Username= Genshushu; Password= 573e385e b7b2e8f8d10d 96a798cb3a40; Salt= 66428; Email=2686224016 @qq.com; Type=0; Status=1; activation_code= 80adbc fdfccc47fea34b 92c9bd21de2e; header_url=	<b>PASS</b>

		nowcoder.com/ head/856t.png create_time= 2020-04-25 19:18:09;	<a href="http://images.nowcoder.com/">http://images.nowcoder.com/</a> head/856t.png create_time= 2020-04-25 19:18:09;	
<b>selectByName</b>	Username=Genshushu	Id=105; Username= Genshushu; Password= 573e385e b7b2e8f8d10d 96a798cb3a40; Salt= 66428; Email=2686224016 @qq.com; Type=0; Status=1; activation_code= 80adbc fdfccc47fea34b 92c9bd21de2e; header_url= <a href="http://images.nowcoder.com/">http://images.nowcoder.com/</a> head/856t.png create_time= 2020-04-25 19:18:09;	Id=105; Username= Genshushu; Password= 573e385e b7b2e8f8d10d 96a798cb3a40; Salt= 66428; Email=2686224016 @qq.com; Type=0; Status=1; activation_code= 80adbc fdfccc47fea34b 92c9bd21de2e; header_url= <a href="http://images.nowcoder.com/">http://images.nowcoder.com/</a> head/856t.png create_time= 2020-04-25 19:18:09;	<b>PASS</b>

<b>selectByEmail</b>	Email=2686224016@qq.com	Id=105; Username= Genshushu; Password= 573e385e b7b2e8f8d10d 96a798cb3a40; Salt= 66428; Email=2686224016 @qq.com; Type=0; Status=1; activation_code= 80adbc fdfccc47fea34b 92c9bd21de2e; header_url= <a href="http://images.nowcoder.com/head/856t.png">http://images.nowcoder.com/head/856t.png</a> create_time= 2020-04-25 19:18:09;	Id=105; Username= Genshushu; Password= 573e385e b7b2e8f8d10d 96a798cb3a40; Salt= 66428; Email=2686224016 @qq.com; Type=0; Status=1; activation_code= 80adbc fdfccc47fea34b 92c9bd21de2e; header_url= <a href="http://images.nowcoder.com/head/856t.png">http://images.nowcoder.com/head/856t.png</a> create_time= 2020-04-25 19:18:09;	<b>PASS</b>
<b>insertUser</b>	Username=AndrewDYF; Password=573e385e b7b2e8f8d10d 96a798cb3a40; Salt=sd15; Email=	1	1	<b>PASS</b>

	duzong007 @163.com Type=0; Status=1; activation_code=222f97 d413a441 059e307a d59bfcb23b; header_url= <a href="https://avatars2.githubusercontent.com/45820270?v=4">https://avatars2.githubusercontent.com/45820270?v=4</a> ; create_time=2020-05-15 00:35:20			
<b>updateStatus</b>	User_id=107; Status=1;	1	1	PASS
<b>updateHeader</b>	User_id=107; header_url= <a href="http://q99qgd8f7.bkt.gdipper.com/">http://q99qgd8f7.bkt.gdipper.com/</a> fd2074de732 c494ca1653d 02a0c232e5	1	1	PASS
<b>updatePassword</b>	User_id=107; Password=123;	1	1	PASS
<b>insertComment</b>	User_id=105; postId=1338;	Create relationship and return 1	Create relationship and return 1	PASS
<b>updateCommentCount</b>	postId=1338; comment_count=1;	1	1	PASS

<b>insertCrawler</b>	List of questions (100)	100	100	<b>PASS</b>
<b>insertCrawlerTags</b>	List of tags(100) and category of tag	Create relationship and return 100	Create relationship and return 100	<b>PASS</b>
<b>selectTagByTagName</b>	Tagname=java	Java	Java	<b>PASS</b>
<b>selectTags</b>	NULL	Return all tags name	Return all tags name	<b>PASS</b>
<b>selectTagsByDiscussPostId</b>	postID=1338;	Raise	Raise	<b>PASS</b>
<b>selectZeroReply</b>	User_id=105;	Return all zero-reply questions posted by users who do not have an account ID of 105	Return all zero-reply questions posted by users who do not have an account ID of 105	<b>PASS</b>
<b>insertDiscussPost</b>	User_id=4; Title=test1; Content=test1; Type=0; Status=0; Create_time=2020-04-24 17:09:33; Comment_count=0; Score=0; Link_url=xxx (The problem comes from the Internet that has this attribute)	1	1	<b>PASS</b>
<b>insertTags</b>	Tag name=c sharp;	1	1	<b>PASS</b>
<b>insertRelationDiscussPost</b>	Post_id=1338; Tagname=raise; User_id=101;	2	2	<b>PASS</b>
<b>updateDiscussPostStatus</b>	Post_id=1338;	1	1	<b>PASS</b>

	Status=0;			
<b>updateDiscussPostScore</b>	Post_id=1338; Status=100;	1	1	PASS
<b>selectTagsTagNumber</b>	Tag name=java	460	460	PASS
<b>insertVisitSecond</b>	postId=1338; userId=107; second=20;	Create relationship and return 1	Create relationship and return 1	PASS
<b>selectVisitSecondByUserId</b>	userId=105; post_id=1338;	Second=20	Second=20	PASS
<b>updateVisitSecond</b>	userId=105; post_id=1338; second=50;	1	1	PASS
<b>selectAllCategory</b>	NULL	Return all category names	Return all category names	PASS
<b>selectAllTagsByCategory</b>	categoryName=opengl	Returns all tag names in this category	Returns all tag names in this category	PASS
<b>insertFollow</b>	User_id=105; entityId=107;	Create relationship and return 1	Create relationship and return 1	PASS
<b>deleteFollow</b>	User_id=105; entityId=107;	delete relationship and return 1	delete relationship and return 1	PASS
<b>insertLike</b>	User_id=105; Post_id=1338;	Create relationship and return 1	Create relationship and return 1	PASS
<b>deleteLike</b>	User_id=105; Post_id=1338;	delete relationship and return 1	delete relationship and return 1	PASS
<b>selectById</b>	User_id=105;	Id=105; Username= Genshushu; Password= 573e385e	Id=105; Username= Genshushu; Password= 573e385e	PASS

		b7b2e8f8d10d 96a798cb3a40; Salt= 66428; Email=2686224016 @qq.com; Type=0; Status=1; activation_code= 80adbc fdffffcc47fea34b 92c9bd21de2e; header_url= <a href="http://images">http://images</a> . nowcoder.com/ head/856t.png create_time= 2020-04-25 19:18:09;	b7b2e8f8d10d 96a798cb3a40; Salt= 66428; Email=2686224016 @qq.com; Type=0; Status=1; activation_code= 80adbc fdffffcc47fea34b 92c9bd21de2e; header_url= <a href="http://images">http://images</a> . nowcoder.com/ head/856t.png create_time= 2020-04-25 19:18:09;	
<b>insertUser</b>	Username=AndrewDYF; Password=573e385e b7b2e8f8d10d 96a798cb3a40; Salt=sd15; Email= duzong007 @163.com Type=0; Status=1;	1	1	<b>PASS</b>

	activation_code=222f97 d413a441059e307a d59bfcb23b; header_url= <a href="https://avatars2.githubusercontent.com/u/45820270?v=4">https://avatars2.githubusercontent.com/u/45820270?v=4</a> ; create_time=2020-05-15 00:35:20			
<b>updateHeader</b>	User_id=107; header_url= <a href="http://q99qgd8f7.bkt.gdipper.com/">http://q99qgd8f7.bkt.gdipper.com/</a> fd2074de732c494ca1653d 02a0c232e5	1	1	PASS
<b>selectRecommendPostByTags</b>	Tag name=java; AuthorId=101	List of user email	List of user email	PASS
<b>selectHotTags</b>	NULL	Return the top ten tag names according to usage ranking	Return the top ten tag names according to usage ranking	PASS
<b>selectPostByTag</b>	Tag name = java; Offset=0; Limit=10;	Return ten questions related to this tag name	Return ten questions related to this tag name	PASS
<b>insertCategory</b>	Category name=CPP	1	1	PASS

### 5.1.2 Crawler test

Function name	Test Data	Expected Outcome	Actual Result	Result Type
CrawlerSearchTest	Key word = java	100 new data inserted in the database	100 new data inserted in the database	PASS

### 5.1.3 ElasticSearch test

Function name	Test Data	Expected Outcome	Actual Result	Result Type
testInsert	postId=300320	1	1	PASS
testDelete	NULL	Clear ALL	Clear ALL	PASS
testSearchByTemplate	Key words=same project	Return the question data, and add the matching keywords to the data	Return the question data, and add the matching keywords to the data	PASS

### 5.1.4 Kafka test

Function name	Test Data	Expected Outcome	Actual Result	Result Type
testKafka	Topic=test; Content1=hello; Content2= what's up;	Automatically receive the contents of two messages in the console	Automatically receive the contents of two messages in the console	PASS

### 5.1.5 E-mail test

Function name	Test Data	Expected Outcome	Actual Result	Result Type
testMail	Email address= <a href="mailto:2686224016@qq.com">2686224016@qq.com</a> ; Subject=test; Content=test;	Receive this email at this email address	Receive this email at this email address	PASS

### 5.1.6 Redis test

Function name	Test Data	Expected Outcome	Actual Result	Result Type
testStrings	redisKey= test:count; value=1;	test: count value is 1	test: count value is 1	PASS
testSet	redisKey= test:set; value1=123; value2=123; value3=789;	test: set values are 123 and 789	test: set values are 123 and 789	PASS

## 5.2 Functional test results

Use case	Test Description	Test Data	Expected Result	Actual Result	Result Type
Register	Enter the correct information	Username=TestAccount3 Password=123 Email= <a href="mailto:xxxxx@gmail.com">xxxxx@gmail.com</a>	Send activation email and return to homepage	Send activation email and return to homepage	PASS
	Enter existing username or email	Username=Genshushu Password=123 Email=2686224016@qq.com	An error message is displayed because the account or email already exists	An error message is displayed because the account or email already exists	PASS
Activate account	Account not activated	Click on the email to activate the address	Show successful activation information and jump to the login page	Show successful activation information and jump to the login page	PASS
	Account activated	Repeat access to activation address	Show account activation information, and jump to the login page	Show account activation information, and jump to the login page	PASS
	Invalid activation code for account activation	Click on the email to activate the address	Show activation code error, and jump to the registration page	Show activation code error, and jump to the registration page	PASS

<b>Sign in</b>	Enter the correct username, password and verification code	Username=Genshushu Password=123456 Verification code=8byy	Sign in success, Jump to homepage	Sign in success, Jump to homepage	<b>PASS</b>
	Enter the wrong username or password and the correct verification code	Username=Genshus Password=123456 Verification code=8byy	Display wrong username or password and return to login page	Display wrong username or password and return to login page	<b>PASS</b>
	Enter the correct username or password and wrong verification code	Username=Genshushu Password=123456 Verification code=821g	Display verification code error message and return to the login page	Display verification code error message and return to the login page	<b>PASS</b>
	GitHub login	GitHub account has set a public mailbox	Sign in success, Jump to homepage	Sign in success, Jump to homepage	<b>PASS</b>
	GitHub login	GitHub account does not have a public email	Jump to the mailbox binding page for mailbox binding	Jump to the mailbox binding page for mailbox binding	<b>PASS</b>
<b>Post question</b>	Enter complete post information	Title=test Content=test Tag=java	Display successful release information and return to the home page	Display successful release information and return to the home page	<b>PASS</b>

	Enter incomplete post information	Question title or content is empty	Display the failure message and refill the problem information	Display the failure message and refill the problem information	<b>PASS</b>
<b>Comment</b>	Comment	abc	The page does not refresh, and the content is displayed in the comment area	The page does not refresh, and the content is displayed in the comment area	<b>PASS</b>
	Reply	abc	The page does not refresh, and the content is displayed in the comment area	The page does not refresh, and the content is displayed in the comment area	<b>PASS</b>
<b>Like</b>	Like for question	Click the like button	The page does not refresh, the number of likes changes, and like becomes liked	The page does not refresh, the number of likes changes, and like becomes liked	<b>PASS</b>
	Like for comment	Click the like button	The page does not refresh, the number of likes changes, and like becomes liked	The page does not refresh, the number of likes changes, and like becomes liked	<b>PASS</b>

	Like for reply	Click the like button	The page does not refresh, the number of likes changes, and like becomes liked	The page does not refresh, the number of likes changes, and like becomes liked	<b>PASS</b>
<b>Unlike</b>	Unlike for question	Click the liked button	The page does not refresh, the number of likes changes, and liked becomes like	The page does not refresh, the number of likes changes, and liked becomes like	<b>PASS</b>
	Unlike for comment	Click the liked button	The page does not refresh, the number of likes changes, and liked becomes like	The page does not refresh, the number of likes changes, and liked becomes like	<b>PASS</b>
	Unlike for reply	Click the liked button	The page does not refresh, the number of likes changes, and liked becomes like	The page does not refresh, the number of likes changes, and liked becomes like	<b>PASS</b>
<b>Follow</b>	Follow users	Click the follow button	Page does not refresh, follow becomes unfollow	Page does not refresh, follow becomes unfollow	<b>PASS</b>

<b>Unfollow</b>	Unfollow users	Click the unfollow button	Page does not refresh, unfollow becomes follow	Page does not refresh, unfollow becomes follow	<b>PASS</b>
<b>Search</b>	Search all resources	Enter keywords	Return the question title and content that match the keyword, and mark the matched keyword in red	Return the question title and content that match the keyword, and mark the matched keyword in red	<b>PASS</b>
<b>Crawler</b>	Crawl question data from StackOverflow	Enter tags that does not exist in this system	The system can host the crawling data and remind users by email after the crawling is completed.	The system can host the crawling data and remind users by email after the crawling is completed.	<b>PASS</b>
		Enter tags that already exists in this system	The system shows that the tag already exists, the email reminds the crawl is complete	The system shows that the tag already exists, the email reminds the crawl is complete	<b>PASS</b>
		Failed during crawling	The crawler is terminated, the	The crawler is terminated, the	<b>PASS</b>

			email reminds that the crawling data failed	email reminds that the crawling data failed	
		Stored procedure failed	The crawler is terminated, the email reminds the reason for the failure to store the data	The crawler is terminated, the email reminds the reason for the failure to store the data	<b>PASS</b>
<b>Sort</b>	Latest sort	View question list order	The latest question is at the top	The latest question is at the top	<b>PASS</b>
	Hottest sort	View question list order	The hottest question is at the top	The hottest question is at the top	<b>PASS</b>
		Like, comment or refine new question	The problem appears at the top of the list after 5 minutes	The problem appears at the top of the list after 5 minutes	<b>PASS</b>
<b>Zero-reply questions</b>	User is not logged in	Visit the question list page	Display the titles of zero-reply questions in ascending order of publication time	Display the titles of zero-reply questions in ascending order of publication time	<b>PASS</b>
	The user is logged in and has liked or	Visit the question list page	Display the titles of zero-reply questions in	Display the titles of zero-reply questions in	<b>PASS</b>

	commented before		ascending order of publication time	ascending order of publication time	
	The user is logged in and has not liked or commented before	Visit the question list page	Display the titles of zero-reply questions that may be of interest to the user in ascending order of publication time	Display the titles of zero-reply questions that may be of interest to the user in ascending order of publication time	<b>PASS</b>
<b>Related questions</b>	There are similar questions in the system	Visit the question detail page	Display the title, author and abstract of the relevant question	Display the title, author and abstract of the relevant question	<b>PASS</b>
	No similar problems in the system	Visit the question detail page	Hide related question areas	Hide related question areas	<b>PASS</b>
<b>Sensitive word filtering</b>	Post question	Include sensitive words in posted content(drug) For example, zbddurgsdfs, zfdsfd@#\$r#u^gsdf, etc.	Zbd***sdfs Zfdsf***%gsdf	Zbd***sdfs Zfdsf***%gsdf	<b>PASS</b>
	Comment				
	Reply				
<b>Recommendation</b>	Push recommended questions via email	When other users post a new question, the system recommends it to the corresponding user through the	The user received a recommendation email	The user received a recommendation email	<b>PASS</b>

		actions of all users, including comments, likes, or visits for more than 20 seconds			
<b>Message</b>	Like message	The user ID is 105, and the user 107 likes the question posted by	User 107 receives system notification	User 107 receives system notification	<b>PASS</b>
	Comment message	User ID 105, comment on the question posted by user 107	User 107 receives system notification	User 107 receives system notification	<b>PASS</b>
	Follow message	User ID is 105, follow the questions posted by user 107	User 107 receives system notification	User 107 receives system notification	<b>PASS</b>
<b>Online chat</b>	Chat message	User ID = 105 sends a message to user ID = 107	Received message sent by user ID 105	Received message sent by user ID 105	<b>PASS</b>
	Online	User ID = 105 sends a message to user ID = 107	User ID is 107 real-time received messages sent by user ID 105	User ID is 107 real-time received messages sent by user ID 105	<b>PASS</b>
<b>Account set</b>	System and GitHub users modify avatar information	Change avatar	Change avatar	Change avatar	<b>PASS</b>
	System user repair account password	change the password	The modification is successful, return to the	The modification is successful, return to the	<b>PASS</b>

			login page, and log in again	login page, and log in again	
<b>Manage questions</b>	Delete question	Delete question with question ID 1338	Successfully deleted, and return to the home page	Successfully deleted, and return to the home page	<b>PASS</b>
	Set as sticky question	Sticky question with question ID 1337	Set up successfully, the top position in the question list page	Set up successfully, the top position in the question list page	<b>PASS</b>
	Set as the essence question	Essence question with question ID 1337	The setting is successful, it is the top position in the popularity sorting page	The setting is successful, it is the top position in the popularity sorting page	<b>PASS</b>
<b>Manage tags</b>	Add category name	category name = C Sharp	Added successfully, and displayed in the list	Added successfully, and displayed in the list	<b>PASS</b>
	Add tag name	Tag name = C#	Added successfully, and displayed in the list	Added successfully, and displayed in the list	<b>PASS</b>
<b>Manage authority</b>	Super Administrator	Log in as super administrator	Manage question and tags	Manage question and tags	<b>PASS</b>

	Administrator	Log in as administrator	Manage question	Manage question	<b>PASS</b>
	User	Log in as user	Post, comment, like questions	Post, comment, like questions	<b>PASS</b>

### 5.3 Usability test feedback

The following is the feedback from the available tests:

- The correct answer appears, or the available solution should be highlighted or reminded to the user that this answer is the correct answer.
- The system should improve user feedback features such as logo, sound and colour.
- The tags displayed on the publishing page should be pagination in pages.
- Improve the display speed of the problem page as much as possible.