Status Report
for
ECMS Data Annotator

Prepared for

Mr. Parma Nand

Mr. Petteri Kaskenpalo

Supervisor

Mr. Lingmin He

Prepared by

Zheyang Cao

Binbin Ni

Penglin Jin

Zixin Jiang

Date: 3/9/2021

Version 1.5

Menu

Executive Summary

This is a mid-term status report on our group's R&D project, which consists of the following main parts.

First is a brief project description, which gives a short account of our project's functions, etc. Next is the approach we have used in the project so far, such as how we have communicated and collaborated. Also, it is very important to show what we have accomplished so far in terms of tasks and functional modules. This is followed by the features that we still need to complete and improve, and the problems that we have as a group, some of which have been solved and some of which are still to be solved.

Finally, the report shows what each member of the group has accomplished and contributed. This is so that we can clearly see where we stand and how we can work better in the future.

Project Description

The initial goal of this project is to annotate ASCII files, which is the most basic and important feature. In addition, as part of this project or as a follow up project, our tool should also add the ability to tag images.

Our project is based on web development in java language, the backend will be developed using SpringBoot framework for functionality and the front-end pages will be developed using Vue.js framework. The basic functionality of the project includes simple registration, login and role management. Users can upload .txt files and json format files and images, and then manually annotate these datasets and images. Collaborative annotation features can be implemented between users on the web page.

We will also add and improve the natural language pre-processing related functions, so that the machine will annotate all the datasets for us in advance when we need it, and the annotator only needs to check whether those annotations need to be modified.

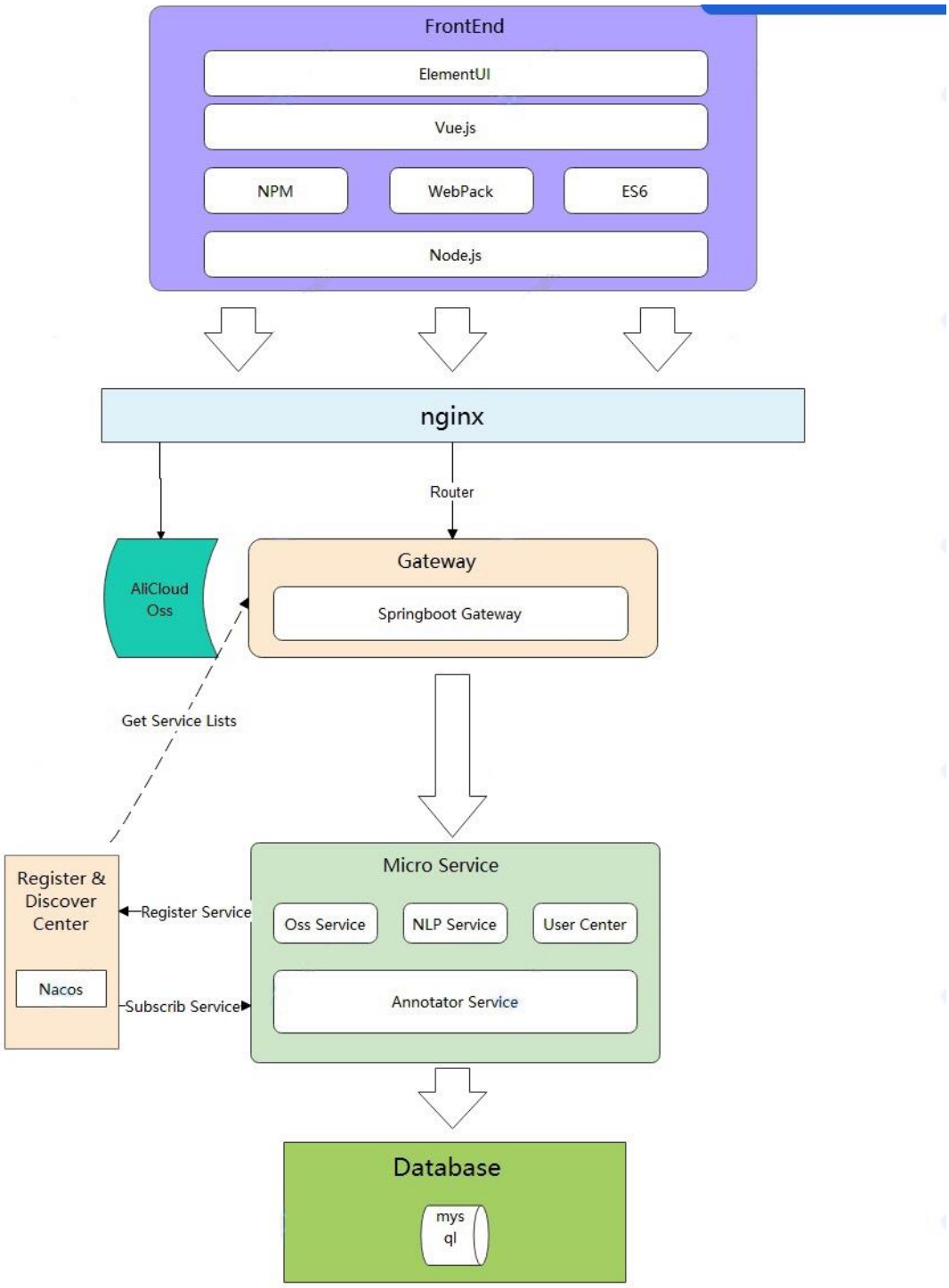

Figure 1 System Structure

Project Scope

Objective

The goal of this project is to build a tool that can be used to import and annotate datasets. The tool should be simple and intuitive so that non-technical users can easily use it. Moreover, this tool is very useful for artificial intelligence and machine learning, which can greatly reduce the cost of machine learning and improve the efficiency of machine learning.

Approach

As planned in the proposal, we used “agile” in the development of the project. We divided the whole project into many sub-projects, such as database building, back-end programming, front-end programming, and so on .For each sub-project, we have a complete build process, from design to implementation to testing. However, for different sub-projects, we still have the order of priority because of the limited team member. What is certain, however, is that in general, all subprojects continue to grow in parallel.

Major Milestones

Milestone	Date	Responsible
Requirement list	2020/12/25	Jiang Zixin
Database construction	2021/1/5	Cao Zheyang
Prototype	2021/1/25	Ni Binbin
UI building	2021/2/5	Ni Binbin
User management module	2021/2/5	Jin Penglin
Authority Management Module	2021/2/17	Jin Penglin
Dataset upload and processing module	2021/2/27	Jin Penglin

Deliverables

Deliverables	
Deliverable 1	Implement a graphical interface, a full - featured annotation tool
Deliverable 2	Implement multi - file or single file upload
Deliverable 3	Implement the output of multiple types of files.
Deliverable 4	Develop management functions to support background log monitoring
Deliverable 5	Implement Multi - person collaborative annotation
Deliverable 6	Use NLP for some data processing and annotation

Response to Recommendations

1. The content of the proposal is too vague, and it fails to analyze and elaborate according to the actual situation of the project.

   Response: Modify the proposal according to the actual situation of the project

2. The proposal does not reflect the characteristics of the tool, in other words, it is the reason why users choose our tool rather than other existing tools.

   Response: In the proposal, feature content is added, features of the tool are summarized and new functions are added. For example, the tool is designed to be front-end and back-end, so users don't need to download the software, and can comment on a Web page.

New variations

1. Multi-person collaborative annotation

The idea comes from the collaborative editing capabilities of office software such as MS Office and WPS.

Given the convenience of collaborative editing of large documents, can collaborative annotation capabilities improve the efficiency of annotation projects for large annotation projects?

With this idea in mind, we developed a collaborative annotation feature.

2. Using NLP for annotation preprocessing

The idea stems from the original purpose of the tool - to provide a prototype for machine learning. So can we use machine learning to help annotate. This can not only improve the efficiency of annotation, but also further improve the learning speed of machine learning. After importing the data set, a preprocessed annotation result is given to the user through NLP, and then the user can modify or accept it directly according to the situation

Current Project Status

Work Completed

Completed Work	Description
Requirement Analysis	Requirements analysis, requirements interface and list making
Database Design & Implement	Database structure design and deployment
Prototype & UI Design	UI interface design
UI Implement	The UI programming by Vue
User management function	Administrators can manage and monitor all users in the background
Permission management function	The user permission grant and the executable action for the corresponding permission
Dataset upload and processing function	Dataset uploading and annotation
Multi-person collaborative annotation function	Multiple users can annotate collaboratively in groups
Image annotation function	Can upload pictures for annotation, and you can display thumbnails after uploading

Uncompleted Work

Uncompleted Work	Description	Planned completion time
Front-end interface	Although the front-end interface has been designed, the page production has not been completed and also	2021.03.25

	need to be optimized.	
Improve the stability	Considering the situation that the tool will face after it is put into use, the stability of the tool currently produced by the team still needs to be improved.	2021.04.27
NLP preprocessing	Preprocessed imported dataset by NLP	2021.4.23

Existing Issue & Improving Recommendations

Existing Issues

1. The initial progress of the project is slow

Before the informal review on January 7, the team has only completed part of the requirements analysis and database design, and the team members have determined the idea of completing the project, but have not discussed with the client.

2. Did not communicate with client in time

The team only communicated with clients at the beginning of the project, and then did not communicate our ideas with clients immediately. When we realize the problem, after completing the needs analysis and UI design, we sent our designed interface and functions to clients for their opinions, which was recognized by clients.

3. The communication among project members was not timely

During the winter vacation, the team members did not complete the work together in the school, we all learned skills at home and completed the follow-up work of the project. So, we did not communicate our respective progress in time, which led to the team members not knowing what other members really completed, and the progress was not unified.

4. Project file management confusion

Team members did not upload files to the sharing platform in time after completing their work. Therefore, team members cannot see what other members have done or are doing on the team's shared platform in time.

Improving Recommendations and Measures

1. Speed up project progress

When aware of the slow progress of the project, team members timely communicate and adjust the current work plan to ensure that the project keeps up with the progress. The current project progress tends to be normal, and then the team will continue to monitor the project progress in real time to ensure that the project is completed on time and with high quality.

2. Strengthen communication with client

Sort out the current status of the project and communicate with client. At present, the team has improved in this aspect. After that, we will continue to communicate with client in the important stage of the project in time, listen to client's opinions and suggestions, and ensure that the project meets client's needs.

3. Regular meetings within the team

Team meetings need to be more frequent and regular. During the meeting, project members report the current progress and later plan. In addition, strive to strengthen team cooperation and improve work efficiency.

4. Project upload GitHub in time

After finishing the current work, upload the work file to GitHub. Inform team members of any significant progress or problems.

Individual Contributions & Learning Achieved

Cao Zheyang

In the R&D project, as the team leader, I was mainly responsible for the process control and task allocation of the project, as well as the compilation and integration of documents. Supervise and control the various processes and sub-projects of the project, keep track of the progress and make adjustments according to the situation. Organize group meetings and guide group members to discuss. Study and analyze document content, build document framework and general content to facilitate team members to understand and write documents during group meetings. Assign tasks properly and review and revise the results. In addition, in the technical aspect, the database structure construction and implementation.

In the process of R&D project development, I have learned a lot of knowledge about project management. Especially with the agile development we use, how to integrate it well with our project is a difficult thing. In addition， it took me a lot of time to learn the logic of database construction.

Ni Binbin

I was responsible for prototyping the project, and since I had no prior exposure to prototyping. So I spent a few days studying and reviewing materials, and I found a lot of good prototyping tools. Among these tools I chose MoDao. It's a free and very popular app and web prototyping tool in China and used by many product managers in companies.

In addition, I was also involved in writing front-end code, because I only know some simple front-end knowledge such as js, css, html, but these skills obviously cannot meet

the needs of the project development, so I learned Vue related technology from some video sites such as Bilibili and applied it to the project code writing.

Jin Penglin

I am responsible for the overall architecture and back-end code writing of this project, and participated in the writing of some front-end Vue components. My main software for back-end development is IntelliJ idea, and the back-end development uses vs code, Both of them are free for students. At present, the main difficulty I encounter lies in the front-end aspect, especially the Vue componentization and front-end asynchronous processing. Now I mainly improve my ability through online documents and video materials, as well as asking for help from students who are familiar with this aspect, and put it into practice in the actual development. In addition, in the back-end aspect, I used springboot as the basic framework of the back-end, combined with the gateway in spring cloud to cooperate with the front-end routing, and used Alibaba cloud OSS platform as the data upload destination.

Jiang Zixin

At present, I am mainly responsible for project management and needs and requirement analysis. I learned this by watching instructional videos. I set up the team work plan and control the project progress, including meeting minutes, project management documents, report writing and so on. This is my first time in charge of project management. I have learned a lot about project management through my practice in recent months. Some problems in the early stage of the project helped me to improve the project plan and realize the importance of time management and effective communication. At the same time, I also continue to learn project management skills, hoping to do better in this area. Now, some problems have been improved, such as communication with client and project progress.

Database Design:

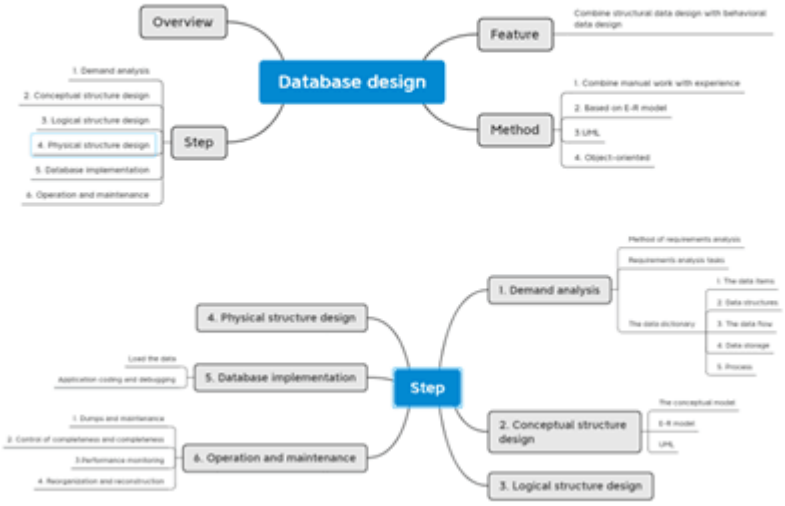

Figure 2 Learning Diary example

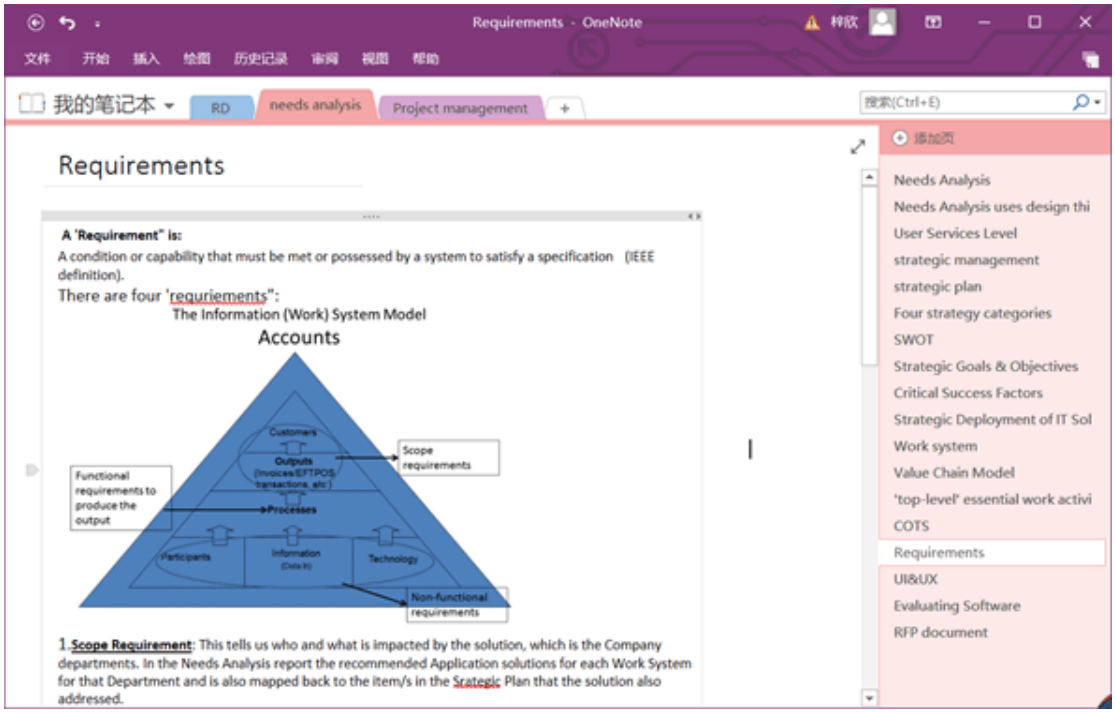

Figure 3 Learning Diary example