

# Progress Report: HyperLink Classification

## **Group Name: TTIS 5**

Captain: Zhengkai Zhang (zz68)

Yuan Chung Ho (ych11)

Wan Feng Cai (wfcai2)

Xinqian Xiang (xinqian6)

Zheng Ma (zhengma3)

## 1) Which tasks have been completed?

For the final project, our main approach is to do the Link Classification.

We defined the database, UI and API for it.

### **Database:**

User table contains ID, name, categories and feedback columns.

Category table contains id, name, and keywords.

### **API:**

For each table, we all have to create, update, delete, getByName, and getAll to get the data from our database.

In addition, we have the API that allows users to send the link and user information to be processed, and the API that receives the user feedback on our decision.

### **UI:**

We have a simple UI for users to send their URL and select the user to identify the classification.

Also, we have the API to create new user records in our database.

For category, we didn't create any UI, but the provided API will allow us to write a script to insert the data.

## 2) Which tasks are pending?

### **Data:**

We need some real data for categories, keywords, and users.

It will require us to write a script to set up the database record, and script to fetch the data from Online.

**Scrape web:**

We need to write a method based on the input url to get all the content inside of that url. It will require work to get the content sometime recursively.

**Get the keywords:**

Based on the content we get in the URL: we will need to find the keywords in the website.

**Decision maker:**

Based on the keyword and database record, we will decide if the user is interested in the url or not.

**UI/API integration and improvement:**

There is still a lot we can do to improve our api and ui, also integrate with later compare algorithms and the scrape part.

### 3) Are you facing any challenges?

Setup for each team member. Since our team has different backgrounds, we need to sync and get everyone on the same page.

Scrape the web and the algorithm to make the decision will be a big part of work for it.

Database insync, we still use local databases. Will need to spend some time to set a remote one or have a unique set up script for our local database.