Cyberminer: A web search Engine

Tingwei Wu  
Computer science  
University of Texas at DallasRichardson,US  
[txw190006@utdallas.edu](mailto:txw190006@utdallas.edu)

Churong Zhang  
Computer science  
University of Texas at DallasRichardson,US  
[cxz173430@utdallas.edu](mailto:cxz173430@utdallas.edu)

Feifei Guo  
Computer science  
University of Texas at DallasRichardson,US  
[fxg180009@utdallas.edu](mailto:fxg180009@utdallas.edu)

Qixiang Wang  
Computer science  
University of Texas at DallasRichardson,US  
[qxw200000@utdallas.edu](mailto:qxw200000@utdallas.edu)

*Abstract*— Cyberminer, which is a simple web search engine, shall accept a list of keywords and return a list of URLs whose description contain any of giver keywords. By implementing this project, Object-Oriented Analysis and Design (OOAD) is applied. This document furnishes the Use Case diagrams, Class diagrams, and Sequence diagrams using UML.

Keywords— Cyberminer, Search Engine, OOAD

# Introduction

A search engine is a [software system](https://en.wikipedia.org/wiki/Software_system) that is designed to carry out [web searches](https://en.wikipedia.org/wiki/Web_search_query). They search the [World Wide Web](https://en.wikipedia.org/wiki/World_Wide_Web) in a systematic way for particular information specified in a textual [web search query](https://en.wikipedia.org/wiki/Web_search_query).[1] Web search engines get their information by web crawling from site to site. The "spider" checks for the standard filename keywords.txt, addressed to it. The keywords.txt file contains directives for search spiders, telling it which pages to crawl and which pages not to crawl. After checking for keywords.txt and either finding it or not, the spider sends certain information back to be indexed depending on many factors, such as the titles, page content, JavaScript, Cascading Style Sheets (CSS), headings, or its metadata in HTML meta tags. For Cyberminer, the spider sends back the title, URL, and page content. These data will be stored in our MySQL database. When a user enters a query in Cyberminer, the search engine will search the database and return a list of results which either its title or its description contains the given keywords. The result can be shown by pages and ascending alphabetical order.

# Domain modeling and Requirements

## Functional requirements

As required, the Cyberminer shall include the below 9 functional requirements.

FR1. Case sensitive search.

FR2. Autofill, while correcting typographical errors

FR3. Specifying OR/AND/NOT Search

FR4. Filtering out symbols that are not meaningful, according to the user configuration

The above four functional requirements are all about the search query, or input keywords. We can combine them into one use case.

FR5. Multiple search engines. Since we use MySQL database, it naturally satisfies this function.

FR6. Setting the number of results to show per page, and navigation between pages.

FR7. Listing of the query result in ascending alphabetical order; most frequently accessed order, or per payment.

FR6 and FR7 are both concern on the search result listing. They can be combined into one use case: search.

FR8. Hyperlink enforcement. When user clicks on one URL, which has been retrieved as the result of a query, the system can take the user to the corresponding web site. To achieve this function, the system’s database should contain URL information of each item, and the URL information shall be returned to page to enforced hyperlink. This also can be one use case.

FR9. Deletion of out-of-data URL. We can set check mechanism in a given frequency, such as one day or one week. It will check if the URL stored in database is or not out of date. If it is out of data and cannot reach again, then delete it. This function has nothing to do with user. This is the server-side work. Therefore, this function should be a single use case.

## Nonfunctional requirements

# Cyberminer shall be easily understandable, portable, enhanceable, and reusable with good performance. This system shall also be user-friendly, responsive, and adaptable. To achieve these nonfunctional requirements, the UI and operation process of our system should be simple, similar to classic web search engine Google. They it will be quite easy for user to use and understand.

* 1. Domain modeling

Fig.1 is this system’s domain model. It contains 4 parts: user, web search engine, server, and database. The main process is like this: a public user open the web search engine, input the query string to search, the sever get the request and search the database, return the results and show it on the web search engine.

Diagram

Description automatically generated

Fig.1 Class Diagram (Domain)

# Design Specification

## Use Case Diagram

By analyze the functional requirement, this system shall include 4 use case: input, search, open URL, and delete of out-of-date URL. Fig2 shows the use case diagram of Cyberminer. There are two actors: public users and server. Here the sever is also means the system.

Diagram

Description automatically generated

Fig.2 Class Diagram (Domain)

## Class Diagram

The system has four classes: user, web search engine, server, and database. Fig.3 shows the class diagram of Cyberminer.

Diagram, schematic

Description automatically generated

Fig.3 Class Diagram

## Sequence Diagram

From the use case diagram and class diagram, we have the sequence diagram, shows in fig.4.

Diagram

Description automatically generated

Fig.4 Sequence Diagram

# Prototype Implementation

We use python flask + MySQL to implement this project.

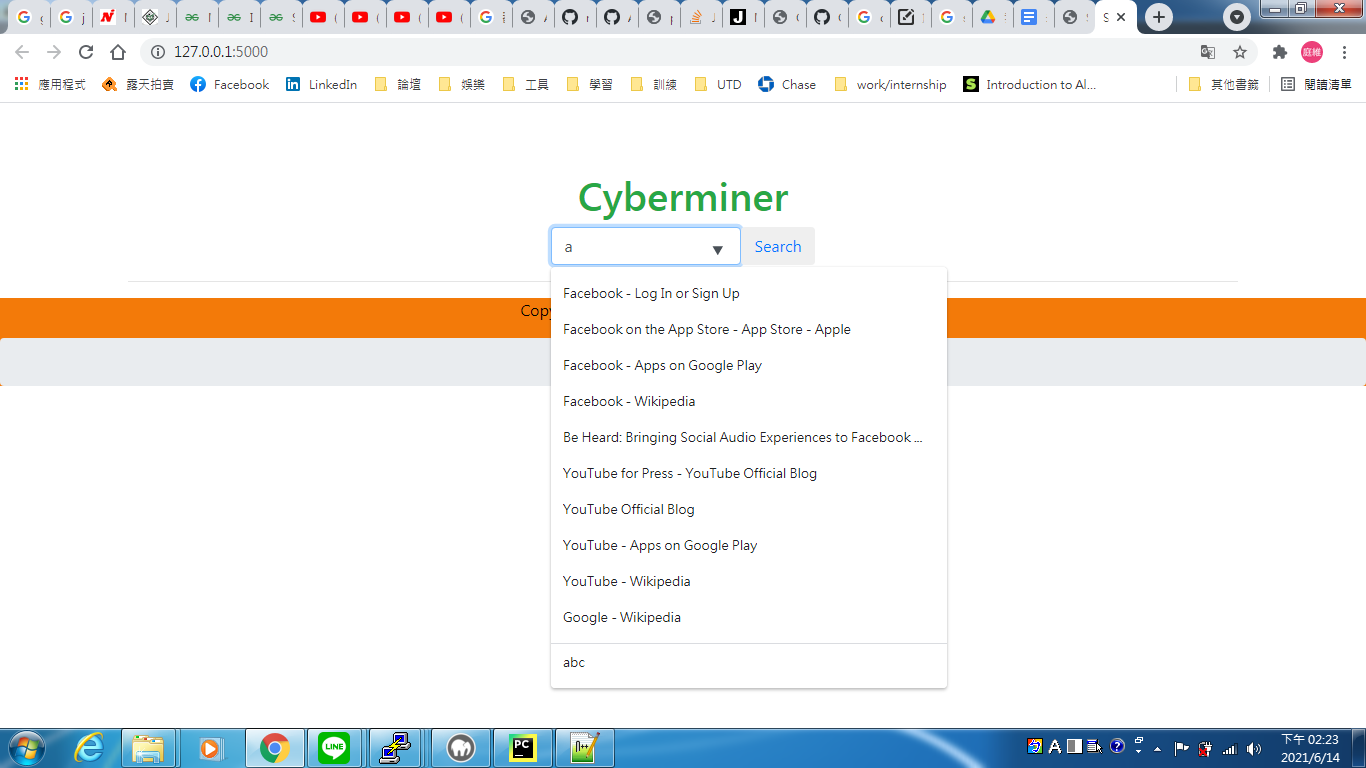


Fig.5 Input query string

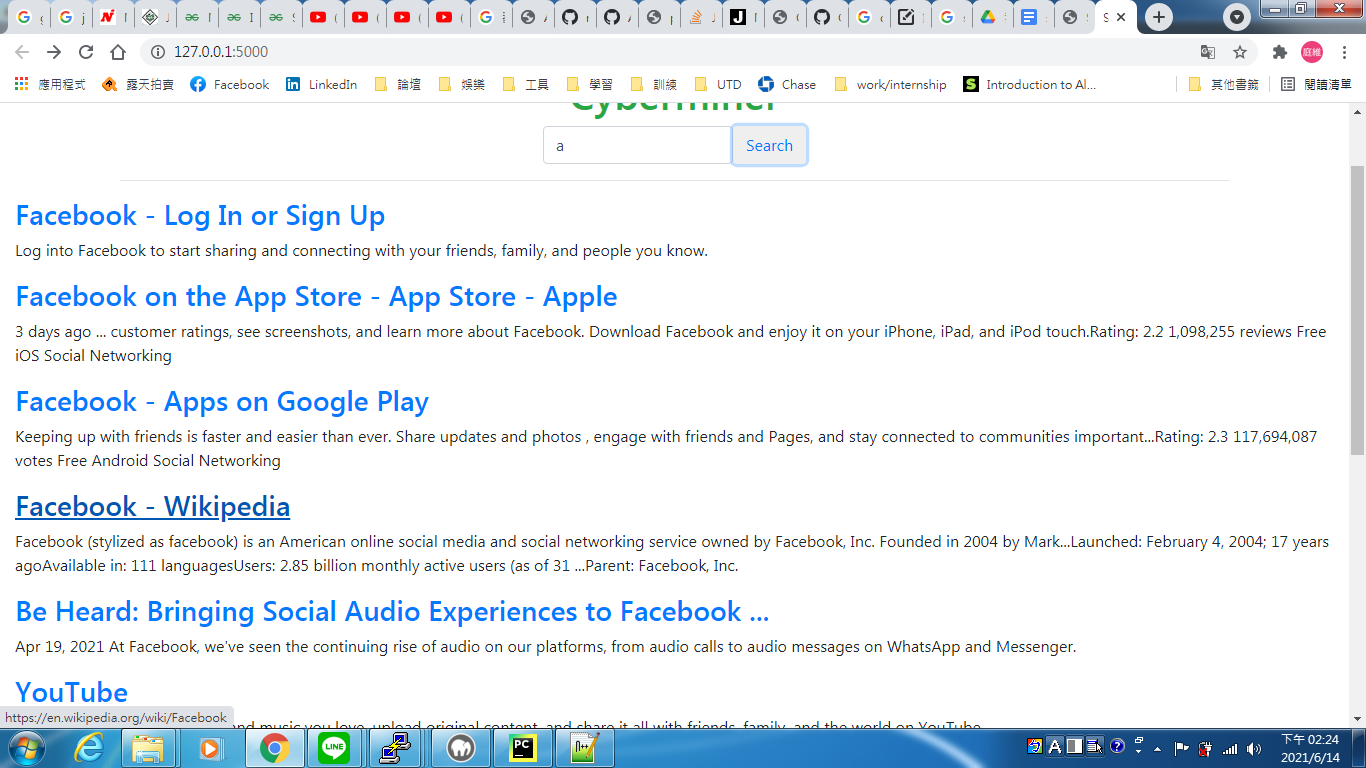


Fig.6 Click search button and show the result

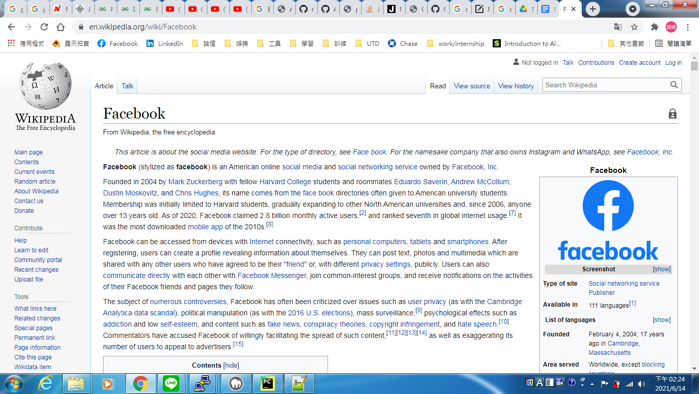


Fig.7 Open URL and hyperlink

# Summary

All the functional and nonfunctional requirements have been enforced.

In future, we can improve in two aspects. First, enlarge the database. Then the search result will be more effective. Second, rate the search result in more effective way. Rathan than listing results in ascending alphabetical order, it shall be listing in as ascending order of correlation index and user frequency.

.

##### References

<https://en.wikipedia.org/wiki/Search_engine>