**CyberMiner - Web Search Engine**

**Phase 1 Final Project Document**

Fredrick Horn

Parker Tate

Collin Matz

Pranit Konda

Conner Carraher

Joshua Brown

Jacob C Danel

Website URL: <https://cs4376-team-a.github.io/Project/>

**Revision History**:

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Revision Date | Description | Author |
| 1.0 | 06/23/2023 | Initial Draft, layout, formatting. | Jacob C Danel |
| 1.1 | 06/27/2023 | Rough Draft of Document: Writing of introduction, adding in various design documents along with descriptions, etc. | Jacob C Danel |
| 1.2 | 6/28/2023 | Updated some diagrams, proofread sections, further fleshed out the implementation section and added associated diagrams as well as a textual and video demonstration of the current iteration of the CyberMiner system being used. | Jacob C Danel, Collin Mitz, Joshua Brown |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. **Introduction**
   1. **Purpose**

The purpose of this search engine, dubbed CyberMiner, is to allow users of the system CyberMiner to be able to search for websites using keywords they submit as input to the CyberMiner system. Said keywords would then be used by CyberMiner to return to the user a list of related websites that the user may choose from. The goal for this process is to be done as quickly and efficiently as possible (i.e., return search results as quickly as possible).

Response time is a very important factor in the development of CyberMiner. Typically search engine/web browser users have come to expect near instantaneous results when utilizing such technologies, so maintaining this expectation with CyberMiner is an important aspect to achieve. In order to effectively do so, the system must be able to promptly obtain input, extract keywords from input, compare keywords to websites’ descriptions in indices table, return relevant websites, generate an easily navigable webpage with a list of the relevant websites, then connect the user to the website whose list entry they interact with. The manner in following this process is fairly straightforward. Input is gotten from a text field which the system then breaks down into a list of keywords by separating the string on typical delimiters (in this case most likely to be empty space). Those keywords are each compared to all entries in CyberMiner’s index table where those keywords are specifically compared to each word in an individual index’s description. This is done for all indices in the table and all relevant websites are stored in a list as this process is happening. Once the relevant website list has been generated it is then displayed on a navigable webpage generated by CyberMiner for the user. The user can then connect to whatever website they choose to click on, as previously stated.

* 1. **Scope of the System**

The scope of the CyberMiner system, in relation to a user’s experience, is as stated ahead. A typical user will open the CyberMiner application on their personal computer which should then display a page similar to other search engines/web browsers where there is a text field where the user can input whatever search queries they may want to make. Once input is given by the user to CyberMiner, CyberMiner would then use that input to generate a page of related websites that the user may then be able to connect to through the CyberMiner application. Once the connection is made the page is updated to be that of the chosen website’s layout and other related information. Should the connection fail however, CyberMiner will instead update the page to display an error message that gives the user a brief description of the error (i.e., connection failure due to no internet connection, a database connection error, etc.)

* 1. **Objectives**

Current objectives:

1. Keep the system as simple as possible for the User, Users do not need to be able to interact with the various components of CyberMiner, just its search functionality.
2. Keep index table search speeds as fast as possible through whatever means necessary
3. Build upon existing version of CyberMiner with added functionality, currently CyberMiner allows for case sensitive searches and hyperlink enforcement  
   1. **References**

* Sample system design document found on Prof. Chung’s website:
* https://personal.utdallas.edu/~chung/CS6354/CS6354\_U07\_source/Team\_1/CAD\_Design\_\_4.doc
* Another Sample system design document found on Prof. Chung’s website: <https://personal.utdallas.edu/~chung/CS6354/CS6354_U07_source/Team_2/deliverable_1_Req_Specification.doc>
* Prof. Chung’s personal webpage where slides and such are found: <https://personal.utdallas.edu/~chung/OOD/syllabus.htm>
* Phase 1 assignment document: https://personal.utdallas.edu/~chung/OOD/Project1.pdf

1. **Domain Model**
   1. **Description**

The domain model is meant to be a more general overview of the system for users to be able to easily understand and to be used as a basis for further design documents and diagrams. In its current iteration the domain model shows the general sequence of events that occur when a user, or admin, interacts with the CyberMiner system. For general users, the functionality needed in order for them to use CyberMiner effectively is primarily giving input and receiving web results that they can then connect to with a click of the mouse. Additional functionality could include such things as allowing users to create bookmarks for websites they frequent often, cataloging and displaying their browsing history, and other basic functionality one would find in other search engines/browsers. However, those added functions aren’t entirely necessary to be able to use CyberMiner effectively, more so they are just quality of life additions.

For Administrators of the CyberMiner system the functionality they would need is a bit more robust compared to general users of the system. On top of the basic functionality needed in order to use the system, admins would need additional functionality that would help them in properly maintaining the system and troubleshooting any problems that may arise during use. For instance, such functionality would be having access to access logs to see where or why a connection was not able to be made with a particular website for a user. Another may be having access to the index table, so that they may make changes to it as needed (creating, updating, fixing entries in the table, etc.).

* 1. **Model** **A picture containing text, screenshot, font, diagram

     Description automatically generated**
  2. **Class Models**

1. **User**

Someone that interacts and utilizes the CyberMiner system in order to search for websites. Primary functionality needed would be what was mentioned before, searching for desired websites based on provided input. There is additional functionality that can be added to the system for users, but it isn’t vital to the use of the system (bookmarks, search history, etc.) Displayed above a user currently has two attribute fields, “name” and “email”. Although not necessary for general use of the system, a user can provide some information like their name and email address.

1. **Admin**

A user that has more permissions/functionality with the CyberMiner system than a typical user would have. Examples of additional permissions/functionality would anything that would help facilitate their ability to maintain the system. For instance, having access to and the ability to change or manipulate the index table, as well as individual indices in the index table. Displayed in the model an Admin has three attributes: “id”, “name”, and “email”. Their “id” attribute would likely be used in the authentication process when an admin would attempt to make changes to the index table. “Name” and “email” are the same as those in user.

1. **CyberMiner**

This the primary system that both users and admins will primarily be interacting with. In this version of the domain model, it has one attribute field called “text\_field”. Text\_field represents the text box in which input will be received, the attribute itself will store the input taken in by the system that is then later broken down into keywords.

1. **SearchQuery**

A search query is generated after the CyberMiner system has been given some input by a user or admin. It has two attributes called “keywords” and “timestamp”. The keywords attribute is essentially a list that is created from the input given to CyberMiner that was stored in the text\_field attribute, CyberMiner breaks up the input based on delimiters in the input and stores the values in between those delimiters into the keywords list. The timestamp attribute is a value which stores the time when the query is generated. This attribute would primarily be used for troubleshooting reasons by an admin.

1. **WebPage**

A webpage in the system is a visual that a user or admin sees which displays the list of websites obtained from the index table that are relevant to the input given by the user or admin. It has a number of attributes, those of which are: “site\_name”, “url”, “description”, “query\_id”. The first three would be visible by any user of the system while the query\_id attribute is primarily for admins for troubleshooting purposes. “Site\_name” is as the name suggests, the generalized name of the website, “url” is the web address link that is used to make connections to said website, “description” is a basic description of what it is the website is about or what the website offers.

1. **Web**

As stated in the diagram above it is the collection of websites that are connected to one another.

1. **Requirements Specifications**
   1. **Description**

The Requirements section, as well as the various models and diagrams utilized within it, provides a visual or conceptual representation of how the system would be used by any of the stakeholders (those being users and admins). In the provided use cases, there is an instance of both an Admin and a User utilizing the system. For an admin a typical use case for them is managing the index table, whether that be by creating a new index to be inserted into the table or by modifying an existing index within the table. Modification of an index would include changing information within an index to better reflect the index as it is currently known or to outright delete and index from the index table. For a user, a typical use case is where they search for a website.

* 1. **Model** A picture containing text, diagram, plan, technical drawing

     Description automatically generated
     1. **Uses Cases and Associated Use Case Texts** A picture containing diagram, text, circle, sketch

        Description automatically generatedA picture containing text, receipt, screenshot, parallel

        Description automatically generated
* For admins their input is whatever new information they aim to provide to the index table through the functions createIndex and modifyIndex and their associated output is thus the additional indices or modifications made to existing indices.
* For users their input is a string that is received by the system and their output is a list of websites that are related to the input they gave.
* Admins are a specialized type of user and thus also have the same functionality that is allowed to users on top of their added functionality
  + 1. **Sequence Diagrams** A picture containing text, diagram, screenshot, parallel

       Description automatically generated

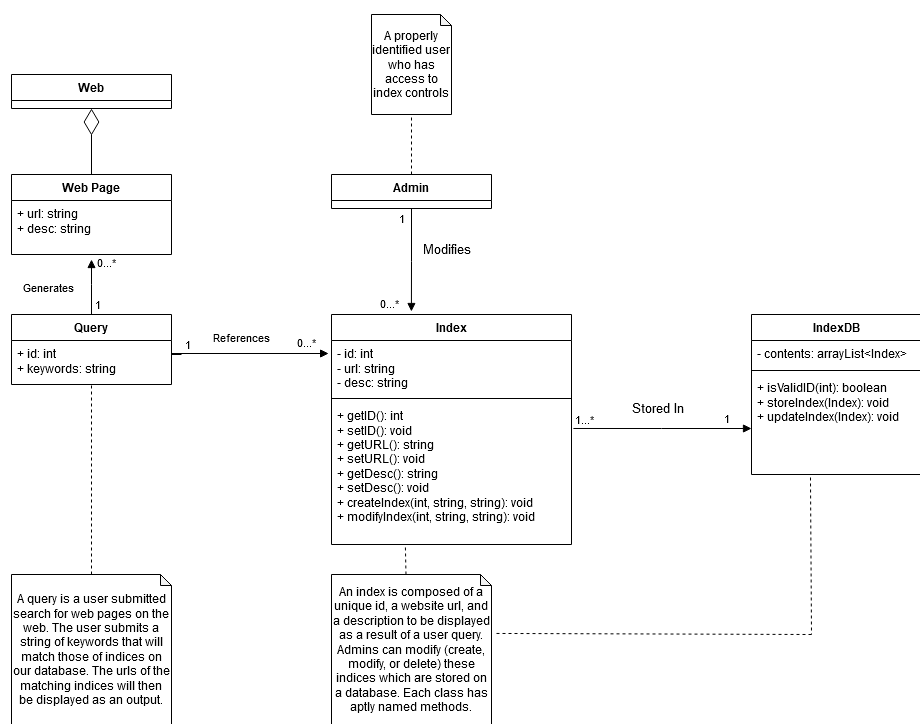
A screenshot of a computer

Description automatically generated with low confidence

* The sequence of events in the sequence diagrams are easy enough to follow in the diagrams themselves.

1. **(Architectural) Design Specifications**
   1. **Description**

The design specifications give a more robust representation of how the system is going to be implemented. In the class diagram it shows a more detailed representation of the class diagram of the system seen in the Domain Model. How CyberMiner’s queries interact with the indices table is shown, where the query will reference a specific index (or indices) from the index table (referred to as the IndexDB in the diagram). The diagram further shows the relationship between admins and the index table as well as some of the functionality they are able to utilize in order to effectively monitor and manipulate the index table as needed (through the functions createIndex and modifyIndex). Furthermore, the sequence diagrams from the Requirements section are also shown with more detail. Specifically, two different instances of where an admin is manipulating the index table through createIndex and modifyIndex.

* 1. **Models**
     1. **Class Diagram**
* The flow of events is roughly, a query is generated by a user, which references an index or indices from the index table (aka IndexDB)
* Using the indices obtained from the IndexDB, the query then generates a web page that displays a list of all sites that are relevant to the users search query
* The user is then able to connect to whatever website they choose from the list generated by clicking on them
* The flow of events for an admin is the same as a user when submitting a search query, for making modifications to the index table it is slightly different, those will be discussed more in the sequence diagrams below
  + 1. **Class models**

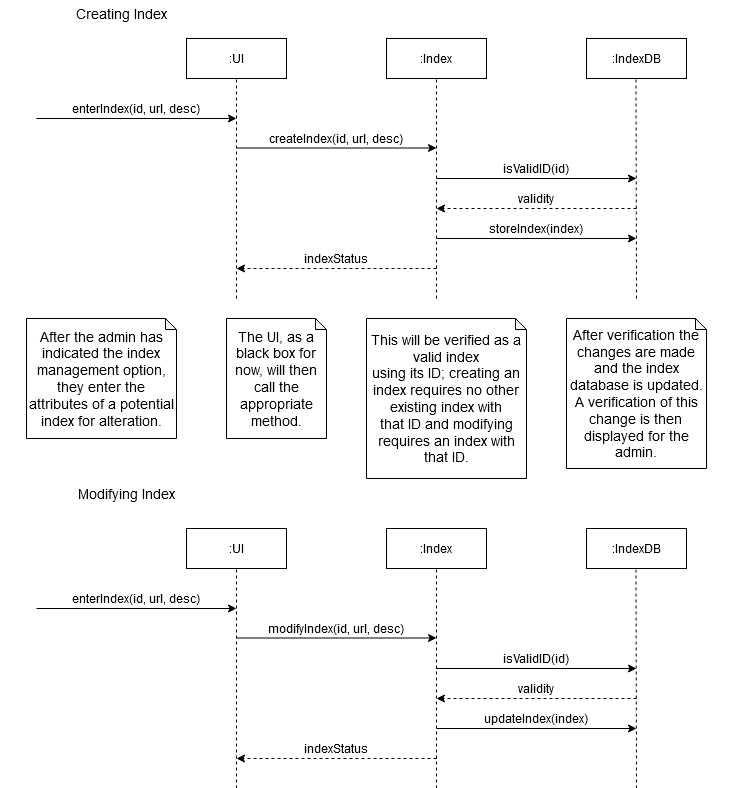
(Those that were not already described in the Domain Model section)

1. **Index**

An entry within the indices table. Contains all the information about an individual website that is stored within the table. Has the attributes “id”, “url”, and “desc”. The “id” attribute refers to a unique identifier value that is used by an admin for troubleshooting reasons primarily, “url” is just the website’s url link, and “desc” is a short description of the website and what kind of information it has or what kind of services it offers. The index class has a number of functions that facilitate the admin being able to properly maintain the system. Included are some basic getter and setter functions as well as the createIndex and modifyIndex functions which allow the admin to make changes to the IndexDB

1. **IndexDB**

The indexDB is responsible for containing the list of all indices. Its attribute “contents” is the list that was previously mentioned. The IndexDB class also has functions that facilitate the admin being able to properly maintain the system. Those being isValid() which checks the validity of an individual index within the table, the storeIndex() function appends a new entry into the table, then the updateIndex() function allows an admin to modify an existing index within the table.

* + 1. **Sequence Diagram**
* Two different instances of an admin making modifications to the indices table is shown above
* The first sequence diagram shows an instance where an admin creates a new index to be stored within indexDB. Once the admin uses the createIndex function the IndexDB checks the validity of the new index using isValid, in this case the created index is valid, so the IndexDB returns a messaging denoting that and then the new index is thus stored within the IndexDB. After which the Index class sends an IndexStatus message to the admin’s UI that confirms the successful creation.
* The second sequence diagram shows an instance where an admin modifies an existing index within the IndexDB. Once the modification has been submitted by the admin the IndexDB checks the validity of this change, in this case the modified index is still valid and is thus stored within the indexDB. After which the Index class sends an IndexStatus message to the admin’s UI that confirms the successful modification.

1. **Implementation**
   1. **Description**

The CyberMiner system is to be implemented using Java and Javafx. The indices will be stored within a database using MySQL. The code for our system has been primarily written by Fredrick Horn and Conner Carraher.

* 1. **Model – Implementation**A screenshot of a computer screen

     Description automatically generated with medium confidence
  2. **Model – Test cases**A picture containing text, diagram, font, parallel

     Description automatically generated
     1. **Test Suite LoginA screenshot of a computer program

        Description automatically generated with low confidence-** Instances of how a login test case would happen, essentially just input name, email, and password to be granted access, if fields don’t match what is on file, then deny access
     2. **Test Suite Generate QueryA picture containing text, receipt, line, font

        Description automatically generated  
        -** Instances of how generation of a query may be handled in the system, essentially the system would read the keywords string list and parse through the index database for each keyword and populate another list containing websites whose known/stored keywords match with the user’s input.
     3. **Test Suite Create/Delete IndexA picture containing text, font, line, number

        Description automatically generated-** Demonstrated in the video in section 5.4
     4. **Test Suite Modify IndexA picture containing text, receipt, screenshot, line

        Description automatically generated  
        -** Demonstrated in the video in section 5.4
  3. **Demonstration**

Below will be a demonstration of the current capabilities of the CyberMiner system. Currently the System has functionality to support manipulation of the index table. There will be a text-based representation with images as well as a video representation of these test cases in case that is easier to digest.

* + 1. **Creating a New Index**

A screenshot of a computer

Description automatically generated with medium confidence

* Navigate to the Index Table, this can be done by hitting the button in the top left corner called “Manage Indexes”

A screenshot of a computer

Description automatically generated with medium confidence

* The fields are filled in with the necessary information, “test” for title, “thisisatest.com” for url, “this is a test” for description, and “test” for keywords. Once these fields are filled it an admin can hit the “Add” button to append the entry as an index to the index table at the bottom.

A screenshot of a computer

Description automatically generated with medium confidence

* Navigating to the bottom of the list shows that the entry has been made into an index and appended to the end of the index table.
  + 1. **Modifying an Existing Index**
* Modification of indices is simple within the CyberMiner systemA screenshot of a computer

  Description automatically generated
* By double clicking on a field (not including ID, Created Date, and delete) for any of the index an admin is able to alter the information in that field. Once the change has been made the admin can hit enter to save the change for that field. A screenshot of a computer

  Description automatically generated with medium confidence
* After hitting enter the change is made, in the case above the description field for the index we created before went from “this is a test” to “this is a modification”
  + 1. **Deleting an Index**

**A screenshot of a computer

Description automatically generated with medium confidence**- Deletion of an index occurs when an admin hits a delete button in the row for the index they intend to delete.

A screenshot of a computer

Description automatically generated with medium confidence- The system after having hit the delete button for the index that was created before, the table is updated automatically and it is shown that it is not there anymore as the last entry in the index table is now 3884, where it was 3891 before.

* + 1. **Video Demonstration**

Link to video: https://youtu.be/hQtH4G20Tcg

1. **Glossary**

Search engine: A tool on the internet that helps one find information. A user can type in what they are looking for and it will show a list of websites and other sources that are related to the user’s search.

User: A person or entity that uses a service or tool in order to perform some task or tasks. In the case of CyberMiner a user is someone who uses the CyberMiner tool in order to search the web for specific websites.

Admin: An admin is a user with additional perks and permissions that allow them to be able to perform maintenance and make changes to the CyberMiner system.

Query: It is roughly a question or request for information that a user asks a search engine. This question or request is then used by the search engine to look for relevant results that match the question or request.

Delimiter: A character or sequence of characters used to separate various parts of a text or data.