CS6362 Software Architecture and Design

SPRING 2013

Project – phase 2 - final

KWIC Software Architecture for Search Engine SYSTEM

Requirement Specifications

Done by:

Abhishek Dey Das

Aishwarya Muthukumaran,

Brian Boyle

Clay Houllion

David Riviera

Gowrav Deivasigamani

Harshita Vissa

Naveen Kumar Devaraj

Pooja Dhondge

Sanjitha Srinivasan

Shiva Malavika Ganesh

Sripadmanabhan Subramanian

Sudharsan Lakshminarasimhan

Web URL: <http://sadteam.abhis.ws/>

Table of Contents

1. Introduction4

1.1 Brief Introduction4

1.2 Purpose4

1.3 Project Overview4

1.4 Team Members5

1. Requirement Specification6

2.1 Functional Requirement6

2.2 Non-Functional Requirement9

1. System Description12

3.1 Product Functionality12

3.2 Operating Environment12

**4. Traceability Matrix………………………………………………………………………………………………………………12**

1. References13
2. Appendix A: Glossary System Description13
3. **Introduction**
   1. **Brief introduction**

This document summarizes the requirements for the Cyberminer search system. Cyberminer is a search engine that utilizes the KWIC indexing system developed in phase I to develop the indexes that will be used for carrying out the search. This document only presents the requirements for Cyberminer. KWIC Indices requirements are given in the Phase I requirements specification document.

* 1. **Purpose**

The document’s purpose is to identifies the issues with the customer requirements, analyzes them and presents a refined set of requirements.

* 1. **Project overview**

The project is going to inplement the Cyberminer using the KWIC indexing system that was constructed in the previous phase.

The KWIC system shall accept an ordered set of lines, where each line consists of two parts:

* *The URL part*, whose syntax is:

URL ::= ‘http://’ identifier ‘.’ Identifier ‘.’ [‘edu’ | ‘com’ | ‘org’ | ‘net’]

identifier ::= {letter | digit}+

letter ::= [ ‘a’ | ‘b’ | … | ‘y’ | ‘z’ | ‘A’ | ‘B’ | … | ‘Y’ | ‘Z’]

digit ::= [‘1’ | ‘2’ | … | ’9’ | ‘0’]

* *The descriptor part*, whose syntax is:

identifier {‘ ‘ identifier}\*

The descriptor part of any line shall be “circularly shifted” by repeatedly removing the first word and appending it at the end of the line. The KWIC index system shall output a listing of all circular shifts of the descriptor parts of all lines in ascending alphabetical order, together with their corresponding URLs. No line in the output list shall start with any noise work such as “a”, “the”, and “of”.

The KWIC system shall allow for two modes of operation:

1. for building an initial KWIC indices
2. for growing the indices with later additions.
   1. **Team Members**

* Abhishek Dey Das
* Aishwarya Muthukumaran,
* Brian Boyle
* Clay Houllion
* David Riviera
* Gowrav Deivasigamani
* Harshita Vissa
* Naveen Kumar Devaraj
* Pooja Dhondge
* Sanjitha Srinivasan
* Shiva Malavika Ganesh
* Sripadmanabhan Subramanian
* Sudharsan Lakshminarasimhan

1. **Requirement specification**

**2.1 Functional Requirements**

**FR1: Handling the input.**

**Description**

Incomplete: It doesn’t account for how the system stores and retrieves the input.

**Alternatives**

1. The input will be converted to lower case or upper case and retrieved as such.
2. Case sensitive search – The system shall store the input as given and retrieve the input as such.

**Choice and Rationale**

A2: Case sensitive search – The system shall store the input as given and retrieve the input as such.

This refines the search results.

**FR2: Providing a way for the user to navigate to the appropriate webpage.**

**Description**

What do you mean by navigating to a page? What is an appropriate webpage**?**

**Alternatives**

1. When the user clicks on the URL, which has been retrieved as the result of a query, the system shall take the user to the corresponding web site.
2. The system shall direct the user to the correct webpage.

**Choice and Rationale:**

A1. The method used for retrieving the correct webpage is being mentioned and hence the proper navigation is mentioned.

This helps the user to see what all the results are and how many are present, thus making it easier to keep track of all the viewed and unviewed results.

**FR3: Different modes of search.**

**Description**

What does mode imply? What are the different modes?

**Alternatives**

1. The system shall allow the user to specify the mode of search, using “OR”, “AND” or “NOT”.
2. A keyword – based search which searches on any of the keywords given is used.

**Choice and Rationale**

A1: The system shall allow the user to specify the mode of search, using “OR”, “AND” or “NOT”.

This helps in better search results; this is because the mode of search provides suitable filtering.

**FR4: Running multiple searches.**

**Description**

Multiple searches are used at the same time.

**Alternative**

1. Run searches at the same time.
2. Multiple searches shall be run concurrently

**Choice and Rationale**

A2: Multiple searches shall be run concurrently.

This makes sure that the search engine can be run concurrently at the same time.

**FR5: Some inaccessible URL’s and their descriptions are deleted.**

**Description**

What does inaccessible mean?

**Alternatives**

1. The out-of-date URLs and their descriptions should be deleted.
2. The URLs that no longer exist shall be deleted.
3. The URLs for which the webpage has been removed or no longer available and the corresponding descriptions for the URLs shall be deleted from the database.

**Choice and Rationale**

A3: The URLs for which the webpage has been removed or no longer available and the corresponding descriptions for the URLs shall be deleted from the database.

It saves space. It helps provide only active links as search results and thus improves search efficiency.

**FR6: Displaying the results in a sequential order.**

**Description**

What is the sequential order?

**Alternatives**

1. The results shall be displayed the order in which it was retrieved.
2. The results of the keyword search query shall be displayed in ascending alphabetical order.

**Choice and Rationale**

A2: The results of the keyword search query shall be displayed in ascending alphabetical order. The results are displayed in the ascending alphabetical order and hence making it easier to retrieve the required data.

**FR7: Showing pages and navigating them.**

**Description**

What is the format for displaying the results in a page?

**Alternatives**

1. The number of pages to be displayed per page shall be set and navigation between the pages shall be enabled.
2. The results per page are fixed and the user can move amongst the pages.

**Choice and Rationale**

A1: The number of pages to be displayed per page shall be set and navigation between the pages shall be enabled.

The reason we chose this alternative is that the result is formatted according to the user’s convenience. Navigation makes it easy to browse through the various pages.

**FR8: Filtering meaningless symbols.**

**Description**

What do you mean by filtering?

**Alternatives**

1. The system shall remove the symbols from the input that are not meaningful.
2. The system shall omit incorrect symbols.

**Choice and rationale**

A1: The system shall remove the symbols from the input that are not meaningful.

This helps in omitting the symbols which might deter the search operations performed by the system.

**2.2 NON-FUNCTIONAL REQUIREMENTS**

**NFR1: The system shall be portable.**

**Description**

What does the term portable mean with respect to the system?

**Alternatives**

1. The system shall run on different environments.

* The system shall run on different processors.
* The system shall run on different operating systems.
* The system shall works on different versions of libraries.

1. The system shall run without errors on all environments.
2. The system shall be maintainable.

**Choice and Rationale**

A1 and A2 are chosen because these options completely define the portability of the system. A3 is not chosen because a system should also be understandable in order to be maintainable.

**NFR2: The system shall be understandable.**

**Description**

The term understandable is ambiguous and needs to be explained clearly.

**Alternatives**

1. The system shall have a low learning curve.

* The systems shall provide users with technical support.
* The user manual shall be in a simple, understandable language.

1. The system shall be easy to use.

**Choice and Rationale**

A1 is chosen. A2 is not chosen because this requirement is relevant only for usability.

**NFR3: The system shall support enhanceability.**

**Description**

The term enhanceability should be clearly defined.

**Alternatives**

1. The system shall support changes that improve existing logic.
2. The system shall support changes that improve existing functionality.
3. The system shall support changes to the system.

**Choice and Rationale**

A1 and A2 are chosen. A3 is not chosen because the changes described are ambiguous.

**NFR4: The system shall be user-friendly.**

**Description**

The term user-friendly means different things to different users. Hence, it needs to be described in detail.

**Alternatives**

1. The system shall have a well developed user interface.
2. The system shall make the web pages easily navigable.
3. The system shall make searching simple with a few steps.

**Choice and Rationale**

A2 and A3 are considered. A1 is not considered because the requirement doesn’t describe how the interface should be well developed.

**NFR5: The system shall be reusable.**

**Description**

The term reusable is very ambiguous and therefore, has to be explained clearly.

**Alternatives**

1. The system shall facilitate the reuse of components.
2. The system shall facilitate the reuse of code.
3. The system shall facilitate the reuse of design.
4. The system shall facilitate the reuse of requirements for other systems.

**Choice and Rationale**

A1, A2, A3 are chosen. A4 is not chosen because requirements may change from one system to another.

**NFR6: The system shall exhibit good performance.**

**Alternatives**

1. The system shall produce the results in optimal time (less than 1 minute).
2. The system shall display correct and relevant results.
3. The system shall not hang in the middle of a search query.

**Choice and Rationale**

Alternatives A1 and A2 are considered. A3 is not considered because a system on which the software is running might have a large number of applications running in the background or might be under a virus attack and therefore, it might hang.

**NFR7: The system shall ne modifiable.**

**Description**

What is modifiability? How is it different from enhaceability?

**Alternatives**

1. Changes in the requirements which require change in the system shall be easily accommodated.
2. The system shall make changes to the existing software to facilitate changes in requirements and not implement additional functionality.

**Choice and Rationale**

A2 is considered since it describes modifiability completely and provides a clear distinction between modifiability and enhanceability.

1. **System Description**

**3.1 Product Functionality**

Cyberminer shall allow for

* *Case sensitive search*: The system shall store the input as given and retrieve the input also as such;
* *Hyperlink enforcement*: When the user clicks on the URL, which has been retrieved as the result of a query, the system shall take the user to the corresponding web site;
* *Specifying OR/AND/NOT Search*: A keyword-based search is usually an OR search, i.e., a search on any of the keywords given. The system shall allow the user to specify the mode of search, using “OR”, “AND” or “NOT”;
* *Multiple search engines*: to run concurrently;
* *Deletion of out*-of-date URL: and corresponding description from the database;
* *Listing of the query result in ascending alphabetical order*;
* *Setting the number of results to show per page, and navigation between pages*; and also possibly
* *Autofill*, while correcting typographical errors,
* Filtering out symbols that are not meaningful.

**3.2 Operating Environment**

The software will work in a web browser environment and it will be accessible from any type of operating systems such as Windows, Linux, UNIX, etc. There are no particular hardware or software requirements.

1. **Traceability Matrix:**

|  |  |
| --- | --- |
| **Functional Requirements** | **Non-Functional Requirements** |
| FR1 | NFR7 - Modifiability, NFR8- Adaptability |
| FR2 | NFR2 – Understandability, NFR4- User Friendly, NFR6 – Performance |
| FR3 | NFR7 – Modifiability, NFR8- Adaptability |
| FR4 | NFR6 – Performance, NFR8- Adaptability |
| FR5 | NFR3 – Enhanceability, NFR6 – Performance, NFR7 - Modifiability |
| FR6 | NFR2 – Understandability, NFR4- User Friendly |
| FR7 | NFR2 – Understandability, NFR4- User Friendly, NFR6 – Performance |
| FR8 | NFR2 – Understandability, NFR4- User Friendly |
| FR9 | NFR2 – Understandability, NFR3 – Enhanceability, NFR7 - Modifiability |

1. **References**

* Lecture Notes
* [www.google.com](http://www.google.com/)
* <http://utdallas.edu/~chung/SA/Project2.pdf>

1. **Appendix A: Glossary**

· KWIC: Key Word In Context