CS 6359: Object-Oriented Analysis and Design

Deliverables

INFINITY ENGINE INTERIM PROJECT II REPORT

Team Details

Name	Net ID
Shreyans Patel	ssp210009
Rathang Rajpal	rxr210009
Shatavari Shinde	svs210001
Arush Sharma	axs200241
Yash Gupta	yxg210002
Prasad Shinde	pxs200111
Pin-Hsuan Yao	pxy200000
Chia Chun Chu	cxc200048
Nayeon Kim	nxk210015
Khushbu Sharma	krs200001
Vanita Bhagwat	vxb210009
Manu Rajeev Karri	mxk200102
Chi Lee	cxl200014

Table of Contents

List of Figures

Introduction

Infinity Engine is a user-friendly search engine from scratch. The search engine will focus on a specific domain of information to provide results confined to the knowledge space of that domain.

The engine will serve through four primary functional units:

1. Crawling/Spider - That will systematically browse the WWW to collect data for indexing.
2. Indexing - refers to collecting, parsing, and storing data for fast retrieval.
3. Searching - matching the user query to provide accurate results.
4. Fronted end - presenting the retrieved data in a convenient and accessible manner.

The domain our search engine focuses on is “Movies”. Our aim is to build a powerful and efficient engine to make custom searches utilizing the attributes of the movie entity and design a novel approach to perform this task.

Example: Lookup movies having “Tom Cruise” and “Brad Pitt”.

Search Engine Link: https://ooad-cyberminer.herokuapp.com/home/
Team Website: https://personal.utdallas.edu/~arush.sharma/

Functional Requirements

FR01: Case Independent Search

This means that a search for a term matches other occurrences of that term, regardless of their case in the original document, and regardless of the case in which the search term is written. For example, a search for Iron Man matches documents that originally contained terms such as IRON MAN, iron man, etc.

FR02: Hyperlink Enforcement

This means that the results returned by the search will have hyperlinks to the actual location of the search result on the Internet.

FR03: Logical Operators Support

Logical operators form the basis of mathematical sets and database logic can be used in the search query. They connect your search words together to either narrow or broaden your set of results. The three basic Boolean operators are: AND, OR, and NOT. For example, a search query containing Angelina Jolie ‘and’ Tom Cruise will return the list of movies they both have acted in or the documents containing both their names.

FR04: Concurrent Operation

Two or more operations performed at the same time (or within a given interval) without failure. The system should function properly in case multiple users perform a search.

FR05: Outdated URL Deletion

Outdated URL deletion is a way in which search engines can provide the most relevant and updated information to its users. To find outdated pages, we can use measures like the website's traffic, or if a URL still exists or not.

FR06: Ranking / Ordering Results

Ranking is the process during which search results are sorted by relevance to the query from the most to the least pertinent. We would be using techniques like TF-IDF to rank the pages. TF refers to term frequency or the number of times the searched word appears in the page and IDF refers to inverse document frequency that checks how important the word that is searched is in our collection. Based on this combination of TF-IDF a score is generated for each page and the pages with the highest score appear on top.

FR07: Number of Results per Page Functionality

The search results for the keyword should be displayed with pagination i.e., the number of results should be limited to a fixed number per page which depends on the size of the UI of the device that the user accesses the application.

FR08: Autofill Support

The search functionality should provide suggestions of the search text to the user based on the input he provides in the search bar. The Autofill should happen in real-time; the suggestions should be provided as soon as the user starts typing.

FR09: Filter Stop Words

The application should filter out the stop words from the user input. The stop words are the most common words in any language (articles, prepositions, pronouns, conjunctions, etc.) which does not add much information to the text. Ex: a, an, the, what, there etc.

Non- Functional Requirements

NFR01: User-Friendly

We are a customer-driven team, and we want to aim at maximum user convenience. We target to design our Infinity Search Engine in a user-friendly interface so that the engine can provide a perfect user experience.

NFR02: Low Latency

We want our system available to be responsive in low latency, this can bring the best experience for our users.

NFR03: Scalable

Our system needs to be scalable and stable when huge amounts of users searching or are on the website at the same time. This maintains the quality of our system and it won't be crushed when a lot of users are using it.

NFR04: High Throughput

We want to execute many tasks and complete them correctly. We are concerned about the robustness and reliability of jobs over a long-time scale. We want to create a reliable system in this project.

NFR05: Understandable to Future Developers

It is important to write readable code for future developers to help maintain or modify the system.

Planned Development Cycle

The current version of the report represents the 1 out of 5 planned iterations as depicted by below mentioned traceability diagram. We plan to achieve the said FRs within each of the projected iterations and we are currently reviewing our plan of action for the FRs to maybe merge a few of them because they are dependent on each other.

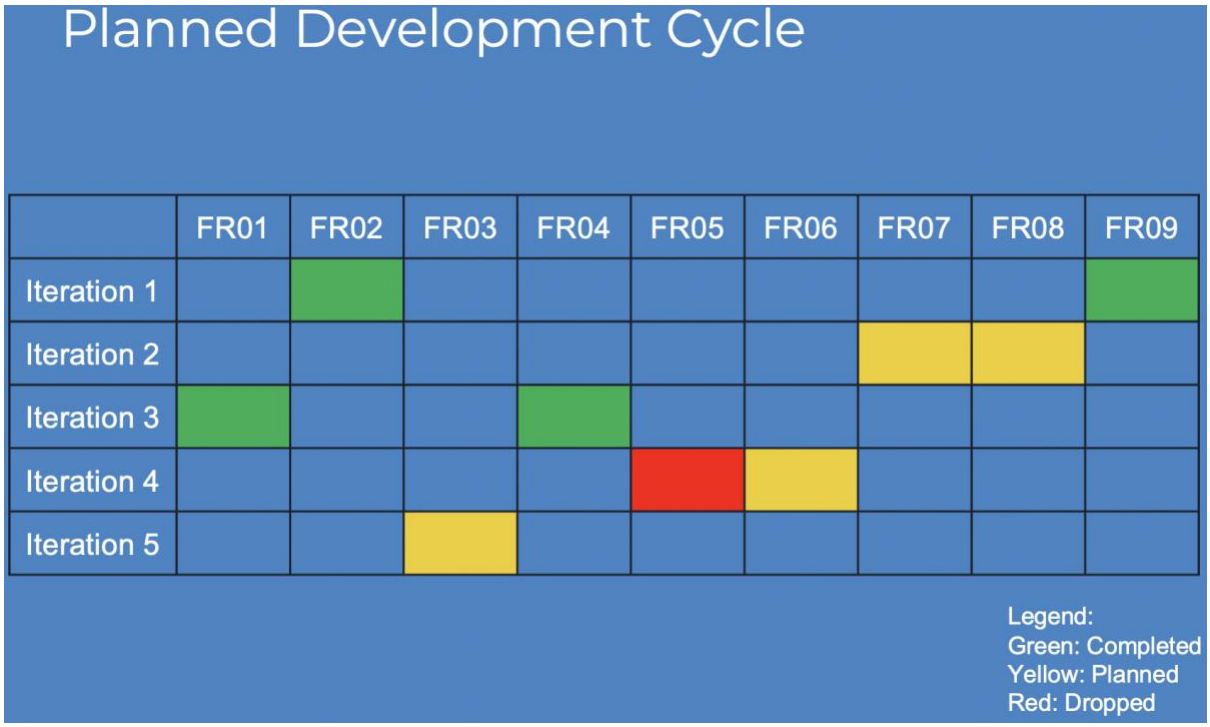

Figure 1: Traceability Diagram

As of now, FR01, FR02, FR04 and FR09 have been completed and the team has worked halfway through on few other FRs.

FR05 has been dropped from the scope of the two-phase project since it can be a feature in pipeline for the phase three of the project which will be implemented post-completion of the course.

Deliverable

- Interim Project I - Initial conceptualization of ideas/designs that can lead to possible solutions along with a ppt file.
- Final Project I - Reporting intermediate deliverables of codebase, designs, and diagrams.
- Interim Project II - Delivering most part of the code base with final designs and implementing improvement suggestions received at the end of Project I along with a ppt.
- Final Project II - Submission of final report with complete code base.

Project Plan

- February 2 (Wednesday) - Preliminary Project Plan
- March 7 (Monday) + March 9 (Wednesday) - Interim Project I

- March 23 (Wednesday) - Final Project I Submission
- April 18 (Monday) - Interim Project II
- April 25 - April 27 (Monday/Wednesday) - Final Project II Submission

Team's Role Play

Frontend Team

This team is responsible for creating a feature-rich and easy-to-use User Interface (UI). This team creates what you will see and interact with while using our search engine. The members of this team are:

- Chia Chun
- Prasad Shinde
- Vanita Bhagwat
- Yao Pin-Hsuan

Backend Team

This team creates is responsible for server-side web application logic and integration of the frontend with the database through custom built APIs. The members of this team are:

- Khushbu Sharma
- Rathang Rajpal
- Shatavari Shinde
- Shreyans Patel

System Design Team

System design team is responsible for collecting data requirements, deriving functional and non-functional requirements, preparing diagrams, and making a detailed design of the whole system. The members of this team are:

- Arush Sharma
- Chi Lee
- Manu Rajeev Kari
- Nayeon Kim
- Prasad Shinde
- Yash Gupta

New Updates

In the second phase of the project, we have completed FR01 and FR04 while work on remaining FRs is under progress. Further discussion on the FRs have led to a conclusion that some FRs might have to be dropped considering the time available before delivery.

Phase II Updates

Domain Model

A domain model is incorporated into the final report as per the suggestions given post first presentation.

Team Formation

The members were internally rotated within the team and the Scraping Team has been transformed to the Frontend Team with a new set of responsibilities.

Functional Requirements

Some new FRs have been completed and one of the FRs has been dropped to ensure team efficiency and timely completion of the project. The dropped FR has been planned for implementation in the third phase of the project which will be post-completion of the course.

Planned Development Cycle

The planned development cycle consisting of the traceability chart has been modified with the latest progress.

Use Case Diagrams

The concept of generalization was incorporated by adding a new physical user. More number of users are added according to the feedback received after the first presentation. More use cases added.

Class Diagrams

Class diagram for the domain model added increasing the total number of class diagrams in the project.

Sequence Diagrams

More sequence diagrams are added for more in-depth insights of the system.

Activity Diagrams

The activity diagram of the search engine has been added which was a new requirement for the second phase.

State Transition Diagrams

The state transition diagram of the search engine has been added which was a new requirement for the second phase.

Other Diagrams

A few more diagrams have been added beyond the scope of the project to better explain the system functionality.

Tools

Added information about the tools which was missing from the presentation of first phase.

PPT Modification

The PPT used for previous presentation has been modified with the new additions and refactoring of previous content.