UNIVERSITI TEKNOLOGI MARA

Universiti Teknologi MARA

Search Engine for Books

Amierul Izzuddin Bin Azman

Thesis submitted in fulfillment of the requirements for

Bachelor of Computer Science (HONS)

Faculty of Computer Science and Mathematical

February 2022

DECLARATION

I certify that this report and the research to which it refers are the product of my own work and that any ideas or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the discipline.

……

AMIERUL IZZUDDIN BIN AZMAN

2020968971

(29 January 2022)

ABSTRACT

Search engine is web-based software that searches for and identifies things in a collection that match the user's keywords or characters, and is mostly used to locate specific websites on the Internet. The basic idea of information retrieval (IR), as is well-known in the computer industry, is to search a given amount of data and retrieve those records that fulfil a set of criteria. For this project, it will be more on developing a search system that focus on the domain of books. Although many search engines available for book searching, most of them still have room for improvement. The reality to find books using search engine are quite challenging because, the users need to know the title of the book in order to retrieved relevant result from the search engine. Therefore, this project aims to propose a search engine that might produce better result by manipulating the indexing structure of the search engine. Software Development Life Cycle also known as SDLC was used as the methodology in this project development. This project applies the vector space model for the matching process, because the chosen software library for this project which is Apache Lucene is using the vector space model as its foundation. In addition, the Bag-of-Words approach was used as the basis for the indexing module. The indexing process indexed the information of the books such as the book title, the author name, publisher, year published, pages count and synopsis of the book. The search engine's assessment criteria include recall and precision. In IR, recall and precision have long been used as standard evaluation criteria. Keenly, the results of this project prove that index files that contain more domain information can increase the relevancy of a search engine.

TABLE OF CONTENTS