Universiti Teknologi MARA

Query Translation Technique in Cross-Lingual Information Retrieval Approach for Intellectual Property Documents.

Muhamad Safwan bin Dermawan

Thesis submitted in fulfillment of the requirements for
Bachelor of Computer Science (Hons.)
Faculty of Computer and Mathematical Sciences

July 2013

ACKNOWLEDGEMENT

ABSTRACT

Nowadays, information in documents can be obtained faster than before. People around the world can obtain information across countries and languages. Thus, search engines such as Google and Bing exist to provide search tools to find documents across the world. These are the examples of information retrieval (IR). IR is the act of finding information in a collection of documents. Cross-lingual information retrieval (CLIR) is much the same as general IR but it involves more than one languages. To find information in a large collection of documents is harder than it looks especially when it involves two or more languages. Translating a collection of documents from one language to another can simplify the process of searching itself but it requires a lot of efforts and time. Query translation is a good technique which can reduce the time used for this project. Query translation is a process of translating user query's language to the documents' language and target output. In this project, search engine prototype that will be produced by using query translation technique will produces relevant list of relevant documents according to user query. In conclusion, this project for searching IP documents by using query translation techniques for CLIR approach is to improve and enhance searching and retrieving of IP documents which involves two languages for students and lecturers in the field of intellectual property and other interested parties.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1   Introduction

This chapter provides the overview and rational for the project. It also discusses about background study, problem statement, project objectives, project scope, project significance, outline of thesis for this project and summary of this chapter.

1.2   Background Study

Information is very important to almost everyone nowadays. With the help of technology such as smartphone especially, people can obtain information almost as fast as the speed of light. Thus, people ought to search for information through the internet. Internet connects information around the world with the people. Furthermore, information in the forms of research papers, journals and others are usually been searched by people from academic area and other interested parties. According to Nasharuddin and Abdullah (2010) in an electronic journal entitled "Cross-lingual Information Retrieval: State-of-the-art," Information Retrieval (IR) is defined as "the act of finding materials, usually in documents of an unstructured form that satisfies an information need within large collections stored in computers."

Google for example is a good search engine which provides tons of information, whether in the form of documents, presentation slides, images, videos and others. A person just needs to type the keywords into the text box and press enter. Then