COMPARISON OF RETRIEVAL SCHEME BASED ON DIFFERENT STRUCTURE AND SIMILARITY MEASURE FOR MEMOS AT TATi UNIVERSITY COLLEGE – CENTRE OF HND STUDIES

BY

AINI ZURIYATI BT ABDUL KADIR

Project Paper Submitted in Partial Fulfillment of the Requirement for the Degree of Master of Information Technology

OPEN UNIVERSITY MALAYSIA

ABSTRACT

Retrieval models and techniques can be applied to retrieve memo information and it does relate to certain queries or concepts. Different method of retrieval shows different relevant documents. Therefore, different methods are used to retrieve, structure and similarity measures being used to retrieve memo. The data are collected at Center of HND Studies at TATi University College. The process that has been applied is digitizing, stop word removal, stemming and building index. The results of this process then stored in database. In this study 50 document of memo being used, where refer to the titles and the document. In order to show the measure of different memo structure three similarity measures being used. The structure refers to the titles of memo, the documents and combination of both title and documents. The results shows that title structure have the lowest performance of result compared to document structure and the combination of both. Meanwhile, in similarity measures the results shows that weighting schemes using cosine perform well compared to Dice and Russell-Rao. The overall result shows that the best performance for retrieving memo is by using document structure coupled with Cosine similarity.

TABLE OF CONTENTS

Chapter 3

METHODOLOGY

Chapter 4

DATA ANALYSIS AND RESULT

Chapter 5

DISCUSSION

Chapter 6

SUMMARY AND CONCLUSION

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Introduction

Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers). As defined in this way, information retrieval used to be an activity that only a few people engaged in: reference librarians, paralegals, and similar professional searchers. Now the world has changed, and hundreds of millions of people engage in information retrieval every day when they use a web search engine or search their email.

The field of information retrieval also covers operations typically done in browsing document collections or further processing a set of retrieved documents. Given a set of documents, clustering is the task of coming up with a good grouping of the documents based on their contents. It is similar to arranging books on a bookshelf according to their topic. Given a set of topics (or standing information needs), classification is the task of deciding which topic(s), if any; each of a set of documents belongs to. It is often approached by first manually