AN AGENT SYSTEM FOR INFORMATION RETRIEVAL IN AN ACADEMIC ENVIRONMENT

KHALIFA CHEKIMA

THESIS SUBMITTED IN PARTIAL FULFILMENT FOR THE DEGREE OF MASTER OF SCIENCE

UMS
UNIVERSITI MALAYSIA SABAH

SCHOOL OF ENGINEERING & INFORMATION TECHNOLOGY

UNIVERSITI MALAYSIA SABAH
2012

AN AGENT SYSTEM FOR INFORMATION RETRIEVAL IN AN ACADEMIC ENVIRONMENT

KHALIFA CHEKIMA

THESIS SUBMITTED IN PARTIAL FULFILMENT FOR THE DEGREE OF MASTER OF SCIENCE

SCHOOL OF ENGINEERING & INFORMATION TECHNOLOGY

UNIVERSITI MALAYSIA SABAH

2012

UMS
UNIVERSITI MALAYSIA SABAH

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN TESIS

JUDUL : _____
_____
_____

IJAZAH : _____
_____

SAYA : _____ **SESI PENGAJIAN** : _____
(HURUF BESAR)

Mengaku membenarkan tesis *(LPSM/Sarjana/Doktor Falsafah) ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:-

1. Tesis adalah hak milik Universiti Malaysia Sabah.
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (/)

☐ SULIT (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di AKTA RAHSIA RASMI 1972)

☐ TERHAD (Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

☐ TIDAK TERHAD

Disahkan oleh:

_____ _____
(TANDATANGAN PENULIS) (TANDATANGAN PUSTAKAWAN)

Alamat Tetap: _____
_____
_____
_____

_____
(NAMA PENYELIA)

TARIKH: _____ TARIKH: _____

Catatan:
*Potong yang tidak berkenaan.
*Jika tesis ini SULIT dan TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh tesis ini perlu dikelaskan sebagai SULIT dan TERHAD.
*Tesis dimaksudkan sebagai tesis bagi Ijazah Doktor Falsafah dan Sarjana Secara Penyelidikan atau disertai bagi pengajian secara kerja kursus dan Laporan Projek Sarjana Muda (LPSM).

UMS UNIVERSITI MALAYSIA SABAH

CERTIFICATION

NAME	**: KHALIFA CHEKIMA**
MATRIX NUMBER	**: PK2008-8352**
TITLE	: **AN SYSTEM FOR INFORMATION RETRIEVAL IN AN ACADEMIC ENVIRONMENT**
DEGREE	**: MASTER OF SCIENCE (COMPUTER SCIENCE)**
VIVA DATE	**:**

DECLARED BY

1. SUPERVISOR

DR.Chin Kim On Signature

_____

2. CO-SUPERVISOR

A.P. DR. Patricia Anthony Signature

_____

UMS
UNIVERSITI MALAYSIA SABAH

DECLARATION

I hereby declare that the material in this thesis in my own except for quotations, excerpts, equations, summaries and references, which have been duly acknowledged.

15 August 2012

_____

Khalifa Chekima
PK2008-8352

UMS
UNIVERSITI MALAYSIA SABAH

ACKNOWLEDGEMENTS

First and foremost I would like to thank Allah the Almighty for his guidance and blessing during my master study.

I would like to express my profound appreciation to my family for their supports during the entire period of my study.

I would like to express my sincere gratitude and thanks to the Dean of School of Engineering and Information Technology, A.P. Dr. Rosalam Sarbatly for providing support during my research work.

I would like to take my opportunity to thank my supervisor, Dr.Chin Kim On for his constant encouragement and useful suggestions.

I would like to acknowledge the help and guidance that I received from my co-supervisor, A.P. Dr. Patricia Anthony.

I would like to show my appreciation to the Ministry of Science, Technology and Innovation (MOSTI), Malaysia for supporting this Master's research.

Lastly, I would like to thank everyone who helped me from the beginning until the final phase of my Master study.

UMS
UNIVERSITI MALAYSIA SABAH

ABSTRACT

MULTI AGENT SYSTEM FOR INFORMATION FOR INFORMATION RETRIEVAL AND ACQUISITION IN ACADEMIC ENVIRONMENT

In the recent years, academic domains just like other domains have undergone a tremendous growth on both content and users. This growth has lead to information overload in which academician are finding it difficult to locate the right academic paper at the right time. Search Engines, were designed originally to be helpful in searching for relevent resluts and returning them to users. Yet, due to thousands of petentially relevant sites, thus search engines are losing their usefulness. To address this problem, a development of a reliable multi-agent system that is able to guide academician through the big ocean of information by filtering the information and recommending them relevant papers is vital. However, recommending an item/academic paper is not easy since it depends on many factors such as the user's current interest, as user's interest changes over time, size of content, and number of users. This thesis presents the development of multi-agent system that helps academicians in the process of retrieving relevant academic papers by recommending them papers based on their current interest. The recommendation is generated using a Hybrid recommendation approach, which is a combination of the two well known recommendation approach, content-based filtering approach and collaborative filtering approach. The system consists of four agents working together. The first agent is Monitoring Agent that monitors User's browsing behavior to implicitly observe users' current interest. The second module is the Categorizer Agent that automatically organizes papers downloaded by users into subcategories based on ACM Association Computing Machinery CCS (Computing Classification System) structure by considering papers' content similarity. The third agent is the Recommender Agent that recommends papers to users based on Hybrid approach, and the last agent is the Search Agent that allows users to search for academic papers locally. The use of multi-agent technology has overcome many problems that a traditional recommendation system suffers from. The accuracy of the Hybrid approach used by the proposed system in the recommendation is then compared with two other common recommendation approaches, content-based filtering approach and collaborative filtering approach by counting the precision value of each approach. Based on the results, the system was able to recommend well based on user's current interest using Hybrid approach. Besides that, the categorizer agent has shown promising results in categorizing of academic papers based on the proposed ACM CCS system.

ABSTRAK

Dalam era masa kini, domain akademik seperti juga domain-domain yang lain telah melalui peningkatan yang tinggi yang berkait rapat dengan pengguna-pengguna dan kandungan. Peningkatan ini telah mendatangkan fenomena perlambakan informasi di mana para akademik mengalami kesukaran untuk mencari kertas akademik yang sesuai pada masa-masa tertentu. Seach Engine, telah direka khas untuk membantu dalam pencarian maklumat yang berkaitan, dan mengembalikan maklumat yang didapati kepada pengguna. Walaubagaimanapun, kerana terdapat beribu-ribu laman web yang relevan, maka Search Engine ini telah kehilangan kegunaan mereka. Untuk menangani masalah ini, pembangunan sistem cadangan agen adalah amat penting dimana ia membantu para akademik. Walaubagaimanapun, cadangan sesuatu kerja atau kertas akademik bukanlah semudah yang disangka memandangkan ianya bergantung kepada beberapa faktor seperti minat semasa pengguna yang berubah dari masa ke semasa, saiz bahan-bahan dan bilangan pengguna. Tesis ini memaparkan pembangunan yang lengkap mengenai kertas penyelidikan system cadangan agen yang mencadangkan kertas akademik kepada para pangguna/akademikus berdasarkan minat semasa mereka. Cadangan Sesutu kerja adalah berdasarkan pendekatan cadangan hibrid iaitu kombinasi daripada dua cadangan pendekatan yang terkenal, pendekatan kandungan berasaskan penapisan dan pendekatan kolaboratif penapisan. Sistem ini mengandungi empat agen yang bekerjasama. Agen yang pertama ialah agen pemantauan yang mana ianya memantau perilakuan pengguna secara implicit dalam mengamati minat pengguna. Modul yang kedua ialah agen kategori yang secara automatik mengkategorikan kertas akademik yang telah dimuat turun oleh pengguna kepada subkategori berdasarkan ACM Association Computing Machinery CCS (Sistem Klasifikasi Komputer)struktur mengingati persamaan isi kandungan kertas penyelidikan. Agen yang ketiga adalah Agen pencadang yang mencadangkan kertas akademik kepada pengguna berdasarkan pendekatan hibrid dan agen yang terakhir adalah agen pencarian yang membolehkan pengguna untuk mencari kertas akademik tempatan. Dengan menggunaan teknologi yang terdapat pada multi-agent, banyak masalah yang dihadapi oleh system tradisi telah diatasi. Ketepatan pendekatan hibrid yang digunakan oleh sistem dalam cadangan ini telah dibandingkan dengan dua daripada cadangan pendekatan yang sedia ada iaitu pendekatan kandungan berasaskan penapisan dan pendekatan kolaboratif penapisan dengan mengira ketepatan setiap pendekatan. Berdasarkan keputusan hasil daripada pembandinagn tiga pendekatan pencadangan, Hibrid didapati lebih berkesan berbanding dua pendekatan yang lain. Selain daripada itu, agen kategori telah memaparkan hasil yang baik dalam mengkategorikan kertas akademik.

CONTENTS

		Page
TITLE		i
CERTIFICATION		ii
DECLARATION		iii
ACKNOWLEDGMENTS		iv
ABSTRACT		v
ABSTRAK		vi
CONTENTS		vii
LIST OF TABLES		ix
LIST OF FIGURES		x
LIST OF APPENDIX		xii
CHAPTER 1: INTRODUCTION		
1.1	Overview	1
1.2	Proposed Agent System Overview	5
	1.2.1 The agent System	5
	1.2.2 Recommendation Approaches	6
1.3	Overview of an Agent	6
	1.3.1 Introduction to an Agent	6
	1.3.2 Characteristics of an Agents	7
1.4	Research Objective	8
1.5	Organization of the Thesis	8
CHAPTER 2: LITERATURE REVIEW		
2.1	Introduction	10
2.2	Problem Statement	10
2.3	Existing Systems and Techniques	11
2.4	Overview of an Agent Technology	15
2.5	Overview of Monitoring Agent	19
2.6	Overview of Categorizer Agent	22
2.7	Overview of Recommender Agent	27
2.8	Overview of Search Agent	32

UMS
UNIVERSITI MALAYSIA SABAH

CHAPTER 3: METHODOLOGY

3.1 Introduction 35
3.2 Proposed Agent System 35
3.3 The Agent System 39
3.3.1 The Monitoring Agent 39
3.3.2 The Categorizer Agent 46
3.3.3 The Recommender agent 53

a) Recommender Agent Techniques 53
I. User Profile Generation Technique 53
II. Recommendation Techniques 55
b) Recommendation Process 62

3.3.4 The Search Agent 65
a) Short Overview of Lucene 65
b) Searching Process 66

3.4 Chapter Summary 67

CHAPTER 4: EVALUATION

4.1 Overview 68
4.2 Categorizer Agent's Evaluation 68
4.2.1 Experimental Setup 68
4.2.2 Evaluation Results 69
4.3 The Recommender System's Evaluation 74
4.3.1 Experimental Setup 74
4.3.2 Evaluation Results 75
4.4 The Agent System performance evaluation 80
4.4.1 Evaluation Setup 80
4.4.2 Evaluation Result 81
4.5 Summary 86

CHAPTER 5: CONCLUSION

5.1 Overview 87
5.2 Future Work 88
5.2.1 Enhancing Categorizer Agent 88
5.2.2 Wider Recommendation Topics 89
5.3 Conclusion 89

REFERENCES 90

APPENDIX A 96

APPENDIX B 97

UNIVERSITI MALAYSIA SABAH

LIST OF TABLES

		Page
Table 2.1	Characteristics of Intelligent Agent	16
Table 2.2	Features, Advantages and Benefits of Agent Technology	17
Table 3.1	ACM Categories list	47
Table 3.2	Users Accessed Papers	57
Table 3.3	List of Subspaces Found in Step-1	58

UMS
UNIVERSITI MALAYSIA SABAH

LIST OF FIGURES

		Page
Figure 2.1	WebMate System Architecture	21
Figure 3.1	Proposed Agent System Architecture	36
Figure 3.2	User Interface (Search Interface)	37
Figure 3.3	PDF Document Monitoring Process	42
Figure 3.4	'userpapers' Database Table	44
Figure 3.5	'doc' Table in Database	45
Figure 3.6	'doccat' Table in Database	46
Figure 3.7	Training Corpus Folders	49
Figure 3.8	Document Classification Example	50
Figure 3.9	PDF Document Categorizing Process	52
Figure 3.10	'userpapers' Table	54
Figure 3.11	'userpaper' Table According to Id Number	55
Figure 3.12	Steps involved in Like-Minded Algorithm	57
Figure 3.13	The subspaces are grouped Together in One Cluster if There is Overlap Between Subspaces	59
Figure 3.14	Using The Subspace Clustering Model to Generate Recommendations	60
Figure 3.15	Recommendation Weight Allocation	61
Figure 3.16	Document Recommendation Process	64
Figure 3.17	The Searching Process	66
Figure 4.1	Results of Categorization Based on ACM CCS Level I and Level II	69
Figure 4.2	Results of Categorization Based on ACM CCS Level I on 500 Computer Science Papers and 500 Non-computer Science Papers Without Threshold.	71
Figure 4.3	Experiment to Identify Reasonable Threshold	73

UMS
UNIVERSITI MALAYSIA SABAH

Figure 4.4	Results of Categorization Based on ACM CCS Level I on 500 Computer Science Papers and 500 Non-computer Science Papers with Threshold	74
Figure 4.5	Comparison of Precision Values Between Various Recommendation Approaches	76
Figure 4.6	Recommendation Snterface Snapshot (User search for "Bidding" Keyword)	77
Figure 4.7	Recommendation Interface Snapshot (User interested in "Artificial Intelligence" Keywords)	78
Figure 4.8	Zero Recommendation Interface Snapshot (User Interested in "Cooking Recipe" Keywords	79
Figure 4.9	Students' Evaluation Results of The Proposed Agent System	81
Figure 4.10	Search Interface	83
Figure 4.11	Paper Recommendation List Interface	84
Figure 4.12	Recommended Paper Details Interface	85

UMS
UNIVERSITI MALAYSIA SABAH

LIST OF APPENDIX

		Page
Appendix A	Accepted Publication	96
Appendix B	Questionnaires	97

UMS
UNIVERSITI MALAYSIA SABAH

CHAPTER 1

INTRODUCTION

1.1 Overview

With the rapid development of World Wide Web (WWW), huge amount of information is now accessible to the web users. This phenomenon has attracted academic users to publish their research papers online, download and share academic papers among them through WWW. Over a period of time, more academic papers have been uploaded by academicians. This phenomenon has led to information explosion in academic domains, whereby academicians are now finding it difficult to get the right and relevant academic papers without spending a considerable amount of time online.

The explosion of information on the Internet can be seen not only on academic domains, but more widely covering almost all types of domains. Montaner*et al.*, (2003), stated that the development of the Internet has resulted in a global information society with a growing number of users around the world. Due to this avalanche of information at our doors, it is more difficult to find what we want, when we need it and in a manner which best meets our requirements. Users are constantly confronted with situations in which they have too many options to choose from, where they need help to explore and to filter out their preferences from the myriad of possibilities. Internet Search Engines, designed originally to be helpful, now commonly return many thousands of potentially relevant sites, thus losing their usefulness (Montaner*et al.*,2003).

According to Mortensen, (2007), the tremendous growth of both information and usage has led to a so-called information overload problem in which users are finding it increasingly difficult to locate the right information at the right time. As a response to this problem, much research has been done with the goal of providing users with more proactive and personalized information services.

An agent systems which are a form or work from specific type of information filtering system technique that attempts to recommend information items (movies, TV program/show/episode, video on demand, music, books, news, books, images, web pages, scientific literature such as research papers etc.) or social elements (e.g. people, events or group) that are likely to be of interest to the user, have proved to help achieve this goal by using the opinion of a community of users to help individuals in the community select more effectively the content of interest from a potentially overwhelming set of choices (Ricci *et al., 2011)*

Two recommendation strategies that are dominant are content-based and collaborative filtering. Content-based filtering relies on rich content descriptions of the items that are being recommended, while collaborative filtering recommendations are motivated by the observation that we often look to our friends for recommendations (O'Donovan and Smyth, 2005).

Systems using agent recommendations have been developed in various research projects. As an example, recommender system has been deployed is in commercial domains. A well-known example is Amazon.com, where a recommender system is used to help people find items they would like to purchase. Many online communities within the movie domain use recommender systems to gather user opinions on movies, and then produce recommendations based on these opinions. Examples are MovieFinder2and Movielens3. New popular music services like Pandora4 and Last.fm5 also make use of recommendations to recommend personalized music players.

In the academic domain, web users surf the Internet in order to search for some information or data that are of interest to them. However searching for the relevant information on the Internet are tedious and time consuming and user does not always get what he wants and when he wants it. A user looking for an academic paper in certain topic may spend a considerable amount of time searching for it.

UNIVERSITI MALAYSIA SABAH

This process is repetitive, so much so that when another person who is looking for the same paper, will need to go through the same process of spending certain amount of time to do so. However, if there is a software agent that is aware that both persons are searching for the same information, then this information search process can be reduced. If this knowledge is captured the first time around, then the second person does not need to search for the same document again.

Another scenario is to find a document or information on the web can take up large amount of time for a user. The user will need to search and read the content of a document to find whether it is related to the user requirement. After the first person finds the document that he wants then he will keep it in his personal portable drive or delete it after using it, therefore there will be no record on the search that has been done by the first user.

A second person who has the same interest as the first person might search for similar documents, not knowing that the first person might have the document that he needs. Had he known that the information can be obtained from the first person, he would have not need to do it again but rather to search for new document again. Furthermore when a second user repeats the search, he may not find everything that the first user has found resulting in the second user to miss some of the valuable information that was found by the first user.

UMS
UNIVERSITI MALAYSIA SABAH

To tackle the problems faced by academicians in the process of obtaining the right documents at a minimum time possible, some researches have been conducted. As an example, CiteULike exploits the use of sets of to recommend papers to users. They created a framework for community-based research paper recommender system using social tagging, where members of this website create their own keywords and attach these keywords to the posted papers they read. Next, these tags are mapped with the users' profiles to provide recommendation. Another similar Website to CiteULike is BibSonomy, which is a bookmarking and publication-sharing system. The shortcoming of these algorithms is that, they suffer from the cold-start recommendation. According to Ricci *et al.*, (2011), cold-start recommendation is when a system is unable to provide

recommendation to the users as the users did not provide enough ratings to compute similarity to other users. Another shortcoming of social tagging websites which rely on collaborative approach in their recommendation is a system will not provide real time recommendation unless the user has interacted with system to reach certain threshold. This can be seen in CiteULike, where the user needs to access at least 20 papers for the system before the system is able to recommend academic papers to user.

Another system based on the ACM Subject Hierarchy is proposed by Wang and Bipin (2007). In their algorithm, they proposed to use Association Computing Machinery Subject Hierarchy to classify text based on content similarity and recommend them to users. An obvious shortcoming of this algorithm is that they did not take into account recommendation based on collaborative approach, as recommendation by content-based only will confront users with vast amount of recommendation. As the number of data increases, the recommendation accuracy drops.

Pan and Li (2010) proposed a recommender system based on topic analysis. In their algorithm, they focused on collaborative filtering whereby this technique matches people with similar interests and then makes recommendations on this basis. The shortcoming of this algorithm is that, it does not take into account the similarity of papers content to be recommended to the user based on the user's previously evaluated papers.

Agrawal *et al.*, (2005), proposed a recommender system that recommends academic papers to academicians within the same lab based on their interest. The recommendation was based on like-minded user. Yet, there is an obvious disadvantage in the system, whereby as the number of users increase, the system functionality drops, and causes a delay in information updating as well as in cutting cost in additional host and network resource. This is due to their implementation of a traditional information retrieval system.

UNIVERSITI MALAYSIA SABAH

Hsieh *et al.,* (2004) proposed a system that recommends users with documents which are related to them. In their algorithm, they employed data mining technique that calculates the support and confidence value of each node in their classifying tree structure. Another algorithm proposed by Kuanget *et al.*, (2011), that use of topic model to generate resource recommendation for educational system. In their technique, they implemented Laten Dirichlet Allocation (LDA) which is an approach to building topic models based on formal generative model of documents.

Guoet al. (2007), presented their auto-regulative personal recommendation system based on P2P network. They applied Auto-regulative algorithm of recommendation threshold (ARPRS). The ARPRS is a pure P2P system based on JXTA, an open source P2P platform. Every peer in the system can join or leave discussion group automatic and can share and search and be recommended document. It also acts as a rendezvous peer when necessary.

Nagori and Aghila, 2011, proposed an integrated personalized document recommendation model based on Latent Dirichlet Allocation (LDA), which is originally proposed as a probabilistic document-topic model in text mining domain.

UMS
UNIVERSITI MALAYSIA SABAH

Bighini et al., (2003), proposed an algorithm based on a hybrid recommender system, InLinx, that combine content analysis and the development of virtual clusters of students and of didactical sources providing facilities to use the huge amount of digital information according to the student's personal requirements and interests.

Most of these algorithms has shortcoming such as suffering from cold-start recommendation. Another shortcoming where system is unable to provide user with relevant recommendation as there is insufficient information on user's current interest as users reluctant to express his interest. Another shortcoming is when some systems are able to recommended users with similar papers of what users has already seen only rather than recommending something that users is interested in

but have not seen yet. More elaboration on the finding and shortcoming of review algorithm will be disused under Chapter Two.

1.2 Proposed Agent System Overview

To tackle the problems faced by academician in the academic paper retrieval and acquisition, such as not getting what they want when they want it, and to overcome some of the shortcomings found in the existing recommender systems such as real time recommendation, an agent system is proposed that utilizes a hybrid recommendation approach in its algorithm. By deploying agent technology, the burden from the traditional recommendation (whereby one system does all jobs) will be replaced by a few agents, which divide the problem to be tackled into sub-problems. Each agent is programmed to tackle a particular sub-problem to deliver better result. The advantages of an agent technology will be discussed in Section 2.4

To increase the accuracy of recommendation and to overcome some of the shortcoming of the techniques/methods implemented by the current systems, a hybrid approach is implemented for the recommender agent which is the combination of content-based filtering recommendation approach and collaborative filtering approach.

UMS
UNIVERSITI MALAYSIA SABAH

1.2.1 The Agent System

Four agents have been developed to function as one system. These agents work together to help academicians obtain relevant academic papers, and provides recommendation to users. These agents are, the Monitoring Agent that monitors all the downloading activities by any user at anytime to get user's browsing behavior as well as user's interest at any time. The Categorizer Agent categorizes any downloaded documents by any user at any time into predefined categories. Next, is the Recommender Agent that exploits the information gathered from both the Monitoring Agent and the Categorizer Agent to recommend academic papers to the users. The final agent is the Search Agent that allows users to perform search within the local repository where all the downloaded documents are kept. If the document is not in the local repository, the search agent will extend the search to the Internet.

1.2.2 Recommendation Approaches

In this thesis, a hybrid recommendation approach is used to recommend documents to users by combining two approaches namely, Content-based filtering approach that recommends items to the user based on the descriptions of previously evaluated documents and Collaborative filtering approach that matches people with similar interests and then makes recommendations on this basis. The hybrid system is believed to enhance the recommendation accuracy compared to the existing recommender systems that utilize either content-based filtering alone or Collaborative filtering alone. This is because purely collaborative technique solves the shortcomings of the purely content based approach.

1.3 Overview of Agent

In this section, a brief introduction to Agent technology is given. Next, a few agent's characteristics are discussed. This can help in expressing the reasons researchers are attracted to pay serious attention to this technology in their researches.

1.3.1 Introduction to an Agent

In Computer Science, a software agent is a program that performs some information gathering or processing task in the background. Typically, an agent is given a very small and well-defined task. According to Russell and Norvig (2003), an agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors. Whereas Maes claimed that, autonomous agents are computational systems that inhabit some complex dynamic environment, sense and act autonomously in this environment, and by doing so realize a set of goals or tasks for which they are designed (Maes, 1995).

A software agent can be defined as a computer program which works toward goals in a dynamic environment on behalf of another entity, possibly over an extended period of time, without continuous direct supervision or control, and exhibits a significant degree of flexibility and even creativity in how it seeks and attempts to transform goals into action tasks.

UNIVERSITI MALAYSIA SABAH

1.3.2 Characteristics of an Agent

There are few characteristics that are present in an agent technology that has attracted many researchers and developers to deploy agent technology in their research or applications. According to Wooldridge and Jennings (1995), among these characteristics are:

1. Autonomy, which refers to agents operating with minimum intervention of humans or others, and having some kind of control over their actions and internal state.
2. Proactiveness, that refers to the ability of the agent to exhibit proactive behavior. It consists of pro-active purposeful that is goal-oriented. It attempts to accomplish its goals but does not simply acts in response to the environment. In other words, preconditions described which procedures need to be satisfied in order to be executed when writing a procedure for objectives design. When preconditions are met and procedures are executed correctly, then the post conditions specified will be true. Besides, agents having proactiveness are able to exhibit goal-directed behavior by taking the initiative.
3. Reactiveness refers to goal-directness, as epitomized via execution of a simple procedure that has two inherent limitations. First, it assumes that while the procedure is executing, the preconditions remains valid which means the environment does not change. Second, it presupposes that the goal and the conditions for pursuing that goal remain valid at least until the procedure terminates. Both assumptions are not realistic in complex, dynamic and uncertain environments. Agents should not only blindly attempt to achieve their own goals, but they should be able to perceive changes in environment and responds accordingly in time. This reactiveness characteristic involves sensing and acting.
4. Social Ability is a characteristic that states that the agent may live and act in an environment along with other agents and human. Social ability simply means that agent is able to coordinate, cooperate, negotiate and even compete with others in order to achieve one's objectives.

By exploiting the characteristics found in the agent technology, such as autonomous, proactiveness and social ability, it is believed that the agent system will help to overcome the shortcomings found in traditional systems.

UNIVERSITI MALAYSIA SABAH

1.4 Research Objectives

The objectives of this research are as follow:

1. To introduce the use of an agent technology in developing an information retrieval system in academic domains.
2. To evaluate the effectiveness of using an agent technology to overcome the shortcoming presents in existing systems that use monolithic/traditional way such as cold start recommendation.
3. To introduce the use of hybrid approach in the recommendation algorithm to improve recommendation accuracy over using a single recommendation approach.

1.5 Organization of the Thesis

The remainder of this thesis is structured in the following way:

Chapter 2 reviews existing systems/applications that are related to this work. Besides that, a review on how an agent technology is being deployed by a variety of domains such as in e-commerce domain, movie domain, news domain, document domain and many more is discussed. The benefits, features and advantages of agent technology are also described in details. Next, a comprehensive and systematic study of an agent system and the methods used by other researchers are described.

UMS
UNIVERSITI MALAYSIA SABAH

Chapter 3 describes the design and development of the agent system by describing the design of each of the four agents involved, by detailing the methods and techniques used in building each of the four agents. Besides that, the agent system architecture is also explained in detail.

Chapter 4 discusses the experimental evaluation. This chapter describes the experiments that were carried out to evaluate the effectiveness and usefulness of the proposed agent system.

Chapter 5 summarizes the research work, revisits the objectives and describes the novelty of the research. It concludes by describing future works that can be undertaken.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The intention of this chapter is first to investigate how information explosion on the web is affecting academic users to find academic papers that are of their interest. Next, is to investigate how researchers are tackling this problem and what techniques are used to help academic users to overcome the problem of finding the correct academic paper.

This chapter also gives an overview of agent technology. An overall introduction to agent technology is discussed as well as how widely agent technology is being deployed by many domains. Besides that, features, benefits, advantages and characteristics of agents are also discussed.

The requirements in building an agent system are also discussed in this chapter. Related work on monitoring agent, categorizer agent, recommender agent and search agent are also described and reviewed.

UMS
UNIVERSITI MALAYSIA SABAH

2.2 Problem Statement

Academic domain just like many other domain, has undergone a tremendous growth on both users and papers. Nowadays, academician are finding it difficult to locate academic papers within their interest, as they are confronted with many results from search engines due to the huge number of relevant sites available on the internet.