

INTELLIGENT INFORMATION AGENT FOR NEWS COLLECTION

A thesis submitted to the Graduate School in partial fulfillment of the requirements
for the degree Master of Science (Intelligent Knowledge Based System),

University Utara Malaysia

By

Sasalak Tongkaw

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a post graduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor or, in their absence, by the Dean of the Graduate School. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for an scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

**Dean of Graduate School
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman**

ABSTRACT (BAHASA MELAYU)

Dalam era teknologi maklumat, banyak sumber maklumat boleh didapati dalam Internet. Bilangan berita yang terdapat dalam Internet juga bertambah. Namun begitu adalah sukar bagi pengguna mendapat berita yang tepat seperti yang dikehendakinya. Sering kali enjin pencari memaparkan terlalu banyak, terlalu sedikit atau tiada langsung laman web yang berkaitan dengan apa yang dicari oleh pengguna. Oleh itu, projek ini dilaksanakan untuk mengatasi kelemahan enjin pencari yang sedia ada dengan membangunkan perisian agen bagi mengumpul dan menganalisa berita dalam Internet. Kajian ini memfokus kepada penggunaan agen pintar yang boleh mengumpul, menapis dan menilai berita berdasarkan kata kunci yang diberi oleh pengguna dan menyimpan maklumat berita dalam struktur ontologi. Spesifikasi agen pintar dibina menggunakan kaedah *Unified Modeling Language (UML)* dan ditulis dalam bahasa pengaturcaraan *Java*. *Protégé-2000* digunakan untuk membina struktur berita ontologi.

ABSTRACT (ENGLISH)

In the Information Technology (IT) era we are living, a growing number of information sources are available online. News that is available on the World Wide Web is also increasing. It is often not easy to search for a particular news that one needs. Most of the times, the search engine presents results with too many web pages to visit and very few, if at all, are relevant to what is actually needed. Therefore, we aim to solve this problem by developing a powerful software agent to gather and analyze news information. This project focuses on an Intelligent Agent (IA), which is able to gather news information, filter, score and rate according to the keywords entered by users and store it into ontology structure. This IA is specified using the Unified Modeling Language (UML) and written in Java. Protégé-2000 was used to create the news ontology structure.

Keywords: Intelligent Agent, Ontology, Information Gathering

ACKNOWLEDGMENTS

I am grateful to Prof. Madya Dr. Ku Ruhana Ku Mahamud, and Ms. Norita Md. Norwawi of University Utara Malaysia, for her considerable contribution in discussing this study.

I am especially indebted to Joseph P. Bigus for his contribution to an earlier version of this work. We are thanking of Protégé-2000 developer team for special software of ontology.

The love and understanding of my husband, Aumnat Tongkaw, my son, Modem, and my beloved mother that made this work possible and worthwhile. Lastly, all my IKBS friends who have given their assistance throughout the study. There is tremendous sense of achievement in completing this study.

Sasalak Tongkaw
Sekolah Siswazah
Universiti Utara Malaysia
July 2002

TABLE OF CONTENTS

	Page
PERMISSION TO USE	i
ABSTRACT (BAHASA MELAYU)	ii
ABSTRACT (ENGLISH)	iii
ACKNOWLEDGMENT	iv
TABLE OF CONTENTS	v
LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER ONE : INTRODUCTION	1
1.1 Problem Statement	3
1.2 Objective	3
1.3 Significant Of The Study	4
1.4 Scope, Assumption And Limitation	4
1.5 Methodology	4
CHAPTER TWO : LITERATURE REVIEW	5
2.1 Introduction	5
2.2 Intelligent Agent	5
2.3 Information Gathering	8
2.4 XML/RDF	9
2.5 Ontology	14
2.6 Summary	16
CHAPTER THREE : ANALYSIS OF THE SYSTEM	17
3.1 Introduction	17
3.2 Problem Formulation	17
3.3 Analysis of the Intelligent Agent	17
3.4 Analysis of the Ontology	18
3.3 Summary	20

TABLE OF CONTENTS

	Page
CHAPTER FOUR : SYSTEM DESIGN	21
4.1 Introduction	21
4.2 System Design	21
4.3 Agent Design	22
4.4 Ontology Design	27
4.5 Summary	33
CHAPTER FIVE : SYSTEM IMPLEMENTATION & TESTING	34
5.1 Introduction	34
5.2 Agent Implementation	34
5.3 Ontology Implementation	38
5.4 System Testing	42
5.5 Summary	43
CHAPTER SIX : CONCLUSION & FUTURE WORK	44
6.1 Conclusion	44
6.2 Future Work	45
REFERENCES	46
APPENDICES	
APPENDIX A : User Manual	
APPENDIX B : Source Code	

LIST OF TABLES

	Page
Table 4.1 Database Design	22
Table 4.2 RDF Classes	28
Table 4.3 RDF Properties	29

LIST OF FIGURES

	Page
Figure 3.1: RDF Description	19
Figure 4.1: System Design	21
Figure 4.2: Use Case Diagram of The System	22
Figure 4.3: Class Diagram	24
Figure 4.4: Collaboration Diagram of URLReaderAgent Class	26
Figure 4.5: News Ontology Structure	32
Figure 4.6: Protégé-2000 Layout	32
Figure 5.1: URLReaderAgentCustomizer Class Layout	36
Figure 5.2: FilterAgentCustomizer Class Layout	36
Figure 5.3: Protégé-2000 Dialog Box for Create RDF Schema	38
Figure 5.4: Relationship between Classes and Inheritance	39
Figure 5.5: RDF	40
Figure 5.6: RDF Schema	40
Figure 5.7: Instances of Top News	41
Figure 5.8: Protégé-2000 GUI for Queries	41
Figure 5.9: One of The Query Results	42
Figure 5.10: GUI of Agent	43

CHAPTER 1

INTRODUCTION

In the Information Technology era we are living in nowadays, it is undeniable that to be successful in most occupations, ones must stay on top of current information. One of the best ways to obtain updated information is to find a way to receive news as much as possible. News that is available on the World Wide Web is increasing importance and popularity because of many reasons. First, it is updated more frequently than many other news media such as the daily newspapers. Thus, information on the Internet is often the most current. An easy example of this is the sport news, which is reported minute-by-minute. Second, users that have access to the Internet, can receive news through this medium at a very low cost and usually for free. Third, unlike other media such as television, radio, or newspaper, the information is accessible at all times on the Internet. Fourth, news on the World Wide Web comes in a wider variety than in other media. One can find details of news that particularly serve their needs, which might not be publicised much in other media. Furthermore, specific news can often be searched without having to go through each news one by one, which would require countless amount of time. There is so much information flooding on the Internet each day, and is still increasing rapidly, it is often not easy to search for that particular news that one needs. The traditional way of searching is to enter queries into a search engine.

Most of the time, the search presents result with too many web pages to visit and very few, if at all, are relevant to what is actually needed. Thus, this kind of search is neither efficient, nor practical for most people. Therefore, we aim to solve this problem by developing a more powerful approach to search and retrieve news information. This particular application allows users to retrieve news faster and more efficiently.

The contents of
the thesis is for
internal user
only

REFERENCES

- Abasolo, J.M. and Melisa, M.G. (2000) "An Ontology-Based Agent for Information Retrieval in Medicine", ECDL 2000 Workshop on the Semantic Web.
- Ambite, J.L. and Knoblock, C.A. (1997) "Agents for Information Gathering", IEEE Expert, September/October, pp. 2-4.
- Bigus, J.P. (2000) "Constructing Intelligent Agents Using Java", Wiley Computer Publishing. pp.285-324.
- Brickley, D. (2002) "RDF Vocabulary Description Language 1.0: RDF Schema", web site: <http://www.w3.org/TR/2002/WD-rdf-schema-20020430/>, accessed date: 30/05/2002.
- Chen, Y.J., and Soo, V.W. (2001) "Ontology-Based Information Gathering Agents Web Intelligence", pp. 423-427.
- Desouza, K. (2001) "Intelligent Agents for Competitive Intelligence: A Survey of Applications", In Press of Competitive Intelligence Review, 12(4).
- Guarino, N., Masolo, C. and Vetere, G. (1999) "OntoSeek: Content-Based Access to the Web", IEEE Intelligent Systems, 14(3), pp. 70-80.
- Gudivada, V.N., Raghavan, V.V., Grosky, W.I. and Kasanagottu, R. (1997) "Information Retrieval on the World Wide Web", IEEE Internet Computing, 1(5) September/October, pp. 58-68,
- Huhns, M.N and Singh, M. (1997) "Ontologies for Agents", In IEEE Internet Computing November-December, pp. 81-83.
- Kauffman, R.J., March, S.T. and Wood, C.A. (1999) "Agent Sophistication: Design Aspects For Data-Collecting Agents", Proceedings of the Workshop on Information Technology Systems, Charlotte, December, pp. 11-12.
- Kauffman, R.J., March, S.T. and Wood, C.A. (2000) "Mapping Out Design Aspects for Data-Collecting Agents", International Journal of Intelligent Systems in Accounting, Finance, and Management, 9(4), pp. 217-236.
- Klusck, M. (2001) "Intelligent Information Agents", ACAI01, Prague, web site: <http://www.dfki.de/~klusck/>
- Levy, A.Y., Sagiv, Y. and Srivastava, D. (1999) "Towards efficient information gathering agents. In Software Agents", Proceedings of the AAI.
- Noy, N.F., Sintek, M., Decker, S., Crubézy, M., Ferguson, R.W and Musen, M. A. (2001) "Creating Semantic Web Contents with Protégé-2000", IEEE Intelligent Systems, 16(2), pp. 60-71.
- Ramesh, B. and Tiwana, A. (2001) "Integrating Knowledge on the Web", IEEE Internet Computing, May/June.
- Smith, H. and Poulter, K. (1999) "The Role of Shared Ontology in XML-Based Trading Architectures", Included within Communications of the ACM, special issue on Agent Software, ontology.org
- Taveter, K. (1998) "Intelligent Information Retrieval Based on Interconnected Concepts and Classes of Retrieval Domains", Proceedings of the Eighth DELOS Workshop on User Interface in digital Libraries, Stockholm, Sweden, October 21-23, ERCIM Workshop Proceedings, No. 99-W001, pp.39-43.
- Woelk, D., Cannata, P., Huhns, M., Shen, W. and Tomlinson, C. (1993) "Using Carnot for Enterprise Information Integration", In Proceedings of Second International Conference on Parallel and Distributed Information Systems. pp. 133-136.

- Wooldridge, M. and Jennings N.R. (1992) "Intelligent Agents: Theories and Practice", Knowledge Engineer Review, pp. 25-71.
- Wooldridge, M. and Jennings, N.R. (1995) "Agent Theories, Architectures, and Languages: A Survey", In M. J. Wooldridge and N. R. Jennings, editors, Proc. ECAI-94 Workshop on Agent Theories, Architectures and Languages. Springer-Verlag, pp.1-31.