



INTELLIGENT SYSTEMS REFERENCE LIBRARY
Volume 8

Edward H.Y. Lim
James N.K. Liu
Raymond S.T. Lee

Knowledge Seeker – Ontology Modelling for Information Search and Management

A Compendium

 Springer

Edward H.Y. Lim, James N.K. Liu, and Raymond S.T. Lee

Knowledge Seeker – Ontology Modelling for Information Search and
Management

Intelligent Systems Reference Library, Volume 8

Editors-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Prof. Lakhmi C. Jain
University of South Australia
Adelaide
Mawson Lakes Campus
South Australia 5095
Australia
E-mail: Lakhmi.jain@unisa.edu.au

Further volumes of this series can be found on our homepage: springer.com

Vol. 1. Christine L. Mumford and Lakhmi C. Jain (Eds.)
*Computational Intelligence: Collaboration, Fusion
and Emergence*, 2009
ISBN 978-3-642-01798-8

Vol. 2. Yuehui Chen and Ajith Abraham
*Tree-Structure Based Hybrid
Computational Intelligence*, 2009
ISBN 978-3-642-04738-1

Vol. 3. Anthony Finn and Steve Scheduling
*Developments and Challenges for
Autonomous Unmanned Vehicles*, 2010
ISBN 978-3-642-10703-0

Vol. 4. Lakhmi C. Jain and Chee Peng Lim (Eds.)
*Handbook on Decision Making: Techniques
and Applications*, 2010
ISBN 978-3-642-13638-2

Vol. 5. George A. Anastassiou
Intelligent Mathematics: Computational Analysis, 2010
ISBN 978-3-642-17097-3

Vol. 6. Ludmila Dymowa
Soft Computing in Economics and Finance, 2011
ISBN 978-3-642-17718-7

Vol. 7. Gerasimos G. Rigatos
Modelling and Control for Intelligent Industrial Systems, 2011
ISBN 978-3-642-17874-0

Vol. 8. Edward H.Y. Lim, James N.K. Liu, and Raymond S.T. Lee
*Knowledge Seeker – Ontology Modelling for Information
Search and Management*, 2011
ISBN 978-3-642-17915-0

Edward H.Y. Lim, James N.K. Liu, and
Raymond S.T. Lee

Knowledge Seeker – Ontology Modelling for Information Search and Management

A Compendium

Edward H.Y. Lim
Department of Computing
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
E-mail: edward@iatopia.com

Raymond S.T. Lee
IATOPIA Research Lab
Suite 1515, Star House, 3 Salisbury Road
Tsim Sha Tsui, Kowloon, Hong Kong
E-mail: raymond@iatopia.com

James N.K. Liu
Department of Computing
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
E-mail: csnkliu@inet.polyu.edu.hk

ISBN 978-3-642-17915-0

e-ISBN 978-3-642-17916-7

DOI 10.1007/978-3-642-17916-7

Intelligent Systems Reference Library

ISSN 1868-4394

© 2011 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

Ontology is being a fundamental form of knowledge representation about the real world. In the computer science perspective, ontology defines a set of representational primitives with which to model a domain of knowledge or discourse (Gruber 2008). A well constructed ontology can help developing knowledge-based information search and management system, such as search engine, automatic text classification system, content management system, etc, in a more effective way. Most of these existing systems are ineffective in terms of its low accuracy in searching and managing information (especially text data), because of lacking knowledge as the core components used in the systems. Ontology is therefore becoming a very important research area for developing those knowledge-based information systems, as ontology is a recognized form of representing a particular knowledge domain.

The most challenge of ontology research is how to create and maintain ontology. Ontology engineering is a kind of this ontology research for developing theories, methods, and software tools that help creating and maintaining ontology. The ontology engineering methods developed in the past mainly consist of manual ontology creation and automatic ontology learning methods. Although many ontology engineering tools have been developed over the last decade, most of them involve manual creating and maintaining ontology which is a time consuming and inefficient task as every process requires deep analysis by domain experts. There is also a problem that domain experts may create ontology by different and subjective view, so that the ontology knowledge is not exact and may not be relevant to all knowledge domains. Therefore, automatic ontology learning is a more practical method in ontology engineering. This ontology learning method tried to develop an automated process to extract knowledge from some computer data and present as a specific form of ontology, with the least or minimum involvement of human work.

Ontology learning from text is the most useful method in formalizing ontology, as text data is a rich and direct source of human knowledge. However, analyzing textual data by computer is a difficult task, as it requires some natural language processing and semantic analysis. Many methodologies on ontology learning from text have been widely developed in recent years (Maedche 2002, Buitelaar and Cimiano 2008). Most of them use artificial intelligent approaches such as machine learning or statistical analysis to develop the methodologies, and try to extract ontology from text learning automatically.

A lot of researches on ontology learning in text have been carried out in recent years. However, most of them applied only on English text, as text is a language dependent data, algorithms applied on English text were found not working

well in Chinese text. Therefore, the research and some related experiments on ontology learning in Chinese text are focused in this project, as we aim to develop an efficient ontology learning method which can be applied to Chinese text data. Information search and management systems that contain mainly Chinese text data hence can be enhanced by the ontology, because many existing ontology are developed in English which cannot be applied to Chinese based information system. In this research project, we aim to develop a comprehensive system framework of ontology learning in Chinese text which can be applied to Chinese based information search and management systems.

The overall objective of this research is to develop an ontology based system framework, called KnowledgeSeeker, which can be used to develop various ontology based information systems. These information systems mainly include Chinese based text information retrieval system, such as search engine and content management system. The final goal of the KnowledgeSeeker system framework is that it can improve the traditional information system with higher efficiency. In particular, it can increase the accuracy of a text classification system, and also enhance the search intelligence in a search engine. This can be done by enhancing the system with machine processable knowledge (ontology). It has been mentioned that ontology is a useful component in developing knowledge based intelligent information systems, but the problem is that lots of research work are still required to find out the method of creating and maintaining a relevant ontology for used in the information system. Therefore we raise the following research questions to define the scope of this research work:

- What format of the ontology can be modeled in computer system?
- What learning method is used to automatically create the ontology data?
- How to generate the ontology for machine and also be visualized for human use?
- What operations can be done with the ontology and how it operates?
- What applications can be developed by using the ontology and how is the performance?

This book is organized in three parts containing twelve chapters:

Part I (Chapters 1-4): Introduction

Part II (Chapters 5-8): KnowledgeSeeker - An Ontology Modeling and Learning Framework

Part III (Chapter 9s-12): KnowledgeSeeker Applications

The book is outlined as follows:

Chapter 1 presents briefly the philosophical question about knowledge and ontology, and also the perspective of ontology in information system.

Chapter 2 presents the ontology engineering approaches in the recent researches, including the fundamental concepts about ontology and ontology learning.

Chapter 3 presents the traditional text information retrieval system and related models and algorithms.

Chapter 4 presents the research about web data semantics and some semantic modeling technologies for developing intelligent systems.

Chapter 5 presents an overview of the system framework called KnowledgeSeeker. It also presents the proposed graphical based model of ontology called Ontology Graph.

Chapter 6 presents the methodology of automatic ontology learning in Chinese text. Relevant examples and experiments are presented to illustrate the proposed methodology.

Chapter 7 presents the Ontology Graph generation process.

Chapter 8 presents different kinds of Ontology Graph based operation and operating methods.

Chapter 9 presents an application, a text classification application with experiment, which adopts the technology of KnowledgeSeeker for classification. It provides experimental result and performance analysis.

Chapter 10 presents a commercial application which adopts the techniques of KnowledgeSeeker system framework called IATOPIA iCMS KnowledgeSeeker, which is ontology based digital assets management system for managing multimedia files.

Chapter 11 presents another commercial application called IATOIPA News Channel (IAToNews), which is an ontology-based news system provided in web environment.

Chapter 12 presents a collaborative content and user-based web ontology learning system, which enhances the ontology learning method by user-based knowledge.

Acknowledgements

Upon the completion of this book, we would like to take the opportunity to express our sincere gratitude to the following parties or persons. We are grateful for the partial support of GRF grant 5237/08E and CRG grant G-U756 of Department of Computing, The Hong Kong Polytechnic University. We also thank IATOPIA Research Lab for providing the facilities and convenience for system testing. Lastly, we thank Prof. Lakhmi Jain of the KES Centre, University of South Australia who encouraged us to carry out the book.

Edward H.Y. Lim^{1,2}, James N.K. Liu¹, and Raymond S.T. Lee²

¹Department of Computing
The Hong Kong Polytechnic University
²IATOPIA Research Lab

Contents

Part I: Introduction

1 Computational Knowledge and Ontology	3
1.1 What Is Knowledge?.....	3
1.2 Computational Knowledge in Information Systems	3
1.2.1 Knowledge Engineering	4
1.2.2 Knowledge Representation	4
1.3 What Is Ontology?.....	6
1.4 Ontology Modeling in Computer Systems.....	7
1.4.1 Computational Ontology Representation.....	7
1.4.2 Top-Level Ontologies	8
1.4.3 Lexical Ontologies	9
1.4.4 Domain Ontologies	11
2 Ontology Engineering.....	13
2.1 Introduction	13
2.2 Ontology Fundamentals.....	14
2.2.1 Ontological Structure	14
2.2.2 Ontological Taxonomy	15
2.2.3 Ontological Relations	16
2.2.4 Lexical Relations	18
2.3 Ontology Engineering Tools.....	19
2.3.1 Protégé ontology editor.....	20
2.4 Ontology Learning from Text.....	21
2.4.1 Learning Terminology	23
2.4.2 Learning Taxonomy.....	25
3 Text Information Retrieval.....	27
3.1 Information Retrieval Model	27
3.1.1 Term Weighting Model.....	27
3.1.2 Text Classification	28
3.1.3 Text Clustering	31

3.2	Feature Selection and Reduction	32
3.3	Search Engine Modeling	34
3.4	Evaluation Methods	35
3.4.1	Performance measurement	36
4	Web Data Semantics	37
4.1	Semantic Web	37
4.1.1	W3C Semantic Web	39
4.2	Semantic Modeling	40
4.2.1	Taxonomy	41
4.2.2	Thesaurus	41
4.2.3	Topic Maps	42
4.2.4	Ontology	43
4.2.5	Ontology Languages for the Semantic Web	44
4.3	Semantic Annotation and Information Retrieval	46
<hr/>		
Part II: KnowledgeSeeker: An Ontology Modeling and Learning Framework		
5	Ontology Modeling Framework	49
5.1	KnowledgeSeeker – The System Overviews	49
5.2	Background of Signs System and Ontology	51
5.2.1	The Semiotics	51
5.2.2	The Sign System for Concept Formation	54
5.2.3	The Meaning of Words	56
5.2.4	The Semantics of Relations	57
5.3	Ontology Graph Overviews	59
5.3.1	Nodes in Ontology Graph	60
5.3.2	Term Nodes in Ontology Graph	62
5.3.3	Words Function	64
5.3.3.1	The Function of Language	64
5.3.3.2	Understanding and Feeling	65
5.3.3.3	Meaning and Information	67
5.4	The Implementation of Ontology Graph	68
5.4.1	The Conceptual Structure of Ontology Graph	68
5.4.1.1	Conceptual Units in Ontology Graph	69
5.4.2	The Class Diagram of Ontology Graph	69
6	Ontology Learning in Chinese Text	71
6.1	The Ontology Learning Method	71
6.1.1	Term Extraction	72
6.1.2	Term-to-Class Relationship Mapping	73
6.1.2.1	Term-to-Class Independency Measurement by χ^2	73
6.1.2.2	Term-to-Class Positive and Negative Dependency Measurement by R	78

6.1.3.	Term-to-Term Relationship Mapping	81
6.1.3.1	Term-to-Term Independency Measurement by χ^2	83
6.1.4.	Concept Clustering	91
6.1.5	Sample Result of Domain Ontology Graph Generation (10 Domains)	96
7	Ontology Graph Generation Process	99
7.1	Ontology in Information Systems	99
7.2	Ontology Graph Generation Process in KnowledgeSeeker	100
7.2.1	Definition of Ontology Graph Structure	101
7.2.2	Domain Ontology Graph Generation Process.....	103
7.2.3	Document Ontology Graph Generation	105
7.3	Automatic Generation of Domain Ontology Graph.....	107
7.3.1	Experimental Setup.....	108
7.3.2	Experimental Results	110
8	Ontology Graph Operations	121
8.1	Ontology Graph Matching and Querying Process	121
8.2	Introduction to Ontology Matching and Mapping	121
8.2.1	Ontology Graph Matching Methods	122
8.2.1.1	Semantic Mapping Function	122
8.2.1.2	Similarity Measurement Function.....	123
8.3	Matching Different Components with Ontology Graph	124
8.3.1	Matching Terms to Domain Ontology Graph	125
8.3.2	Matching Text Document to Domain Ontology Graph	128
8.3.3	Ontology Graph Based Similarity Measurement	132
8.3.4	Matching Two Document Ontology Graphs.....	134
8.4	Overviews of Ontology Graph Based Querying	137
8.4.1	Ontology Graph Querying Methods	137
8.5	Operations in Ontology Graph Based Querying	137
8.5.1	Querying Document with Document Ontology Graph	138

Part III: KnowledgeSeeker Applications

9	Ontology Graph Based Approach for Automatic Chinese Text Classification	145
9.1	Introduction	145
9.2	Methodologies	147
9.2.1	Ontology Graphs Reviews	147
9.2.2	Classification Algorithm.....	148
9.3	Experiments	149
9.3.1	Experiment Description	150
9.3.1.1	Evaluate the Performance of Ontology-Graph Based Approach (Experiment 1).....	150

9.3.1.2	Evaluate the Optimal Size of Domain Ontology Graph for Best Classification Result (Experiment 2).....	151
9.3.1.3	Evaluate the Effects for Setting Different Thresholds of weight of Domain Ontology Graph (Experiment 3)	151
9.3.2	Evaluation Method.....	151
9.3.3	Performance on Ontology Graph Based Text Classification	152
9.3.4	Experimental Results	153
9.3.4.1	Performance on Ontology Graph Based Text Classification (Experiment 1)	153
9.3.4.2	Performance on Using Different Size of Terms (Dimensionality) for Text Classification (Experiment 2)	156
9.3.4.3	Result of Using Different Thresholds of Weight of Domain Ontology Graphs (Experiment 3).....	161
9.3.4.4	Combining the Results and Optimizing the Parameters for the Text Classification Experiments	163
10	IATOPIA iCMS KnowledgeSeeker – An Integrated Content Management System and Digital Asset Management System (DAMS).....	165
10.1	IATOPIA iCMS KnowledgeSeeker.....	165
10.1.1	System Features	165
10.1.2	System Model and Architecture.....	166
10.1.3	Ontology System.....	168
10.1.4	Search Indexing System.....	168
10.1.5	IATOPIA Search Engine	169
10.2	IATOPIA Digital Asset Management System (DAMS).....	170
10.2.1	DAMS System Architecture Overview.....	170
10.2.2	IATOIPA iCMS Databank Cluster in DAMS.....	170
10.2.3	Ontology System in DAMS	171
10.2.4	DAMS and Web Channel Interface Examples.....	172
11	IATOPIA News Channel (IAToNews) – An Intelligent Ontological Agent-Based Web News Retrieval and Search System.....	175
11.1	Introduction.....	175
11.2	IAToNews System Architecture Overview	175
11.3	Ontology System in IAToNews.....	176
11.3.1	Article Ontology	177
11.3.2	Topic Ontology	178
11.3.3	Ontology Based Content Indexing.....	178
11.4	IAToNews Web Interface Examples	179

12 Collaborative Content and User-Based Web Ontology Learning System181

12.1 Introduction.....181

12.2 Background.....182

 12.2.1 Problem of Building a Generalized Ontology to Satisfy Daily Life.....182

 12.2.2 Semantic Web.....182

 12.2.3 Web Channels.....182

 12.2.4 BuBo (Feedback, Personalization with User Ontology)183

12.3 Methodology.....183

 12.3.1 Overview of System Architecture.....183

 12.3.2 Content-Based Ontology Learning Process184

 12.3.3 User-Based Ontology Personalization Process187

12.4 Implementation192

 12.4.1 Architecture of the Collaborative System Implementation.....192

 12.4.2 Structure of the Specified Domains Ontology with Generalized OG193

 12.4.3 Ontology-Based Search Engine Within BuBo.....194

12.5 Conclusions.....194

References195

Appendix211

List of Figures

1.1	SUMO top level.....	8
1.2	OpenCyc upper-level ontology hierarchy (OpenCyc 2003)	9
1.3	Graphical expression of Chinese lexical concept in HowNet.....	10
1.4	Basic data - Taxonomy in HowNet.....	11
1.5	Ontology sample of entertainment domain.....	12
2.1	Subject-relation-object representation	15
2.2	Taxonomical relations and non-taxonomical relations in ontology	15
2.3	Taxonomy and partonomy relations in ontology	16
2.4	Taxonomy and partonomy relations in ontology	16
2.5	Ontological relations by universals and particulars example.....	17
2.6	Ontological relations by universals and particulars example.....	18
2.7	The Protégé-Frames editor for OKBC ontology modeling.....	20
2.8	The Protégé-OWL editor for OWL ontology modeling	21
2.9	Degree of knowledge formalization in seven levels	22
3.1	8 x 7 term-by-document matrix for 7 documents and 8 terms	28
3.2	kNN classification approach.....	30
3.3	Inputs, output and function in a neuron	30
3.4	3-layer structure of neural network classification.....	31
3.5	Search engine process and its components	34
4.1	Classification model and its sub-processes.....	38
4.2	Semantic Web stack.....	40
4.3	The semantic modeling techniques in levels.....	40
4.4	Example of a financial site map organized in a taxonomy hierarchy	41
4.5	The semantic modeling techniques in levels.....	42
4.6	XTM example.....	43
4.7	Language for ontology modeling.....	44
4.8	RDF subject-predicate-object triple.....	44
4.9	RDF triple relations	44
4.10	RDF example.....	45
4.11	OWL example.....	45
5.1	Four modules in KnowledgeSeeker system framework.....	49
5.2	Knowledge components of ontology learning from text.....	50
5.3	A sign represented by Icons.....	52

5.4	Saussure's sign model.....	53
5.5	Signification between concept and word	53
5.6	Signification between concept and word	53
5.7	Relations between signs.....	54
5.8	The Meaning Triangle (Ogden and Richards 1923)	54
5.9	The meaning triangle of the symbol "Dog"	55
5.10	Signs and relations	55
5.11	Ontological triples	56
5.12	Word network by word triples	57
5.13	Sentence triple for word network.....	57
5.14	Conceptual units as a network model	60
5.15	A formation of meaning with terms - example	61
5.16	A formation of meaning with symbols - example.....	61
5.17	Example of meaning formation	64
5.18	Jakobson's model of the function of language.....	64
5.19	The functions of language and knowledge	65
5.20	The functions of language and knowledge	66
5.21	Relation types expressing different nature of knowledge	67
5.22	Conceptual structure of Ontology Graph (OG) in KnowledgeSeeker	68
5.23	The class relationship of Ontology Graph	70
6.1	Bottom-up approach of the Ontology Graph learning process	72
6.2	Term mapping to classes.....	73
6.3	Term-to-class Relationship mapping steps	74
6.4	8 x 10 term-by-document matrix for 10 documents and 8 terms	75
6.5	Term mapping to classes from the example.....	76
6.6	Updated term-to-class relationship mapping steps	79
6.7	Terms mapping to each other.....	82
6.8	Create document links to each term in the term-group	86
6.9	Term-to-term relationship mapping steps	87
6.10	Ontology Graph created for 4 terms in class c_1 ($k=4$)	91
6.11	Terms-to-concept clustering steps	92
6.12	Final Ontology Graph created for class c_1	92
6.13	The distribution of terms in their POS	96
6.14	<i>DOG</i> (文藝 Arts and Entertainments)	97
6.15	<i>DOG</i> (政治 Politics).....	97
6.16	<i>DOG</i> (政治 Traffic).....	97
6.17	<i>DOG</i> (教育 Education).....	97
6.18	<i>DOG</i> (環境 Environment)	97
6.19	<i>DOG</i> (經濟 Economics)	97
6.20	<i>DOG</i> (經濟 Military).....	98
6.21	<i>DOG</i> (醫療 Health and Medical).....	98
6.22	<i>DOG</i> (電腦 Computer and Information Technology)	98
6.23	<i>DOG</i> (體育 Sports).....	98

7.1	Three types of Ontology according to their level of abstraction.....	100
7.2	Types of ontology in different levels	100
7.3	Domain Ontology Graph generation process.....	103
7.4	A graphical representation of the <i>DOG</i> of “ <i>entertainment</i> ”	105
7.5	Document Ontology Graph generation process	106
7.6	The steps of automatic domain ontology graph generation	107
7.7	The interface of yED Graph Editor (yED 2009).....	108
7.8	Sample of a generated GraphML document of <i>DOG</i>	109
7.9	Number of nodes generated for different threshold values	110
7.10	The rates of nodes generated for different threshold values	110
7.11	Number of edges generated for different threshold setting.....	112
7.12	The rate of edges generated for different threshold setting.....	112
7.13	$\theta_u = 20, \theta_v = 200$	114
7.14	$\theta_u = 20, \theta_v = 150$	114
7.15	$\theta_u = 20, \theta_v = 100$	114
7.16	$\theta_u = 20, \theta_v = 90$	114
7.17	$\theta_u = 20, \theta_v = 80$	115
7.18	$\theta_u = 20, \theta_v = 70$	115
7.19	$\theta_u = 20, \theta_v = 60$	116
7.20	$\theta_u = 20, \theta_v = 50$	116
7.21	$\theta_u = 20, \theta_v = 40$	117
7.22	$\theta_u = 20, \theta_v = 30$	117
7.23	$\theta_u = 20, \theta_v = 20$	117
7.24	$\theta_u = 20, \theta_v = 10$	117
7.25	$\theta_u = 20, \theta_v = 0$	117
7.26	$\theta_u = 120, \theta_v = 200$	118
7.27	$\theta_u = 120, \theta_v = 160$	118
7.28	$\theta_u = 120, \theta_v = 120$	118
7.29	$\theta_u = 120, \theta_v = 100$	118
7.30	$\theta_u = 120, \theta_v = 80$	118
7.31	$\theta_u = 120, \theta_v = 60$	118
7.32	$\theta_u = 120, \theta_v = 40$	119
7.33	$\theta_u = 120, \theta_v = 20$	119
8.1	Ontology matching overviews	122
8.2	Terms mapping example.....	122
8.3	Components of Ontology Graph mapping process	124
8.4	Components of concept formation process.....	125
8.5	Process of concept formation.....	125
8.6	Ontology Graph OG_A	126
8.7	OG of Concept c_A	127
8.8	OG of Concept c_B	127
8.9	OG of Concept c_C	127
8.10	OG of Concept c_D	128
8.11	OG of Concept c_E	128

8.12	Mapping overlapping terms in document and Domain Ontology Graph	128
8.13	Components of Document Ontology Graph generation process.....	129
8.14	Process of Document Ontology Graph (<i>DocOG</i>) extraction	129
8.15	<i>DocOG</i> for $d_1 - OG_{d_1}$	132
8.16	<i>DocOG</i> for $d_2 - OG_{d_2}$	132
8.17	Components of Document and Domain Ontology Graph comparison	133
8.18	Process of Document and Domain Ontology Graph comparison	133
8.19	Components of comparison between two documents.....	135
8.20	Process of documents comparisons	135
8.21	OG of mapping result	136
8.22	Ontology querying overviews.....	137
8.23	Components of Ontology Graph querying process.....	138
8.24	Components of document ontology graph formation process	139
8.25	Process of Document Ontology Graph extraction	139
8.26	<i>DocOG</i> for $d_1 - OG_{d_1}$	140
8.27	<i>DocOG</i> for $d_2 - OG_{d_2}$	141
9.1	Conceptual structure of Ontology Graph.....	147
9.2	Comparison of document and domain ontology graph for similarity measurement.....	148
9.3	Scoring terms inside a document.....	149
9.4	Result of precision for different approaches	155
9.5	Result of recall for different approaches.....	155
9.6	Result of F-measure for different approaches.....	156
9.7	Result of precision and recall for <i>tf-idf</i> approach.....	157
9.8	Result of precision, recall and f-measure for <i>tf-idf</i> approach.....	158
9.9	Result of precision and recall for term-dependency approach.....	159
9.10	Result of precision, recall and f-measure for term-dependency approach.....	159
9.11	Result of precision and recall for ontology-graph approach.....	160
9.12	Result of precision, recall and f-measure for ontology-graph approach	161
9.13	Precision and recall for different threshold values used	162
9.14	Result of classification in precision, recall and f-measure.....	163
9.15	Result comparison of precision and recall for the three approaches.....	163
9.16	Result comparison of f-measure for the three approaches	164
10.1	The system architecture of iCMS KnowledgeSeeker	166
10.2	Process flow of IATOPIA iCMS KnowledgeSeeker	167
10.3	DAMS architecture overview	170
10.4	IATOPIA iCMS databank cluster for DAMS.....	171
10.5	DAMS interface for movie domain	172
10.6	Web channel linked to DAMS content for movie domain.....	172
10.7	Example of Chinese Opera ontology tree maintained by domain experts.....	173

11.1	IAToNews system architecture and information flow	176
11.2	Ontology system in IAToNews	177
11.3	Article ontology in IAToNews ontology system	178
11.4	Main page displaying classified news in IAToNews	179
11.5	Display news content and related info in IAToNews	180
12.1	Ontology agent application	184
12.2	Four content-based ontology learning processes	184
12.3	Ontology Graph	185
12.4	Ontology graph containing k number of Concepts	187
12.5	Structure of the user preferences XML.....	188
12.6	The top level Ontology Graph	189
12.7	The example structure of Ontology Graph	189
12.8	SG interception with the POG and form new POG	190
12.9	The process of customization and how the user uses the BuBo to interact with the User's Personalization Ontology Search Agent	191
12.10	The ontology learning server and web channel server.....	192
12.11	BuBo Structure and responsibilities.....	193
12.12	The example of the topic-base generalized OG	193

List of Tables

1.1	Knowledge representation formalisms (Guarino 1995).....	5
1.2	Aristotle's categories	6
1.3	Kant's categories	6
2.1	Transitive relations characteristic	18
3.1	Content of terms and documents	28
3.2	Term-to-category contingency table	33
3.3	The table of retrieval result.....	35
4.1	Examples of semantic relations in thesaurus	42
4.2	Components of an ontology.....	43
5.1	Coherence Relations (Louwerse 2001, Mancini and Shum 2006).....	59
5.2	Summary of POS in IG.....	63
5.3	Types of relations - example.....	63
5.4	Factors of Communication and Functions of Language	65
5.5	Referential and Emotive function of Verb and Adjective.....	66
5.6	Factors of Communication and Functions of Language	67
5.7	Class Relations of Ontology Graph Implementation	69
6.1	2 x 2 term-to-class contingency table of term t and class c	73
6.2	Content of term and documents distribution.....	75
6.3	Term-to-class table - 10 documents in 5 classes with 8 candidate terms.....	75
6.4	χ^2 mapping of 5 classes and 8 candidate terms	76
6.5	2x2 term-to-class contingency table of term t_6 and class c_1	77
6.6	2x2 term-to-class contingency table of term t_6 and class c_3	77
6.7	2x2 term-to-class contingency table of term t_6 and class c_1	77
6.8	2x2 term-to-class contingency table of term t_6 and class c_3	77
6.9	$R_{t,c}$ mapping of 5 class and 8 candidate terms.....	79
6.10	Dependency values of term t_1	80
6.11	Dependency values of term t_2	80
6.12	Dependency values of term t_3	80

6.13	Dependency values of term t_4	80
6.14	Dependency values of term t_5	80
6.15	Dependency values of term t_6	80
6.16	Dependency values of term t_7	80
6.17	Dependency values of term t_8	80
6.18	2x2 term-to-term contingency table of term t_a and term t_b	83
6.19	Term dependency $n\chi^2_{t_a, t_b}$ of class c_1	84
6.20	Term dependency R_{w_a, w_b} of class c_1	84
6.21	Term-groups created for class c_1 to c_5	85
6.22	Content of terms and documents distribution of the document corpus D_2	86
6.23	Document link from D_2 to C_{c_1}	88
6.24	Term-to-term table (10 documents for the term-group of class c_1).....	88
6.25	2x2 term-to-term contingency table of term t_1 and term t_1	88
6.26	2x2 term-to-term contingency table of term t_1 and term t_2	89
6.27	2x2 term-to-term contingency table of term t_1 and term t_3	89
6.28	2x2 term-to-term contingency table of term t_6 and term t_4	89
6.29	χ^2 term-to-term mapping of 4 terms in class c_1	90
6.30	Final Result - terms dependency $n\chi^2_{t_a, t_b}$ of class c_1	90
6.31	Final Result - terms dependency R_{t_a, t_b} of class c_1	90
6.32	Class label (Chinese & English) in document corpus D_1	93
6.33	Document distribution of the ten classes (D_1)	93
6.34	Document distribution for learning and testing	94
6.35	Documents distribution of corpus D_2	94
6.36	Statistics of term extraction results	95
6.37	Results of term-to-class relationship mapping.....	95
6.38	Ratio of positive to negative terms in each class	96
7.1	Term list of a sample DOG of domain “entertainment”	104
7.2	Details of the rates of nodes generated for different thresholds values.....	111
7.3	Top 20 terms of domain “文藝” in ontology learning	113
7.4	DOC generation result of the domain “文藝 (Arts and Entertainments)”	113
8.1	Table of the terms and relations in OG_A	126
8.2	Mapping of concept c_A	127
8.3	Mapping of concept c_B	127
8.4	Mapping of concept c_C	127
8.5	Mapping of concept c_D	128
8.6	Mapping of concept c_E	128
8.7	C_A, d_1	130
8.8	C_B, d_1	130
8.9	C_D, d_1	131

8.10	C_{D, d_2}	131
8.11	C_{E, d_2}	131
8.12	Terms and relations in OG_{d_1}	132
8.13	Terms and relations in OG_{d_2}	132
8.14	Terms and relations in OG_{d_1}	134
8.15	Terms and relations in OG_{d_2}	134
8.16	Mapping result	136
8.17	Terms and relations in OG_{d_1}	140
8.18	Terms and relations in OG_{d_2}	141
8.19	Result of q_1 -to- OG_{d_1}	141
8.20	Result of q_1 -to- OG_{d_2}	141
8.21	Result of q_2 -to- OG_{d_1}	141
8.22	Result of q_2 -to- OG_{d_2}	141
8.23	Result of q_3 -to- OG_{d_1}	141
8.24	Result of q_3 -to- OG_{d_2}	141
9.1	The table of retrieval result	151
9.2	Class label (Chinese & English) of training document set	152
9.3	Document count for training and testing document sets	153
9.4	Details of classification result of each class.....	154
9.5	Classification result for <i>tf-idf</i> approach.....	157
9.6	Classification result for term-dependency approach.....	158
9.7	Classification result for ontology-graph approach.....	160
9.8	Experimental result of using different threshold values	162
9.9	Summary of the optimized performance.....	164
11.1	Article data in article ontology	177
11.2	Semantic data in article ontology.....	177
11.3	Topics in IATOPIA KnowledgeSeeker news channel.....	178
11.4	Types of ontology based content index of news article	179
A.1	Top 30 ranked terms selected in Domain Ontology Graph (文藝)	213
A.2	Top 30 ranked terms selected in Domain Ontology Graph (政治)	214
A.3	Top 30 ranked terms selected in Domain Ontology Graph (交通)	215
A.4	Top 30 ranked terms selected in Domain Ontology Graph (教育)	216
A.5	Top 30 ranked terms selected in Domain Ontology Graph (環境)	217
A.6	Top 30 ranked terms selected in Domain Ontology Graph (經濟)	217
A.7	Top 30 ranked terms selected in Domain Ontology Graph (軍事)	219
A.8	Top 30 ranked terms selected in Domain Ontology Graph (醫療)	220
A.9	Top 30 ranked terms selected in Domain Ontology Graph (電腦)	221
A.10	Top 30 ranked terms selected in Domain Ontology Graph (體育)	222

A.11	Terms dependency in Domain Ontology Graph (文藝).....	223
A.12	Terms dependency in Domain Ontology Graph (政治).....	224
A.13	Terms dependency in Domain Ontology Graph (交通).....	225
A.14	Terms dependency in Domain Ontology Graph (教育).....	226
A.15	Terms dependency in Domain Ontology Graph (環境).....	227
A.16	Terms dependency in Domain Ontology Graph (經濟).....	228
A.17	Terms dependency in Domain Ontology Graph (軍事).....	229
A.18	Terms dependency in Domain Ontology Graph (醫療).....	230
A.19	Terms dependency in Domain Ontology Graph (電腦).....	231
A.20	Terms dependency in Domain Ontology Graph (體育).....	232

Part I

Introduction

Chapter 1

Computational Knowledge and Ontology

Abstract. The definition of knowledge is an augmentative topic in philosophy. We do not try to find out an explicit meaning of philosophical knowledge, but the most important thing is that we should know about what is knowledge in computer system, as called computational knowledge. In this chapter, we introduce some researches and definitions related to knowledge and computational knowledge. Ontology is a word used in both philosophy and computer system to describe the formalization of knowledge. We shall look into the definition of ontology in brief and also introduce its formalization methods in computer system.

1.1 What Is Knowledge?

“Knowledge” has been discussed by many philosophers since the Greek ancient times. It is not an easy task to find out a high abstraction of definition about “knowledge”. However, in very generally speaking, knowledge can be said as a meaningful resource that makes us know about the world. Theories of knowledge define what is about the world, how is it encoded, and in what way we reason about the world. Similar definition can be applied in computer and information system, unless we are defining those in the aim of computer processing instead of human understanding, and it is called computational knowledge.

1.2 Computational Knowledge in Information Systems

Computational knowledge in computer system has been represented as a hierarchy of data-information-knowledge in many knowledge management theories (Daft, 2004; Devenpart and Prusak 2000). Data refers to a string of bits, numbers or symbols that are only meaningful to a program. Data with meaning, such as computing words, texts, database records, etc, define information that is meaningful to human. Knowledge is the highest level of abstraction, which is encoded in some form inside information.

Creating computational knowledge is a study of artificial intelligent (AI) - an area of computer science focusing on making a computer to perform tasks with more intelligence (Genesereth and Nilsson 1987). Advanced information systems, such as information retrieval system, forecasting system, resource management system, online shopping system, personalization system, etc, always require computational knowledge to perform tasks with more intelligence. Traditional

information systems are lacking of intelligence because they process data and information without analyzing the knowledge behind. To enable a computer understand and process knowledge, we need to discover and represent the knowledge from raw data to a computable form for processing. Intelligent information system with the ability to process knowledge is so called a knowledge-based system.

1.2.1 Knowledge Engineering

Knowledge engineering grew out rapidly with the increased desire of knowledge-based system in the past decade. Knowledge engineering is a process to find out a way or approach to extract useful knowledge from computer data. It requires processes of analyzing and discovering patterns of data and transforming them to a format that is understandable to either human or computer, or both. Over the years, knowledge engineering researches have been focusing on the development of theories, methods, and software tools which aid human to acquire knowledge in computer. They use scientific and mathematical approaches to discover the knowledge. The approaches can be simply defined as an input-process-output system: Input – the set of computer data such as texts and database records; process – the method for the transformation of input data to knowledge; output – the desired knowledge in a specific form of knowledge representation (such as ontology).

1.2.2 Knowledge Representation

A general view of knowledge representation can be summarized in five basic principles (Randall et al. 1993):

1. A knowledge representation is a surrogate - a substitute of a thing (a physical object, event and relationship) itself for reasoning about the world.
2. A knowledge representation is a set of ontological commitments - an ontology describing existences, categories, or classification systems about an application domain.
3. A knowledge representation is a fragmentary theory of intelligent reasoning - a theory of representation that supports reasoning about the things in an application domain. An explicit axioms or computational logic may be defined for intelligent reasoning.
4. A knowledge representation is a medium for efficient computation - other than the knowledge represented logically. It also must be encoded in some sort of format, language, which enables a computer to process it efficiently.
5. A knowledge representation is a medium of human expression – a knowledge representation that can be understood by human. It is used by knowledge engineers or domain experts to study and verify the knowledge.

In the area of information system, knowledge representation defines a computable form of knowledge in Computer. It applies the theories and techniques from other fields including (Sowa 1999): 1. Logic, it defines a formal structure and rules; 2. Ontology, it defines the kinds of existence in a domain of interest; and 3.