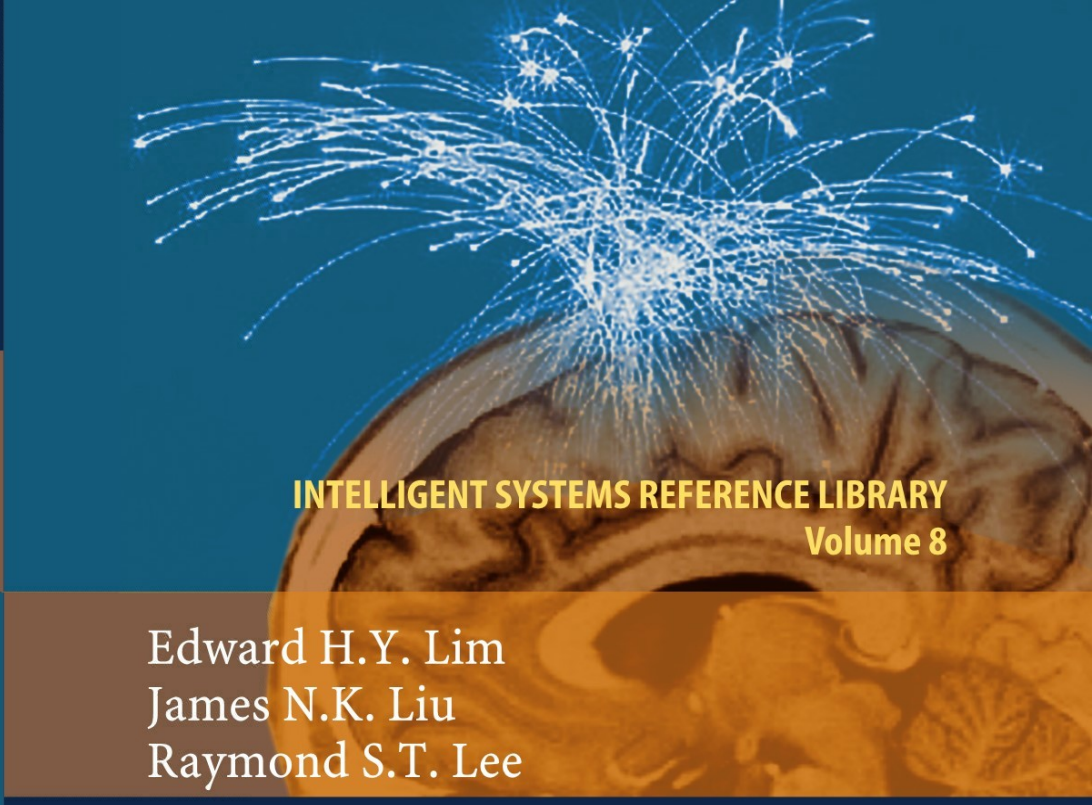

INTELLIGENT SYSTEMS REFERENCE LIBRARY
Volume 8

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

Knowledge Seeker – Ontology Modelling for Information Search and Management

A Compendium

Springer

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 Scheding
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

Springer

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

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 processeable 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. Liu[1], and Raymond S.T. Lee[2]

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

Contents

List of Figures

List of Tables

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

E.H.Y. Lim, J.N.K. Liu, and R.S.T. Lee: Knowledge Seeker, ISRL 8, pp. 3 – 12.
springerlink.com

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.