Ahlam F. Sawsaa

Ontological Engineering approach of developing Ontology of Information Science

Anchor Academic Publishing

disseminate knowledge

Sawsaa, Ahlam F.: Ontological Engineering approach of developing Ontology of Information Science, Hamburg, Anchor Academic Publishing 2015

Buch-ISBN: 978-3-95489-448-2
PDF-eBook-ISBN: 978-3-95489-948-7
Druck/Herstellung: Anchor Academic Publishing, Hamburg, 2015

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Bibliographical Information of the German National Library:
The German National Library lists this publication in the German National Bibliography. Detailed bibliographic data can be found at: http://dnb.d-nb.de

Hermannstal 119k, 22119 Hamburg
http://www.diplomica-verlag.de, Hamburg 2015
Printed in Germany

In the name of Allah, the Beneficent, The Merciful

(رَبَّنَا آتِنَا فِي الدُّنْيَا حَسَنَةً وَفِي الآخِرَةِ حَسَنَةً وَقِنَا عَذَابَ النَّارِ)

"Our Lord gives unto us in the world that which is good, and guard us from the doom of Fire."

(The Holy Quran, 2:201)

To my family

The near and far

Acknowledgments

First of all a special thank to my Lord Almighty Allah. We always keep the deficit in the description of words of thanks, especially to people who provide with tenderness without borders, the lines of thanks are always extremely difficult to formulate.

I would like to thank Professor Zhongyu (Joan) Lu, for her support, encouragement, and guidance from the initial to the final level, enabling me to develop an understanding of the subject, and her excellent feedback on the related paper and articles produced during the research project. Her encouragement has a big impact on my learning in finding pieces of knowledge and solving challenges.

I am heartily thankful to Julie Wilkinson and Violeta Holmes for their guidance and advice during these years.

Also, great thanks to all the staff of the Computing and Engineering School at the University of Huddersfield who inspired me to do my research. A big thank to everyone who participated in Ontocop's project.

Also, all the people were at my side; my sincere and faithful husband for his great and magnificent support, and the three essences of my life: Khalifa, Tasnim and Noor.

A great thanks also to all who gave me a helping hand to reach the end of my journey.

Abstract

Ontology has been a subject of many studies carried out in artificial intelligence (AI) and information system communities. Ontology has become an important component of the semantic web, covering a variety of knowledge domains. Although building domain ontologies still remains a big challenge with regard to its designing and implementation, there are still many areas that need to create ontologies. Information Science (IS) is one of these areas that need a unified ontology model to facilitate information access among the heterogeneous data resources and share a common understanding of the domain knowledge. The objective of this study is to develop a generic model of ontology that serves as a foundation of knowledge modelling for applications and aggregation with other ontologies to facilitate information exchanging between different systems. This model will be a metadata for a knowledge base system to be used in different purposes of interest, such as education applications to support educational needs for teachers and students and information system developers, and enhancing the index tool in libraries to facilitate access to information collections. This thesis describes the process of modelling the domain knowledge of Information Science IS.

The building process of the ontology of Information Science (OIS) is preceded by developing taxonomies and thesauruses of IS. This research adopts the Methontology to develop ontology of Information Science OIS. This choice of method relies on the research motivations and aims, with analysis of some development ontology methodologies and IEEE 1074-2006 standards for developing software project life cycle processes as criteria. The methodology mainly consisted of; specification, conceptualization, formalization, implementation, maintenance and evaluation. The knowledge model was formalized using Protégé to generate the ontology code. During the development process the model has been designed and evaluated.

This research presents the following contributions to the present state of the art on ontology construction;

- The main achievement of the study is in constructing a new model of Information Science ontology OIS. The OIS ontology is a generic model that contains only the key objects and associated attributes with relationships. The model has defined 706 concepts which will be widely used in Information Science applications. It provides the standard definitions for domain terms used in annotation databases for the domain terms, and avoids the consistency problems caused by various ontologies which will have the potential of development by different groups and institutions in the IS domain area.

- It provides a framework for analyzing the IS knowledge to obtain a classification based on facet classification. The ontology modelling approach is based on top-down and bottom–up. The top-down begins with an abstract of the domain view. While the bottom-up method starts with description of the domain to gain a hierarchal taxonomy.
- Designing Ontocop system a novel method presented to support the developing process as specific virtual community of IS. The Ontocop consists of a number of experts in the subject area around the world. Their feedback and assessment improve the ontology development during the creating process.

The findings of the research revealed that overall feedback from the IS community has been positive and that the model met the ontology quality criteria. It was appropriate to provide consistency and clear understanding of the subject area. OIS ontology unifies information science, which is composed of library science, computer science and archival science, by creating the theoretical base useful for further practical systems. Developing ontology of information science (OIS) is not an easy task, due to the complex nature of the field. It needs to be integrated with other ontologies such as social science, cognitive science, philosophy, law management and mathematics, to provide a basic knowledge for the semantic web and also to leverage information retrieval.

Publications

1. Sawsaa, A. & Lu, J (2009). A Generic Model of Knowledge Mapping Through Virtual Communities of Practice in Information Science (IS). Conference proceeding. World Congress in Computer Science, Computer Engineering, and Applied Computing, Las Vegas, Nevada.12-15 July 2009.
2. Sawsaa, A. & Lu, J (2010). Ontocop: A virtual community of practice to create ontology of Information science (IS). Conference proceeding. World Congress in Computer Science, Computer Engineering, and Applied Computing, Las Vegas, Nevada 12-15 July 2010.
3. Sawsaa, A. & Lu, J (2010). Ontology of Information Science (IS) based on OWL conference proceeding. The International Arab Conference on Information Technology (ACIT2010).
4. Sawsaa, A. (2010). A virtual community. The 3th Scientific Research Symposium for Libyan Students in UK Universities. Sheffield Hallam University, 12th June 2010.
5. Sawsaa, A. & Lu, J. (2011) 'Extracting Information Science concepts based on Jape Regular Expression'. In: WORLDCOMP'11 - The 2011 World Congress in Computer Science, Computer Engineering, and Applied Computing, 18-21 July 2011, Las Vegas, Nevada, USA
6. Sawsaa, A. & Lu, J. (2011) 'Virtual Community of Practice Ontocop: Towards a New Model of Information Science Ontology (OIS)' International Journal of Information Retrieval Research, 1 (2), pp. 55-78. ISSN 2155-6377
7. Sawsaa, A., ZHAOZONG, M. & LU, J. (2012) Using an Application of Mobile and Wireless Technology in Arabic Learning System. IN LU, Z. J. (Ed. Learning with Mobile Technologies, Handheld Devices and Smart Phones: Innovative Methods. USA, IGI Global. pp. 171-186. ISBN978-1-4666-0936-5
8. Sawsaa, A. & Lu, J. (2012) Developing a Domain Ontology of Information Science (OIS). IN SHONIREGUN, C. A. & AKMAYEVA, G. A. (Eds.) International Conference on Information Society (i-Society 2012) June 25-28, 2012,. London, UK, i-Society 2012 Technical Co-Sponsored by IEEE UK/RI Computer Chapter. pp 462-467.
9. Sawsaa, A. & Lu, J. (2012) Extracting Information Science Concepts based on JAPE Regular Expression. International Journal of advanced Computer Science (IJEC) , in press

10 Sawsaa, A. & Lu, J. (2012) Building an advance domain ontology model of Information Science (OIS). Journal of the American Society for Information Science and Technology (JASIST), in press

11 Sawsaa, A. & Lu, J. (2012) Building Information Science (OIS) Ontology with Methondology and Protégé. Journal of Internet Technology and Secured Transactions (JITST) 2 (1/2), ISSN 2046-3723. In press.

12 Sawsaa, A. & Lu, J. (2013) Extracting Occupational Therapist concepts to develop domain ontology. The Seventh International Conference on Digital Society (ICDS 2013) February 24 - March 1, 2013 - Nice, France.

Posters:

---

1 Sawsaa, A. and Lu, J. (2011) 'ONTOCOP: Virtual Community Of Practice to build Ontology of Information Science'. *University of Huddersfield 2009*

2 Sawsaa, A. and Lu, J. (2009) A Generic Model of Knowledge Mapping Through Virtual Communities of Practice in Information Science (IS)

Table of contents

List of Figures

List of Tables

Acronyms

AI	......Artificial Intelligence
ANNE	......A Nearly New Information extraction System
CLIPS	......C Language Integrated Production System
COC	......Community of Commitment
CoI	......Community of Interest
Cops	......Communities of practice
DARPA	......Defence Advanced Research Projects Agency
GATE	......General Architecture for Text Engineering
IE	......Information Extracting
IS	......Information Science
JAPE	......Java Annotation Patterns Engine
KE	......Knowledge Engineering
KR	......Knowledge Representation
LOOM	......Language for Object Oriented Methods
Nop	......Network of practice
OCML	......Operation Conceptual modelling Language
OIS	......Ontology of Information Science
Ontocop	......Community of Information Science website
OWL	......Web Ontology Language
RDF	......Resource Description Framework
SGML	......Standard Generalized Mark-up Language
VCops	......Virtual Community of practice

VCs ....Virtual community

W3C ....World Wide Web Consortium

XML ....Extensible Mark-up Language

Part 1: Fundamental Issues

1 Chapter 1: Introduction

Recently, the development of domain ontologies has become increasingly important for knowledge level interoperation and information integration. They provide functional features for AI and knowledge representation. Domain Ontology is a central foundation of growth for the semantic web that provides a general knowledge for correspondence and communication among heterogeneous systems. Particularly with a rise of ontology in the artificial intelligence (AI) domain, it can be seen as an almost inevitable development in computer science and AI in general.

Ontologies are useful for different applications to be able to share information between heterogeneous data resources. They are also essential for enabling knowledge-level interoperation of agents, when these agents are interacting to share a common interpretation of the vocabulary. Moreover, it is useful for human understanding and interaction to reach a consensus amongst a professional community.

Although there are a range of domain ontologies on the semantic web such as Gene Ontology (GeneOntology, 2009), Biological science ontology (Sabou 2005), CIDOC-CRM ontology of culture heritage documentation, FRBR in Bibliographic and NCI cancer ontology (Golbeck et al., 2008), there still exists a lack of domain ontologies, which has led to the loss of knowledge in specific domains. This is a significant problem for scholars and researchers who need to be able to access information within their interest area.

Ontology provides a vocabulary for metadata description with machine understandable terminology. Ontology provides a format for explaining and understanding terminology and the knowledge contained in a software system. By using shared concepts and terms in accordance with a specific approach, a lot of information remains in people's heads. It is discussed in 2.3.

However, information science (IS) is a fast paced discipline and communication technology is rapidly increasing, so it is imperative to take advantage of this development. IS is a multidisciplinary field and it has gained the fundamental root of its theory from different related fields. The analysis includes the three branches of the field, which are; Library Science, Archival Science and Computer Science. Meanwhile it overlaps with other sciences, as stated in Section 2.2, e.g., communication, cognitive science, philosophical science, management, social science and marketing. More precisely, the relationships between information and marketing can be subdivided into marketing information, marketing information services, marketing of library services.

These kinds of relationships need logical ontology to clarify their relations and the science boundaries, amongst others. Therefore, Information Science still needs identity.

However, there is a lack of IS ontology representing the unified model that combines all concepts and their relationships. Moreover, IS as any domains which use the natural language. It contains a lot of jargon which needs to be in a formal language for programming or logic. Alternatively, integration of the computer with the internet has led to the emergence of new concepts in the field of IS such as , Electronic Library, Virtual Library, Library Without Walls, Digital Library and Information Management, as well as Nerve Centres. Even the information concept itself has strong and complex relations with other concepts, for example some people have defined it as fact, energy, data, and symbols. Also, it can be composed with other words such as; information age, information revaluation, information crisis, information explosion. However, there are 400 definitions for information in the literature (Yuexiao, 1988). It is hard to differentiate between these concepts. Even within the same field, there is still confusion over defining information - everyone defines it based on his background, for example librarians know it in term of facts, and data can be in containers such as journals, books and documents. The computer scientists conceive it as small units such as bits and bytes.

Consequently, modelling the IS domain necessarily assumes the need to represent the correct picture of the whole domain, and any changes in the domain will have to be added to keep the model up to date (Mommers, 2010, Yuexiao, 1988).

Our consideration is that in developing an ontology of Information science OIS to define its boundaries, and avoid ambiguous concepts.

Therefore, there is a lack of unified model of domain knowledge, because of the inconsistency in structure of domain which led to difficulty of using and sharing data in syntax and semantic level.

1.1. Problem Identification

Information Science is seeking its identity and it is one of the many domains which use natural language including much jargon. Also, integration of the computer with the internet has led to emerging concepts in the field of IS such as , Electronic Library, Virtual Library, Library without walls , digital Library It is hard to differentiate between them.

Furthermore, its structure led to lack of a unified model of domain knowledge. This led to lack of a unified model of domain knowledge, and difficulty of using and sharing data at syntax and semantic levels. The OIS ontology provides a standard terminology and shared representation of domain concepts.

Therefore, the ontology of information science is missing in ontological engineering area. Our consideration is that developing ontology of Information science to define its boundaries, and to avoid the concepts ambiguous.

The research problem of the study was defined as the following:

> **Q. How an ontology of Information Science (OIS) model can be developed to visualise the IS domain, and how the model could capture and represent this knowledge?**

To achieve the primary objective, the researcher asks questions to be answered through this study such as:

- What domain knowledge does the ontology represent?
- What is the level of knowledge that the ontology will represent?
- Which knowledge representation techniques and languages should be used?
- What are the relations that will be used to structure the knowledge, and which structure for the ontology will it have e.g. tree, graph, and its main components of ontology (e.g., classes, instances, relations, rules)?
- What is the value of tools such virtual community of practice ontocop? Could they be valuable in supporting the developing process?
- Does the developing process of the ontology follow designing criteria?
- Is the ontology evaluated based on specific criteria?

1.2. Aims and Objectives

The aim of this research is to develop a generic model of ontology that visualize domain knowledge of IS that serves as a foundation of knowledge modelling for applications and aggregation with other ontologies.

The visualisation stage provides an extensible and commonly understood semantic framework by describing the terminology of the domain. Achieving this aim in the current study will fulfil the following Objectives:

- Building a conceptual model for establishing a better analysis framework to understand, classify and compare various classes of Information Science.