University of Huddersfield Repository

Sawsaa, Ahlam and Lu, Joan

Building an Advance Domain Ontology Model of Information Science (OIS)

Original Citation


This version is available at http://eprints.hud.ac.uk/21161/

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

http://eprints.hud.ac.uk/
Building an Advance Domain Ontology Model of Information Science (OIS)

Ahlam F. Sawsaa & Joan Lu
School of Computing & Engineering
University of Huddersfield -UK
swsa2004@yahoo.com & J.lu@hud.ac.uk

ABSTRACT

The paper describes the process of modelling domain knowledge of Information Science (IS) by creating an Ontology of Information Science domain (OIS). It also reports on the life cycle of the ontology building process using Methontology, based on the IEEE standard for development software life cycle process, which mainly consists of: specification, conceptualization, formalization, implementation, maintenance and evaluation. The information resource used in acquisition and evaluation has been obtained from Information Science. The conceptualization consists of identifying IS concepts and grouping them into a hierarchy tree based on a faceted classification scheme. The OIS ontology is formalized by using the ontology editor Protégé to generate the ontology code. The achieved result is OIS ontology which has fourteen facets: Actors, Method, Practice, Studies, Mediator, Kinds, Domains, Resources, Legislation, Philosophy & Theories, Societal, Tool, Time and Space. The model is evaluated using ontology quality criteria to check the ontology’s usefulness, and how it could be transferred into application ontology for Information Science education.

KEYWORDS
Ontology - Knowledge Representation- Semantic web - Information Science- Web Ontology Language-Protégé.

1. INTRODUCTION

In recent years ontology has gained attention in both academic and industrial fields. The word ontology has been defined in different ways, originally taken from philosophy, where it means the basic characteristics of existence in the world. Ontology is applied in various domains such as medicine, movies, cooking, and management, to provide a formal model that structures knowledge. The Information Sciences IS domain appeared as an interesting research area due to the fact that it is a multidisciplinary science emerging from Library science, documentation and computer science. Inconsistency in the structure of the IS domain led to the lack of a unified model of domain knowledge. This lack makes data at syntax and semantic level difficult to use and share.

Many technologies offer a good solution for data sharing at syntax level, instance XML, but it cannot solve. Ontology offers a good solution for data use and sharing at semantic level. Ontology is a moulding tool that provides a formal description of concepts and their relations as a foundation for semantic integration and interoperability.

The lack of domain ontologies in computer-based applications has led to loss of knowledge in specific domains. In this sense, the problem is vital for scholars and researchers, who need to access information in efficient ways to meet their interests. The problem has been defined to as requiring an artifact for a solution. Ontologies can lead to solutions to this problem due to the fact that they give some sort of notion of meaning about terms. It has the potential to overcome the problem and make the conceptualization of domain Information Science explicit and understandable.

Information Science IS as an interdisciplinary science needs to be defined. However, it became necessary to develop OIS ontology to represent the domain knowledge. The ontology of Information science is discussed in this paper.

The goal of the paper is to study the terminology of information science to create domain ontology. Many ontologies have been created and published. However, the OIS ontology is missing. OIS ontology is a new research direction in the IS field. This study is devoted to clarifying the basic concepts and framework in the IS field. This study is devoted to clarifying the basic concepts and framework in the IS field. It presents a formal semantic explanation for IS meta-data. This
paper is organized as follows: in section 2 we discuss the theoretical foundation of ontology. In section 3 we discuss the method of building IS ontology and how it has been constructed. Section 4 presents ontology of Information Science development and implementation, followed by discussion and evaluation in section 5. Finally, the conclusion and future work will be presented.

2. BACKGROUND

2.1. Ontology Definition

Ontology has different definitions in available literature [1,2,3]. Basically, ontologies are used in different communities. Ontology emerged from the philosophical field as an area of study introduced by Aristotle. In recent years this term has borrowed from computer science community uses to represent the knowledge required to understand the real world. Developers have been developing a conceptual base for building technology to construct knowledge components to be reusable and sharable.

So, ontology has been defined from different perspectives. The philosophical perspective defined ontology as the science or study of being[4], while the Artificial Intelligence (AI) community defined it in 1991:

“Ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary.” [5]. This definition is a brief definition, which indicates that ontology is providing definitions of terms that are explicitly defined and the relations and rules to unite them, but ontology is more than that. It can provide inferred new terms using the rules. In 1993 Gruber defined ontology as: “An ontology is a specification of a conceptualization.” [6]. His definition has been developed to be more accurate for defining ontology which is: “Formal explicit specification of shared conceptualization.” The definition can be explained as follows:

A formal: ontology should be machine readable and processed by an Artificial Intelligent (AI) system. We do not need there to be communication devices between people and people even people and machine. Ontology should be formally defined using a formal language [7].

Specification: means writing specifications of language syntax to satisfy certain criteria such as precision, non-unambiguity, consistency, completeness and implementation as independent statements [8]; it should provide a communication device to enable users to share knowledge in consensual mode.

Shared: means ontology represents a consensual knowledge that has been arranged and agreed on by groups typically as the result of a social network rather than an individual’s view.

Conceptualisation: this is an abstract model of domain knowledge driven by application for users, and represents ways in which it is committed by knowledge-based systems.

We can formulate definitions that we can understand based on the above; ontology should be formally defined to process by machine. The ontology is a specific type of information object or artifact. The ontology construction refers to clear classes, relations and their instances, which play roles of explicit specification of conceptualisation. In other words, the back bone of the ontology is composed of the specification of concepts.

However, ontology is not software and it cannot run as a program, but it can be used by programs. A far more interesting question is what information systems could learn from philosophical ontology. It is a shared belief there is a similarity inherent in ontology from philosophical and applied scientific perspectives. Philosophical ontology is describing the real world as it exists, while computational ontology is describing the world as it should be [9].

According to Gruber’s definition(1995) OIS ontology is the formal explanation of shared conceptualization of the IS domain. The concepts in IS are represented by the ontology model. It is more interesting that the IS knowledge will be conceptualized by defining classes and certain relationships, to make it machine readable [10]. In this paper we focus on the conceptual ontology
that is being used in the semantic web. The aim of this study is:
1. Providing a visualisation of the Information Science area.
2. Sharing a common understanding of Information Science theory.

2.2. Domain Ontologies
The number of studies of ontology has been growing rapidly in recent years. Gartner points out that integration of the semantic web could be the greatest impact on the technologies in the next few years. Ontology is used as a basis for enabling interoperability through the semantic web [12]. Bhatt shows an approach of extracting sub-ontology to meet the user needs, based on the unified medical language system (UMLS), by designing onto Move to exploit the semantic web. It used language of RDFs and OWL [13]. Onto CAPE is a large scale ontology for chemical processes for use in an industrial field [14]. Du et al have proposed onto Spider which is a novel ontology for extracting ontology from the HTML web; nevertheless, the lexical semantics and natural language have a negative effect on the result due to a difference of outcome when words or links are missed [15], [16], [17].

Domain ontology plays a vital role by defining terms which could be used as meta-data. Sabou’s work is about creating ontology from an OWL-s file for describing a web service [18], particularly in a specific domain such as biomedical ontologies which play a fundamental role in accessing the heterogeneous sources of medical information, and using and sharing patients’ data. GALEN (Generalised Architecture for Languages Encyclopaedias and Nomenclatures) provides reusable terminology resources for clinical systems. It contains 25,000 concepts used to represent a complex structure of descriptions of medical procedures [19]. Furthermore, GENE ontology (GO) was developed by the National Human Genome Research Institute in 1998. It presents a control vocabulary of gene and gene products attributes. It contains (30,000) concepts and is organized as follows; biological process, molecular function, and cellular component. The GO ontology is regularly updated and it is available in several formats [20], [21], [22].

Standardized Nomenclature for Medicine-clinical terminology (SNOMED) is ontology for health care terminology. It contains 350,000 terms that represent clinical meanings. Each concept has a number, ID and full specific name (FSN). SNOMED has the ability to automate functions related to medical record administration and to facilitate data collection for research purposes (Jepsen, 2009). Toronto Virtual Enterprise (TOVE) was developed in the Integration Laboratory at University of Toronto. It provides a shared terminology to be understood and shared between commercial and public enterprise. TOVE is implemented in C++ and Prolog for axiom. It covers activities, time, parts and resources [23].

Economic ontology is constructed to define the economic domain from economic documents. It uses the OntGen tool to semi-automatically construct ontology. The ontology is based on machine learning methods [24].

2.3. Methodology Employed
2.3.1. Theoretical Bases
The nature of the ontology is a concept model. The concept model represents the relationship of concepts within the domain; to gain a better understanding of OIS ontology development and its role in the semantic web, the framework is established to describe the main theoretical base. The methodology is based on Category Theory, which is foundation theory for mathematics. A number of thinkers such as Hartmann and Husserl asserted that ontology relies on category theory.[25], [26]. Thus, ontology is a model of a domain,

\[ O = \{C, R, A\}', \text{ Where } C \text{ is a concept} \]

\[ R \subseteq C \times C, \text{ Where } R \text{ is relations. For } r = (C1, C2) \in R. \text{ Or } R(C1) = (C2) \]
Whereas, Concepts: set of entities within a domain. Relations: interactions between concepts in the domain. Axioms: explicit rules to constrain the use of concepts. Instances: concrete examples of concepts in the domain [28,29,30].

2.3.2. Techniques and Tools

2.3.2.1. Web Ontology Language (OWL)
OWL is designed to represent information about objects and how these objects are organised and interrelated within specific domains. OWL is derived from descriptive logic that aims to bring reasoning and expressive power to the semantic web.

2.3.2.2. Protégé
Protégé is an open source tool that was developed at Stanford University by Stanford Medical Informatics. The core of Protégé is an ontology editor, which provides a suite of tools to construct domain models using various formats. It can also be extended by using plug-ins to add more functions such as import and export ontology language (XML, OIL, FLogic). The platform of Protégé is supporting two ways of modelling ontologies.

Building Information Science ontology OIS follows Methontology based on IEEE standard criteria to design an ontology life cycle process. The IEEE 1074-2006 is a standard for developing a software project life cycle process [31],[32], using methodology to capture the domain knowledge and to establish the creation of the glossary of domain knowledge [33].

Methontology is a chosen methodology to develop the Information Science ontology OIS. This methodology uses an iterative approach which allows us to refine the ontology to create a more accurate model of the IS domain. The OIS ontology methodology is constructed as follows:

1. **Building the conceptual model** which is established from:

   a- Determining the domain’s scope, interest, goals, strategy and boundary, which need to answer the following questions:
   - Q1. What are the general characteristics of ontology of IS?
   - Q2. What is the purpose of ontology of IS?
   - Q3. Will it cover the general domain or specific?
   - Q4. What about its size and formalism used?
   - Q5. Does it use formal axioms or order logic?

   To answer these questions we should describe the contents of the ontology. These contents includes: taxonomic structure, concepts it will covered, top-level division, and the internal structure of the concept.

   b. Acquiring domain knowledge and developing the glossary that contains the key concepts in the field. It requires the integration of all relevant terms, which include concepts, instances, attributes, relations

   c. Create concepts dictionary to identify the terminological concepts and relations

   d. Modelling concepts in a hierarchical taxonomy and their relations( subclasses, super classes)

2. **Convert the conceptual model to Computational model**, which starts from:

   b. Formalising Ontology by ontology Protégé editor.

   c. Evaluation and maintenance of the computational model.

   d. Documentation of the ontology life cycle, is shown diagrammatically in Figure 1.
3. IMPLEMENTATION

3.1. Conceptual Model of OIS Ontology
The conceptual model reflects on the computational model. It could be a communication device for experts in the domain. It shows the entity classes, attributes and their relationships with OIS ontology. We develop the main relationships among the defined classes. The conceptual model is a base of domain ontology which helps to build the OIS ontology. Figure 2 shows the conceptual model of domain ontology.

3.2. Computational Model of OIS Ontology
OIS ontology is structured in natural language to be suitable for data modelling and knowledge representation. It is intended to express the unambiguous and complete specification of domain concepts. It provides a dictionary of concepts with relations between them and organises them in super-types and sub-types of hierarchy.
OIS ontology is encoded by the Protégé editor to formalize the OIS, due to the fact that Protégé provides plug-ins and play environments for developing prototypes and applications. Furthermore, ontology in Protégé can be exported to different formats, as seen in List 1.

List (1) OIS ontology in OWL.

The root class in OWL is thing (owl: Thing) which is the root of all classes such as Resources in RDF ( rdfs: resources) The list below displays a simple hierarchy of the main classes of OIS ontology by OWL language, as shown in diagram 2.
The Upper-level of classes was created based on a taxonomy of IS. The OIS ontology is basically organized into several classes that correspond to different kinds of things that describe the science. The first layer is a meta-class level has concepts; Actors, Domains, Kinds, Practice, Studies, Mediator, Method, Resources, Legislation, Philosophy & Theories, Tools, Societal, Time & Space as shown in Figure 4. Each sub class is grouped under main upper-class such as “Education of Information science”, “Education of Computer Science”, “Education of Library Science”, all grouped under the Education class.

Furthermore, the current version is defined by a large number of classes - about 687 - and consists of approximately 170 assertions including more than 67 rules and relations to determine the rich semantic expression capability of the language.

4. EVALUATION
Ontology evaluation means taking into consideration that which guarantees the stability and accuracy of the ontology. Evaluation of the ontology avoids concept duplication, excessiveness and inconsistent relationships to create a better understanding. In this study the evaluation process is based on interim and completion evaluation. The evaluation is used at development stage to improve the design and implementation of the project. The OIS ontology was evaluated during the development process to ensure its completeness and consistency of meaning.

The OIS was evaluated by the domain’s experts to identify their level of satisfaction, based on predefined criteria. The first criterion was ontology consistency. (64%) of respondents indicated level 3 of satisfaction, and others expressed levels 2 and 4 by (20%,12%) respectively. The second criterion was consistency of is-a and part-of relationships. (14) of the participants indicated their satisfaction with the consistency of ontology relations at level 3 (56%) while 6 of them (24%) pointed to level 2. For the third criterion the majority of participants identified level 3 to indicate their level of satisfaction with assessing completeness of OIS ontology which is (48%), in comparison with levels 1 and 5.

The fourth criterion was clarity of OIS ontology, as illustrated in Table 1.
The FaCT++ was implemented to ensure errors free that the in ontology syntax was errors-free. If there is a class incorrectly classified it will appear in red colour in a root class called Nothing, e.g. the class Analytics, ArchitectureLibrary, Dissemination, and DocumentationCenter appears inconsistent in the class category. They appeared as main classes that were organised under the main root (Nothing). The subclass of ElectronicDocumentdelivery and Information Diffusion are classified by the reasoner under the Domain while they are a subclass of Information Service that is structured under the Practice class.

It also ensures there are no confounding and contradictory concepts, and ensuring that terms have consistency of meaning with clarity. Ontology should provide mapping according to the meaning of its contents. However, the consistency and the syntax of the generated OWL file can be verified by using an OWL ontology validator. The OIS ontology was verified by using OWL validation as well, for more testing and validation. Once the ontology was uploaded to the validator, the abstract syntax –Full OWL - form says Yes:Why, which means the ontology has succeeded and the results are good. Figure 3 shows a segment of the verification results.

5. DISCUSSION AND ANALYSIS
The OIS ontology that represents the domain knowledge is introduced in this paper. It enables us to understand the domain knowledge and the conceptual relationship between its branches. The theoretical base of the model is based on a faceted analytic-synthetic system. The model is structured around the domain conceptualization based on Methontology. The OIS is structured from 14 meta-classes that are based on fact classification, to define the key elements of the domain and possibly to be linked with other domains. This structure can be used for structuring IS, organising the sub-classes in the domain. For example, the meta-class (Mediator) structures all types of mediator in the IS domain such as: Archives, Libraries, media Centres, Documentation centres, Information Centres, Museums and Websites. Meanwhile the class (Library) could be extended to offer the following list of sub-classes; such as Library – Academic, University library, College library, Higher education institutions, Department library, Library – International, Library – School.

<table>
<thead>
<tr>
<th>Ontology Criteria</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent of ontology</td>
<td>0.64%</td>
</tr>
<tr>
<td>Consistency of is_a and part_of_relationships</td>
<td>0.56%</td>
</tr>
<tr>
<td>Completeness</td>
<td>0.48%</td>
</tr>
<tr>
<td>Clarity</td>
<td>0.40%</td>
</tr>
<tr>
<td>Generality</td>
<td>0.44%</td>
</tr>
<tr>
<td>Semantic data richness</td>
<td>0.48%</td>
</tr>
</tbody>
</table>
The current version of OIS ontology contains 687 classes - about 44% - and 700 subclasses; about 45%.

In addition, we note that classes and subclasses feature in the OIS more than other components such as data property that is 1%, object property which is 4% while individuals make up 6%. This is because this model is a generic model that structures the IS domain as the base of application ontologies that will be developed. The model has data properties that indicate the semantic relations between classes and subclasses. The model has different relationships (object property) such as, hierarchical relationships (isPartOf, IsA, ), inverse relationships ( hasA, hasPart), equivalency and associative relationships. These relationships are representing the core relations between the concepts. In comparison to other relations the hierarchical relations were used more than their functional equivalent, while the transitive relations were used less than others. We describe some classes of the OIS ontology to clarify some of these relationships.

6. CONCLUSION

6.1. Achievements

The following are the main achievements presented through this study:
- Constructing an Ontology of Information Science (OIS), and methods of building it.
- Demonstrating the strategy of building and designing the conceptual model in the domain using ontology technique. Ontology of information science will help to identify the features of this science, which mainly consist of the overlapping sets of science that make it difficult to determine its boundaries.
- The resulting ontology covers three main areas of the domain knowledge: library science, archival science and computing science. The vocabularies of these branches are formalized in class hierarchy with relations which are interconnecting concepts from all these areas, in order to define a sufficient model of the Information Science domain.
- The phases of the methodology were specification, knowledge acquisition, conceptualization, formalization, and evaluation, of all which are essential in order to attain the results. The domain knowledge was formalized by using the Protégé ontology editor, which can also be used to automatically generate the ontology code. The ontology was evaluated and validated by using FaCT++ reasoner. The evaluation report was used to check if the ontology was consistent and satisfied needs.

6.2. Future Works

The reusing, sharing, and maintaining of the ontology for future issues that relate to our ontology need to be considered. In the OIS module there is always space for improvement, at least adding the additional of new or missing concepts and adding new classifications based on different criteria and perspectives. Although most Information Science concepts were considered, there are concepts that need to be added. Another, more interesting, possibility would be to link this general model with others that are related to the domain in order to be integrated with other ontologies in an ontology library to use for specific applications.

7. REFERENCES

3. STARLAB, F. Systems Technology and application Research Laboratory home page . Faculty of Science, Department of Computer Science , Vrije University Brussel. (2003).
7. MORBACH, J., WIESNER, A. & MARQUARDT, W. OntoCAPEis a large –scale ontology for
27. GÓMEZ-PÉREZ, A., CORCHO, O. & FERNÁNDEZ-LÓPEZ, M. Methodologies, tools and languages for building ontologies. Where is their meeting point? Data & Knowledge Engineering, 46, 41-64. (2003)