Building Information Science ontology (OIS) with 
Methontology and Protégé

Ahlam F. Sawsaa
School of Computing & Engineering
University of Huddersfield
Huddersfield UK

Joan Lu
School of Computing & Engineering
University of Huddersfield
Huddersfield UK

Abstract

Ontology is the backbone of the semantic web and can overcome semantic barriers. Domain ontology provides a common understanding of the knowledge of a particular domain. Information Science, meanwhile, is an interdisciplinary science that is yet to be defined. It is necessary to develop Ontology of Information Science (OIS) to represent the unified domain knowledge.

This paper presents a representation of specific domain knowledge by providing a definition, scope, and boundaries of Information Science (IS). The methodology followed is Methontology, which is based on the IEEE standard for the development of a software life-cycle process using the ontology editor Protégé. The OIS ontology has fourteen facets: actors, method, practice, studies, mediator, kinds, domains, resources, legislation, philosophy and theories, societal, tools, time and space. It provides a broader base of classes to offer the opportunity to enrich OIS ontology. Also, it could be a basis for multiple ontologies to be built. OIS is structured at class level and subclass but does not provide the individual level. The model was evaluated by domain experts based on specific criteria and using the FaCT++ reasoner to check the ontology usefulness, and how it could be transferred into application ontology for Information Science education. The paper then discusses the OIS ontology, particularly its structure and development.

Keywords: ontology, ontology engineering, knowledge visualization, knowledge representation, semantic web, information science

1. Introduction

In recent years, ontology has received attention from both academic and industrial fields. The word ontology has been defined from different perspectives, having originated in the field of philosophy, where it is used to mean the basic characteristics of existence in the world. Information science is a multidiscipline consisting of a number of different branches, including library science, computer science, and archival science. Thus, it lacks a unified model of domain knowledge.

Inconsistencies in the structure of the domain have led to difficulties in using and sharing data at the syntactic and semantic levels. Many technologies offer good data-sharing solutions for the syntactic level, for example XML, but do not work effectively at the semantic level. Ontology offers a good solution for using data and sharing it at the semantic level. Ontology is a modeling tool that provides a formal description of concepts and their relations, as a foundation for semantic integration and interoperability [1; 2].

In this paper, we focus on the ontology used in the semantic web as a compatible independent model. The aims of this study are as follows:

- to provide a visualization of the IS area
- to share a common understanding of IS theory
- to describe the terminology and a conceptual model of IS, including concepts, examples of them, and the relationships between them [17], in other words a logical model of OIS.

The Ontology of Information Science OIS is a new research direction within the field of Information Science (IS). It provides a formal semantic explanation for IS data. This paper is organized as follows: In section 2, we discuss the theoretical foundations of ontology. In section 3, we describe the method used to build the IS ontology. In section 4, we present its development and implementation. Section 5 contains our evaluation and discussion of the results. Section 6 concludes and presents suggestions for future work.

2. Background

The World Wide Web (www) is enormous and the semantic web is still in an early phase. The semantic web needs semantic interoperability between metadata connected with web information. The www needs smart tools to improve information retrieval and to integrate all the information that users need. So, metadata makes it easier for search engines to get web pages.

Nowadays, many sources of data and information are available on the Internet, and therefore it has
become imperative for computer scientists and IA community to facilitate information access and improve information retrieval on the Internet, by using techniques, applications and programming languages such as: Extensible Markup Language (XML), Resource Describe Framework (RDF, RDFs) and Web Ontology Language (OWL). The semantic Web provides a good opportunity for researchers to access the required information. Particularly with an increase in information on the Internet, it can be said that the Semantic Web has revolutionized the world of search browsers on the internet.

The semantic web is defined as “an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.” [3P580]. Its aim is to convert a large amount of data and information resources that are available on the internet rather than just units consisting of (0) (bits) to understandable data by the computer programs using specific languages mentioned above.

Tim Berners-Lee, in his article describing the semantic web, said it is an attempt to develop languages that express information in a form accessible to human understanding. This brings to mind to ask this question: what is the importance of the semantic web? Berners–Lee is correct that Google is a wonderful tool for humans, but it does not serve machines, it is only understandable for humans; we also need these pages and their contents to be understandable for machines [4].

The semantic web offers semantic annotations that describe web resources explicitly. These annotations are based on ontologies that represent domain knowledge through defining concepts and the semantic relations between those concepts. Besides it provides a machine processable representation of ontology. The standards for this purpose have been defined by W3C, such as RDF, OWL. Ontology is a foundation and is central to the growth of the semantic web that provides a common knowledge for correspondence and communication among heterogeneous systems. Furthermore, it is useful for different applications to share information among heterogeneous data resources [5].

Therefore, an ontological infrastructure to support the semantic web has been developed. Particularly with a rise of ontology in the artificial intelligence domain, it can be seen as an almost inevitable development in computer science and AI in general. Ontology plays an important role to use as a source of shared defined terms; such metadata can be used in a specific domain.

A. Ontology of Information Science (OIS)

Ontology of IS facilitates data exchanging information integration, and search of IS data. Nowadays, ontologies have become mainstream in several domains. Ontology has been defined from different perspectives, computer science and philosophy and has various definitions in the literature [6; 7; 8; 9]. The philosophical perspective defines ontology as the science or study of being, what is existing, introduced by Aristotle [10; 11].

The term has been borrowed by computer science and is used to represent knowledge or understanding of the world. The Artificial Intelligence AI community defines the term as “a formal explicit specification of shared conceptualization” [12]. It is defined as a specification of conceptualization. Ontology represents knowledge of a specific domain in the way that concepts are defined in unique manner and connected with relationships.

According to Gruber's definition, OIS is the formal explanation of a shared conceptualization of the domain of IS. That is, the concepts of IS are represented by the ontology model. More interestingly, IS knowledge is conceptualized into defined classes and relationships to make it machine readable.

The OIS ontology has developed to overlap the field problematically. Information science IS has interdisciplinary relationships with different sciences and it needs to be determined how it will be defined.

Furthermore, IS still seeking to identify its identity and boundaries against others fields, because a lack of scientific methodology and philosophy led to big problems, particularly when information scientists attempted to establish the basic area of the science [13]. IS is concerned with collecting and organizing information resources to be retrieved by users in information centers and libraries [1].

B. Theoretical foundation of OIS ontology

Theories can help to define formal ontological properties that contribute to characterizing the concepts. Meanwhile, ontologists nowadays have a choice of formal frameworks which derive from formal logic, as well as algebra, category theory, Mereology, set theory and Topology. The methodology is based on Category Theory. In the interim, ontology adopts a categorical framework that means it searches for what is universal in both specific and general domains [14]. The ontology’s
meaning emerges basically from a reliance on the 
theory of category as a grounding of mathematics[15].

Bill Lawvere claims that category theory is 
foundation theory for mathematics, which is based on 
knowing what the universal in the world is. In the 
mean time, ontology seeks to look for what is the 
universal[15]. Category theory is a branch of 
mathematics which is used in the field of computer 
programming, software engineering, and semantic, 
Artificial Intelligence and knowledge management, 
also using its mathematic base for ontology structure 
[16]. It was introduced by Eilenberg and MacLane in 
1945. Category theory is the most famous theory, 
even though it was very simple for axiom set. [17; 
18] This is designed to describe a range of structural 
concepts in a consistent way from different 
mathematical grounds[19].

The category theory is described by Eilenberg and 
MacLane, 1945 as follows “is a powerful language, 
or conceptual framework, revealing the universal 
components of a family of structures of a given kind, 
and how structures of different kinds are interrelated” 
[20] Category theory indicates that the class object is 
part of the universal.[21].

Ontology is by nature a concept model, which means 
that it represents the relationships between concepts 
within the domain. According to the characteristics of 
IS, it can be defined as follows:

\[ OIS = \{ C, R, A, X, I \} \]

where:

- \( C \) is concepts of IS objects
- \( R \) is the relationships between those concepts
- \( A \) is the attributes of the IS object
- \( X \) is the axioms of the concepts [20]

To build the OIS it follows the process of Methontology, which is based on the IEEE standard 
criteria, to design the ontology life-cycle process. The 
IEEE 1074-2006 is a set of standards for developing a 
software project life-cycle process [21].

![Figure 1 IEEE 1074-2006 Software development life-cycle process [22]](image)

3. Methodology

The Methontology methods developed in 
Madrid in a Artificial Intelligence laboratory a the 
Polytechnic University are used to create and develop 
ontologies. Methontology could be followed to create 
ontologies from scratch or reuse an existing one. It is 
proposed by Fernandz in 1997, consisting of many 
stages to construct the ontologies at the knowledge 
level [23].

Methontology provides guides for constructing 
ontology in specific domains through the 
specification, conceptualization, formalization, 
implementation, evaluation, refinement, maintenance 
and documentation; the next section describes these 
activities in a brief manner:

**Specification** phase is to identify the goal of 
developing ontology, for which purpose it will be 
built, and its scope of coverage knowledge. Does it 
cover the whole domain or is part of domain to be 
applied for a specific reason? Also, knowledge is 
acquired in this phase to develop the conceptual 
model to construct the knowledge in an informal way.

**Conceptualization** activity consists of:
- Creating glossary of domain terms to 
determine which terms are included in the 
ontology, in the glossary each term has a 
description, acronyms, and synonyms.
- Identifying the binary relations among 
classes of the ontology.
- Build the concept classification tree to 
indicate support-classes and sub-classes.
- Create concepts dictionary containing 
meanings of concepts, and description of 
instances, and attributes.

**Formalization** activity is transforming the conceptual 
model into a formal mode by using ontology editors.
such as Protégé to represent the model in a formal language such as Web Ontology Language (OWL).

**Evaluation activity** ontology needs to be evaluated to ensure clarity and comprehensiveness. Also, evaluation stage guarantees completeness of information that the ontology covered, and that each class is defined correctly in consistent manner.

Documentation is a very essential stage facilitating the reusability of the ontology that has been designed. Furthermore, once the ontology has been designed, it needs to be maintained and up-to-date within the Refinement and maintenance stage.

### 3.1 Methods Employed

Our task is to capture the domain knowledge and create a glossary of that knowledge as well as a concepts dictionary. We do this using Methontology, which is a chosen methodology to develop the Ontology of Information Science OIS. This methodology uses an iterative approach allowing it to refine the ontology to create a more accurate model of the domain of IS.

The ontology moves slowly from knowledge level to implementation level to be understandable by a machine. We begin by introducing the method used to construct the OIS, which comprises two stages: building the conceptual model and converting it in to a logical model or computational model.

To build the conceptual model, it starts by identifying the purpose and scope of the OIS ontology which has been defined in the specification.

- **Specification**

The specification is establish the knowledge that ontology will cover which in the Information Science domain consists of three branches namely, archival science, Library science and computer science. The reason behind that is that IS is a broad domain and overlaps with other sciences, and these branches are the root of the science. The developing OIS ontology is providing a unified domain model that cover these branches of the IS field such as; library science, archival science and computer science. This stage is very important to the development process as it helps to avoid any inconsistencies between the classes of the ontology.

- **Conceptualization**

1. **Knowledge acquisition**: the information is collected from the Ontocop community. Ontocop is a virtual community of practice that has been designed to support ontology development. Besides creating corpus it contains publications of the ontocop members which are used to speed up the process [24].

### I. Building the IS Glossary:

The conceptualization consists of identifying IS concepts in a glossary of terms to ensure that each term is described with its synonyms and acronyms. This is the initial stage in building the conceptual model of the ontology. The glossary can include many terms that refer to the same concept. It contains all related terms in the domain, including synonyms and acronyms. Simple descriptions are given for each term included in the ontology. Relations are built between them to avoid any ambiguity. For example these concepts;

- Community of informatics.
- Compact Disk
- Compact Disk-Recordable
- Compact Disk-Rewritable
- Computer
- Computer aided design
- Computer aided manufacturing.

Each concept has its description and synonym and acronym to be used as a base of conceptual model of OIS ontology. A fragment of the IS glossary is illustrated in Table 1.

### Table 1 Information Science Glossary

<table>
<thead>
<tr>
<th>Name</th>
<th>Synonyms</th>
<th>Acronyms</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geographic</td>
<td>GIS</td>
<td>is the field of</td>
<td>study of developing , using and application of geographic information system</td>
</tr>
<tr>
<td>inform</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>science</td>
<td>GIS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electronic</td>
<td>Em</td>
<td>online newsletter</td>
<td>published on the internet</td>
</tr>
<tr>
<td>Email</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electronic</td>
<td>publish</td>
<td>Publishing of</td>
<td>material in electronic form such as publishing on the web or CDs,DVDs.</td>
</tr>
<tr>
<td>publish</td>
<td>bli</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ing</td>
<td>shi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ng</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cybermetrics</td>
<td>We</td>
<td>science of describe and measure the internet and its impact as communication tool</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>bm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>etri</td>
<td>cs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Copyright © 2012, Infonomics Society

II. Taxonomy and hierarchy building: this involves identifying classes and relationships in the ontology (including subclasses and superclasses).

The concepts were grouped based on a faceted classification scheme for building classifications after creating the concepts dictionary. The two main approaches implemented to modeling OIS ontology were top down approaches and bottom up; the reason for choosing them is to model the concepts and relationships at general level and to control the level of details.

Identifying the terminological concepts and relations in the IS concept dictionary provides concept names such as public library, while an example of a class attribute would be Library Type, an instance attribute could be name, size or service, and typical relations could include is part of or has a relation with. These aspects all provide information that describes the classes and their value.

- **Formalization and implementation:**

  The knowledge model is formalized by using Protégé the ontology editor. Using the Protégé ontology editor helps to generate ontology codes that explicitly conceptualizes the domain concepts captured in previous stages. The formal language used is OWL language and we use the WebProtégé for ontology run time.

- **Evaluation:** as the ontology provides a fundamental data used to conceptualize knowledge, it has to be evaluated based on predefined criteria. An ontology evaluation requires consideration of the stability and accuracy of the ontology.

- **Maintenance** of the conceptual model and the computational model. Ontology needs to be refined and maintained to ensure usefulness of its usability during the time.

- **Documentation** of the ontology life-cycle to help users and developers to follows its structure and development process.

3.1 Tools and Techniques

Protégé is an ontology editing tool that was developed in Informatics at Stanford University. Protégé facilitates ontology construction by using different formats and plug ins to add more functions. These plug-ins include importing and exporting ontology language (XML, OIL, FLogic) and reasoning for instance. The platform of Protégé supports two ways of modeling ontologies:

- Protégé frame editors enable end users to develop and populate ontology support by Open Knowledge Base Connectivity protocol (OKBC).
- Protégé OWL editor allows the user to create and develop ontologies using web ontology language.

WebProtégé is a web interface providing a flexible environment for experts to work collaboratively. It is a tool to develop the ontology’s process and make the ontology accessible from any web browser. There is a difference between WebProtégé and other tools such as Wikis. It supports OWL 2.0 which is compatible with Protégé 4.

4. Implementation

Ontology is a controlled vocabulary and a formal language that means that knowledge can be expressed in a way that is computer interpretable. The OIS is structured as a combination of an upper and domain ontology. The upper ontology contains meta-generic, abstract and philosophical ideas about the domain.

![Figure 2 Class tree of OIS](image-url)

"Every specific domain ontology must be used as framework for an upper-level ontology which describes the most general domain as independent categories of reality"[25]. Its aim is to support semantic interoperability among heterogeneous information systems [26].

The main feature of our OIS ontology is to provide a dictionary of defined concepts and the relations
between them, and to organize them into a hierarchy of classes and sub-classes.

The main feature of OIS ontology consists of a large number of classes - 687 - and approximately 70 assertions, including more than 67 rules and relations, providing the capability for rich semantic expression. Additionally, the OIS ontology provides a rich hierarchy of associative relations defining a complex structure, as can be seen in Figure 2.

The OIS ontology is providing a rich hierarchy of associative relations used to define a complex structure. The main components of the OIS ontology are classes such as Tools, Domains, Resources, and Methods; these classes are objects that abstract the real world. The second level in the ontology subclass represents more details of the superclass, such as Public Library, Standards, and Information Privacy. The second component is properties. It is also called relationships in OWL - between superclasses and subclasses; data properties and object properties are used. In OIS ontology two types of properties are used to link between two classes and linking between general classes and their instances. These relations have been used <Is-A>, <is Part Of>, and <has A>. These relations are created in the Object property tab, as shown in Figure 3.

Figure 3. OIS ontology: object properties

The object properties used in the ontology such as: AccessBy, Collect, has Apart, includesOf, isDescribeA, link the classes and subclasses, adding constraints to defining classes.

The third component is instances of the OIS ontology, which are created in the individuals tab. Each instance can be described in the description tab, by specifying its type and name. For example, the Institute of Electronic and Electronic Engineering is described as the type “computing standard” and given the name IEEE, see Figure 4.

Figure 4 OIS ontology instances

The class LibraryAssociation class has an instance, CanadianLibraryAssociation which is a type of LibraryAssociation.

The ontology model is structured based on the IS taxonomy to ensure ontology richness. The upper level consists of fourteen entities; Actors, Method, Tools, Practice, Studies, Mediator, Kinds, Domains, Resources, Legislation, Philosophy & theories, Societal, Time, Space. Each entity has subclasses which are grouped under to describe and define their relationships. For example Actors Class is defined as actor in the domain such as a Person or Organization, and the subclass Person is divided into two subclasses; User, Employee. The Abstract sub-class is part-of the Class Tools that defines a representation of the contents of documents. Whereas, AbstractJournal is sub-class of the class Abstract, as shown in the ontology fragment below.

<!--

<rdf:subClassOf rdf:resource="#Tools"/>
<rdf:comment >representation of the contents of document.</rdf:comment>
The defined class in the ontology has unique definitions and is restricted by specific constraints. The Figure 5 shows defining classes and relationships between subclasses. If we define the concept computer

**Figure 5 Defining classes and relationships**

Define Class Computer (? Computer)

“Machine has hardware and Software”

Define – subclass Software (? Software)

“a technology which controls the functions of a computer’s hardware”

Define axiom

(and (subclasses-of Software Computer)
  ( Values - Facts)

“Software program version “

The usage of classes and individuals in OIS ontology can be checked through using the usage class feature in Protégé. for example there have been six usages of the class Access in OIS ontology. One usage is subclass of Practice and 5 usages as superclass of AccessService, Accession, DirectAccess and InformationGap. See Figure 6).

**Figure 6 Class usage in OIS ontology**

5. Evaluation

An evaluation could look at the terminology used from a technical or expert perspective, to ensure terms are defined accurately and the correct methods have been used during the development.

Ontology evaluation should also ensure that we avoid concepts duplication, excessiveness and inconsistent relationships to make our understanding simpler.

The evaluation can be done at the developing process and running time. At the development process the interim evaluation is applied to ensure its completeness and consistency and to improve it.

Ontology consistency can be checked by using a reasoner such as FaCT++ Pellet, or Hermit. The reasoner is one of the main advantages of using a logic-based language such as OWL–DL, which is supported by Protégé 4. The reasoner can be used at the point of developing ontology, publishing time, and run time in applications as a querying mechanism.
In OIS ontology FaCT++ is applied several times to classify the ontology structure correctly, and it indicates any classification errors. If a class is incorrectly classified it will appear in red in a root class called *Nothing*. For example, the class *Government Library* appears as inconsistent in the class category, as shown in figure 7.

![Inferred class hierarchy: Thing](image)

**Figure 7. Inconsistent classes**

Also, after classifying the ontology we find the sub-class *ElectronicDocumentDelivery* is classified under the *Domain* class, which it should be under the *InformationService* Class.

Alternatively, a taxonomical evaluation could be carried out to ensure that the ontology is classified in a manner that is consistent and complete. For example, if we say that *ComputerScientist* is a subclass of *information* and we also say that *GovernmentLibrary* and *SchoolLibrary* fall under the class of *libraries* and the ontology indicates that both are *PublicServices*, we will find an error because *PublicService* will be a subclass of both libraries.

The error occurs because a class in a partition cannot have a common subclass: *professional* and *information professional* are subclasses of *ComputerScientist*. This is known as circularity.

The OIS ontology is evaluated based on predefined ontology criteria by the domain experts. These criteria are namely:
- Completeness refers to ensuring that the ontology includes all concepts, relationships, and axioms relevant to the domain.
- Consistency means avoiding including contradictory information or confounding concepts.
- Clarity to make sure that we have defined our terms with clarity.

The Ontocop members are asked to answer many questions based on web surveys to indicate their satisfaction on the OIS ontology. The OIS ontology is published on WebProtégé to be available for users and evaluators as illustrated in Figure 8.

![OIS ontology on WebProtégé](image)

**Figure 8. OIS ontology on WebProtégé**

WebProtégé searching on the ontology provides good results. The querying on concept information can display (64) result. Also, it is possible to select any concept in the tree to gain its annotations and relations with other components of the ontology.

### 5.1 Evaluation Results

After two months of the evaluation period the domain experts point out some changes to classes in the ontology structure such as adding the subclasses *EvidentialValue* and *HistoricalValue*, and changing the class to *Person* instead of *People*.

Most of the participants assess their satisfaction with the ontology’s consistency at the level 3 (64%). Others express it level 2, and 4 (20%, 12%) respectively.

Diagram 8 shows that most of the participants indicated their level of satisfaction at level 3 for the rest of the criteria. That means the ontology is not fully satisfactory and not unsatisfactory, as none of them choose level 1 or 5. The consistency of relationships of OIS ontology was satisfied, according to (56%) of the responses, whereas, completeness, generality and clarity is (48%, 44%, 40%) respectively.
6. Discussion

The OIS is based on a conceptual model that indicates that there are several classes in the IS domain. The class resource, for example, has several properties: creator name, location name, resource description. A specific book could belong to this class. Book, as a subclass of resource, has various data type properties that are used to identify its class, such as book title, or author name. Meanwhile, the object property is used to establish the domain relationships among the classes and subclasses.

The object property RetrievesBy, for example, tells the Library catalogue to explore all the resources that are stored in it. The object property requiresDevice indicates that a digital resource stored in a computer should be used to retrieve the required resources. The Computer class describes the tools required in the IS domain. It is divided into Software and Hardware subclasses. The software class uses data types to indicate certain features, such as the name and version of the software, or to distinguish among the various web browsers (e.g., Safari, Lynx and Firefox).

Our Hardware subclass includes features that describe the class, such as CPU type which determines the capacity to receive inputs and produce outputs. The computermemory is also included. Information stored on cards such as memory, video and audio cards is also captured. The object property (has A) shows what kind of hardware is used (e.g., a specific monitor).

7. Conclusion and Future works

Creating ontology of information science is not an easy task, due to the complex nature of the field. This paper has presented such an OIS, and the methods used to construct it. We have also demonstrated the strategy used to build and design the conceptual model for the domain using the ontology technique. This OIS can help to identify the features of this science, which is made up of an overlapping set of different sciences, making it difficult to determine its boundaries.

The OIS needs to be integrated with other ontologies so as to provide a basic knowledge for the semantic web, to leverage the retrieval of information. Such ontologies include social science, cognitive science, philosophy, law management and mathematics.

6.2 Future Work

We aim to create a tool that will enable the mapping of the OIS onto other domain ontologies so as to create an upper ontology. This ontology could then be used by knowledge engineers or domain analysts. It will require a search module to provide a basic mechanism for searching the OIS using natural language or keywords. Advanced research is needed to retrieve the specific knowledge that users are seeking.

7. References


[7] F. Starlab. (2003), Systems Technology and application Research Laboratory homepage, Faculty of Science, Department of Computer Science, Vrije University Brussels.


