A review of ontologies with the Semantic Web in view
Ying Ding
Journal of Information Science 2001; 27; 377
DOI: 10.1177/016555150102700603

The online version of this article can be found at:
http://jis.sagepub.com/cgi/content/abstract/27/6/377

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
CILIP
Chartered Institute of Library and Information Professionals

Additional services and information for Journal of Information Science can be found at:

Email Alerts: http://jis.sagepub.com/cgi/alerts
Subscriptions: http://jis.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.co.uk/journalsPermissions.nav
Citations http://jis.sagepub.com/cgi/content/refs/27/6/377
A review of ontologies with the Semantic Web in view

Ying Ding

Free University, Amsterdam, The Netherlands

Received 6 May 2001

Abstract.

The World Wide Web is currently starting to move from the first generation to the second generation: the Semantic Web. Ontologies are the backbone for this Semantic Web. This paper aims to introduce and give the readers an overview of ontology in general. It discusses the definitions of ontology, kinds of ontology, ontology tools, ontology language and some important ontology projects, both current and completed.

1. Introduction

The World Wide Web is currently starting to move from the first generation to the second generation: the Semantic Web. Berners-Lee introduced the vision of a Semantic Web that provides many more automated services based on machine-processable semantics of data and heuristics that make use of these metadata [1]. Ontologies that provide shared and common domain theories will be a key asset for such a Semantic Web. They can be seen as metadata that explicitly represent the semantics of data in a machine-processable way. Ontology-based reasoning services can operationalize these semantics for providing various services. Ontologies help people and computers to access the information they need and to effectively communicate with each other. They have, therefore, a crucial role enabling content-based access, interoperability and communication across the Web, providing it with a qualitatively new level of service: the Semantic Web. This will weave together a net linking incredibly large parts of human knowledge and will complement it with machine processability.

Ontology is a philosophical theory about the nature of existence. Artificial intelligence (especially knowledge acquisition and representation) researchers are reincarnating this term as their own jargon for expressing ‘a shared and common understanding of some domain that can be communicated between people and application systems’ [2]. A typical ontology has a taxonomy defining the classes and their relations and a set of inference rules powering reasoning functions [3].

Since the early 1990s, ontologies have become one of the popular research topics investigated by several artificial intelligence research communities, including knowledge engineering, natural language processing and knowledge representation. The notion of an ontology is also becoming visible in fields such as intelligent information integration, information retrieval, knowledge management, web standards, online databases, multi-agent systems etc. The reason for ontologies becoming so popular is the lack of standards (shared knowledge) for communication syntactically and semantically both from human and computer perspectives. This problem will be exacerbated with the exponential increase of information overload, inefficiency of current web search engines and online information retrieval, the relative ambiguity of the universals of discourses etc. Ontology, as a formal explicit specification of a shared conceptualization, provides a promising way to tackle this problem.

This paper aims to introduce and to give the readers an overview of ontology in general. It comprises the following parts: ontology definitions, kinds of ontology,
ontology tools, ontology language and some important ontology projects both current and completed.

2. Ontology definitions

Ontology is a term used to denote the shared understanding of some domains of interest, often conceived as a set of classes (concepts), relations, functions, axioms and instances [4]. Guarino [5] established a comprehensive survey of ontology definitions from the highly cited relevant works in the knowledge-sharing community [2, 5, 6–10]. In the knowledge representation community, the commonly used or highly cited ontology definition is from Gruber [4]: ‘an ontology is a formal, explicit specification of a shared conceptualization. “Conceptualization” refers to an abstract model of phenomena in the world by having identified the relevant concepts of those phenomena. “Explicit” means that the type of concepts used, and the constraints on their use are explicitly defined. “Formal” refers to the fact that the ontology should be machine readable. “Shared” reflects that ontology should capture consensual knowledge accepted by the communities’.

3. Kinds of ontology

Loosely speaking, any organized set of objects can be considered as an ontology according to the ontology definition discussed above, for instance, catalogues, indexes from the information retrieval community; entity-relationship (ER) models from the database community; dictionaries, thesauri from the computational linguistic community; and object-oriented class definitions from the software engineering community [11]. Some of the main contributions to the classification of ontologies are:

- Uschold and Gruninger [12] and Uschold [13] classified ontology from three dimensions: (1) formality (highly informal – expressed loosely in natural language; semi-informal – expressed in a restricted and structured form of natural language, e.g. the text version of the ‘Enterprise Ontology’ [14]; semi-formal – expressed in an artificial formally defined language, e.g. the Ontolingua version of the Enterprise Ontology; rigorously formal – meticulously defined terms with formal semantics, theorems and proofs of such properties as soundness and completeness, e.g. TOVE); (2) purpose (communication, inter-operability, systems engineering benefits (reusability, knowledge acquisition, reliability, specification)); and (3) subject matter (subject such as domain (domain ontology), subject of problem solving (task, method, or problem solving ontology), subject of knowledge representation languages (representation ontology or meta-ontology)).

- Guarino [5] classified the ontology according to two dimensions: (1) level of detail, e.g. meta-level ontology (frame ontology [4]), reference ontologies, shareable ontologies, domain ontologies; and (2) level of dependence, e.g. top-level ontologies, task ontologies and application ontologies.

- Gomez-Perez and Benjamins [15] classified ontologies as follows: (1) knowledge representation ontologies (e.g. formal ontology); (2) general/common ontology (CYC ontology); (3) top-level ontology (Sowa’s boolean lattice, PANGLOSS, Penman Upper Level, Mikrokosmos, WordNet upper level, etc.); (4) meta-ontology/generic ontology/core ontology (e.g. mereology ontology); (5) domain ontology (e.g. EngMath ontology, PhysSys, Enterprise, TOVE, [KA]², etc.); (6) linguistic ontology (e.g. WordNet, Sensus, GUM); and (7) task ontology (e.g. domain-task ontology, method ontology, application ontology).

4. Important ontologies

There are some important ontologies developed by the artificial intelligence and language engineering communities. These ontologies not only reflect the development of the ontology engineering field but significantly influence the current ontological researches as well.

4.1. Cyc

The Cyc ontology is one of the important parts of the CYC project. It was created on the basis of microtheories. Each microtheory captures the knowledge and reasoning required for some particular domains, such as space, time, causality, or agents [16].

4.2. TOVE

TOVE ontologies constitute an integrated enterprise model, providing support for more powerful reasoning based on the interaction of the ontologies found at www.eil.utoronto.ca/tove/toveont.html: (1) core ontologies – product ontology, service ontology, activity ontology, organization ontology, and resource ontology; (2) derivative ontologies – transportation
ontology, inventory ontology, quality ontology, product design ontology, goals ontology, scheduling ontology, operating strategies ontology, product requirements ontology, information resource ontology, intended action ontology, and electro-mechanical product ontology; (3) enterprise ontologies – enterprise design ontology, project ontology, material flow ontology, and business process ontology [17].

4.3. Enterprise Ontology

The Enterprise Ontology is a collection of terms and definitions relevant to business enterprises and was developed in the Enterprise Project. The major contents of the Enterprise Ontology are as follows: (1) meta-ontology and time – terms used to define the terms of the ontology (e.g. entity, relationship, role); (2) activity, plan, capability and resource – terms related to processes and planning; (3) organization – terms related to how organizations are structured; (4) strategy – terms related to high level planning for an enterprise; and (5) marketing – terms related to marketing and selling goods and services [9].

4.4. KRSL Plan Ontology

KRSL (Knowledge Representation Specification Language) plan ontology is part of the research at the ARPA/Rome Laboratory Planning Initiative (ARPI). This ontology has two major aspects: an abstract ontology setting out major categories (such as space, time, agents, actions, reasoning and plans), and a set of modular specialized ontologies which augment the general categories with sets of concepts and alternative theories of more detailed notions commonly used by planning systems [18].

4.5. WfMC (Workflow Management Coalition)

This is a standard terminology containing technical definitions for terms to be used in the WfMC specifications and discussions. For each term, a definition, a discussion of usage, and a set of possible synonyms are provided. This serves as a semi-informal ontology for the interoperability between different workflow systems [12].

4.6. STEP

STEP (Standard for the Exchange of Product Model Data) is a kind of semi-informal ontology used for the interoperability and exchange of products among different computer systems. The STEP ontology describes product data throughout the life cycle of a product, which facilitates implementation, product-sharing and archiving [12].

4.7. EL ontology

EL ontology [19] is an example of an ontology developed for artificial intelligence and natural language processing. It is a very liberal ontology including kinds, ideas, facts and events, and also an inference engine capable of efficiently making complex inference.

4.8. SENSUS

This contains a simple taxonomic structure (no meaning axioms) of about 50,000 nodes, mostly resulting from the merging of the WordNet thesaurus [20] into the PENMAN top-level ontology [21].

4.9. WordNet

This is a general linguistic ontology formed by synsets – terms grouped into semantic equivalence sets, each one assigned to a lexical category (noun, verb, adverb, adjective). Each synset represents a particular sense of an English word and is usually expressed as a unique combination of synonymous words. WordNet allows for sense disambiguation, and various kinds of semantic relations are maintained among synsets, for instance hypernymy, hyponymy and antonymy etc. [20]. It also provides multilingual supports, such as EuroWordNet.

5. Ontology language

An ontology has to be represented by predefined languages. Currently available ontology representation languages are either logic-based (first-order logic), frame-based (frame logic), or web-based (RDF, XML, HTML). A language absorbing the advantages from these three aspects will become a good language to represent ontology.

- **OIL**: OIL (Ontology Interchange Language) is a standard language proposed by the OnToKnowledge project (www.ontoknowledge.org). It fused three paradigms: frame-based modelling with semantics based on description logic and syntax based on web standards such as XML and RDF schema. OIL has been successfully applied in several areas, such as knowledge management, e-commerce and so on.
[22–25]. Currently DAML+OIL as the standard ontology representation language was released by the Joint Committee (Joint US/EU ad hoc Agent Markup Language Committee) in March 2001 (www.daml.org/2001/03/daml+oil-index).

- **KIF**: Knowledge Interchange Format (KIF) is a language designed for use in the interchange of knowledge among disparate computer systems based on first-order predicate logic. KIF has many essential characters: declarative semantics, logically comprehensive representation of knowledge about knowledge, implementability and readability.

- **CycL**: CycL is a formal language whose syntax derives from first-order predicate calculus. The vocabulary of CycL consists of terms: semantic constants, non-atomic terms (NATs), variables, numbers, strings, etc. Terms are combined into meaningful CycL expressions, ultimately forming meaningful closed CycL sentences (with no free variables.). A set of CycL sentences forms a knowledge base.

- **LOOM**: LOOM is a high-level programming language based on first-order logic which belongs to the KL-ONE family. The LOOM language provides an expressive and explicit declarative model specification language, a powerful deductive support, several programming paradigms, and knowledge-base services.

- **CML**: CML (Conceptual Modelling Language) was created by the KACTUS project. This language was originally developed as part of the KADS and CommonKADS projects. CML is different from most other ontology formalisms in that it makes an explicit distinction between domain knowledge, inference knowledge, task knowledge and problem-solving knowledge. CML uses a notation that is mostly informal so that knowledge modelled in CML cannot be executed by a program.

- **Conceptual representation**: Roux, et al. [26] adopted this conceptual graph to store ontology as a semantic graph, which is described in a lattice where each node is related semantically to the other nodes along an ‘ISA’ relation. Faure and Nedellec [27] represented their ontology using the directed acyclic graph (DAG). Borgo et al. [28] adopted a concept graph (lexical semantic graph) to represent the knowledge base and the user queries. Guarino et al. [29] proposed the lexical conceptual graph (LCGs, the same as the lexical semantic graph) as an oriented connected graph. The rationale behind the LCG’s lexical and semantic constraint is bound to the choice of exploiting a linguistic ontology to clarify their intended meaning and check their consistency. Mitra et al. [30] adopted a graph-based model to represent ontologies.

### 6. Ontology tools

Some important ontology tools are available and briefly described below (for a comprehensive survey of ontology tools, see Duineveld et al. [31]).

- **Ontolingua**: Ontolingua is a set of tools, written in Common Lisp, for analysing and translating ontologies, which was developed in the early 1990s at the KSL of Stanford University [32]. Ontolingua consists of a server and a representation language. This server provides a repository of ontologies to assist the users in creating new ontologies and amending the existing ontologies collaboratively. The ontology stored at the server can be converted into different formats [31].

- **WebOnto**: WebOnto was developed by the Knowledge Media Institute of the Open University [33]. It was designed to support the collaborative browsing, creation and editing of ontologies. It provides a direct manipulation interface displaying ontological expressions and also an ontology discussion tool called Tadzebao which could support both asynchronous and synchronous discussions on ontologies [34].

- **Enterprise toolsets**: these are implemented using an agent-based architecture to integrate off-the-shelf tools in a plug-and-play style. The components of the Enterprise Toolset are: a Procedure Builder for capturing process models, an Agent Toolkit for supporting the development of agents, a Task Manager for integration, visualization, and support for process enactment, and an Enterprise Ontology for communication [9].

- **KACTUS toolkit**: VOID, the KACTUS toolkit, is an interactive environment for browsing, editing and managing (libraries of) ontologies. VOID supports the theoretical and application oriented work packages by providing an environment in which one can experiment with theoretical issues (e.g. organization of libraries of ontologies, translating between different ontology formalisms) and also performing practical work (e.g. browse, edit and query ontologies in various formalisms). In order to support reuse of ontologies, the toolkit can handle various ontology formalisms (CML, EXPRESS and Ontolingua) and can perform (partial) translations between these formalisms.
7. Ontology projects

7.1. OntoKnowledge (www.ontoknowledge.org)
This is a European Union-funded project started in 1999 and focusing on content-driven Knowledge-Management through evolving ontologies. The technical backbone of OntoKnowledge is the use of ontologies for the various tasks of information integration and mediation. The application focus of OntoKnowledge is knowledge management in large and distributed organizations through three case studies in the areas of the organizational memory of a large company, help desk functionality of a call centre, and the knowledge management in a virtual enterprise. The OntoKnowledge project will develop methods and tools and employ the full power of the ontological approach to facilitate knowledge management. These tools can help knowledge workers who are not IT specialists to access company-wide information repositories in an efficient, natural and intuitive way.

7.2. OntoWeb (www.ontoweb.org)
This is another European Union-funded Thematic Network project starting from this year. The extended name of this project is Ontology-based Information Exchange for Knowledge Management and Electronic Commerce. The goal of the OntoWeb Network is to bring together researchers and industrials coming from the research and applications areas above, promoting interdisciplinary work and strengthening European influence on Semantic Web standardization efforts such as those based on RDF and XML.

7.3. OntoBroker (ontobroker.aifb.uni-karlsruhe.de)
The Ontobroker project uses ontologies to annotate and wrap web documents and provides an ontology-based answering service. The Ontobroker supports clients that query for knowledge as well as providers that want to enhance the accessibility of their web documents. The overall architecture of Ontobroker includes four basic engines: query engine (receiving queries and answering them by checking the content of the databases provided by information and inference agents); information engine (collecting various factual knowledge from the web); inference engine (deriving and inferring additional factual knowledge); and database manager (linking and transferring data to the above three engines) [25].

7.4. Cyc (www.cyc.com)
Cyc is a project of the Microelectronics and Computer Technology Corporation (MCC) in Austin, Texas that provides a foundation for common sense reasoning by developing ontologies for a wide variety of domain-specific applications [16]. All of the knowledge in Cyc is represented declaratively in CycL (a variant of first-order logic). The Cyc knowledge base contains simple assertions, inference rules, and control rules for inference. The inference engine can be used to derive new assertions based on this knowledge base. The ontologies underlying Cyc are organized into sets of modules known as microtheories.

7.5. TOVE (www.eil.utoronto.ca/tove/ontoTOC.html)
The TOVE (TOronto Virtual Enterprise, University of Toronto) project was conducted by the Enterprise Integration Laboratory from the Department of Mechanical and Industrial Engineering of the University of Toronto, Canada. This project focused on enterprise modelling, concurrent engineering and integrated supply chain management [17, 35]. TOVE has been used to model two enterprises: a computer manufacturing enterprise and an aerospace engineering company.

7.6. Enterprise (www.aiai.ed.ac.uk/~entprise/enterprise)
The Enterprise project is the UK government’s major initiative to promote the use of knowledge-based systems in enterprise modelling. This project is focused on management innovation and the strategic use of IT to help manage change. It is aimed at providing a method and computer toolset which will help capture aspects of a business and analyse these to identify and compare options for meeting the business requirements. The framework for integrating methods and tools is solidly based on an ontology for enterprise modelling [9]. Real-life applications have been conducted in Unilever, IBM, Lloyd and AIAI.

7.7. OntoSeek (www.ladseb.pd.cnr.it/infor/Ontology/Papers/OntoSeek.pdf)
OntoSeek is a general purpose tool for ontology-based information retrieval being developed by CORINTO (Consorzio di Ricerca Nazionale Tecnologia Oggetti, IBM Semea, Apple Italia and Selfin Spa) with the cooperation of LADSEB-CNR, as part of a project on retrieval
and re-use of object-oriented software components [28]. One of the key choices of this project has been avoiding the construction of an ontology from scratch, relying instead on a large ontology built for purposes of natural language translation [5].

7.8. PIF (www2.ics.hawaii.edu/~jl/pif.html)

The goal of the Process Interchange Format (PIF) project is to support the exchange of business process models among different process representations by developing PIF (an interlingua to unify heterogeneous process representations) along with local translators between PIF and local process representations. There is a core PIF ontology with which all translators operate. In addition, there are different extensions of this core ontology which various ontologies may share. In PIF, these extensions are captured by partially shared views, so that ontologies that have a partially shared view in common can translate without loss of expressiveness.

7.9. KACTUS (www.swi.psy.uva.nl/projects/Kactus/home.html)

KACTUS is a European ESPRIT project aimed at the development of a methodology for the reuse of knowledge about technical systems during their life-cycle, so that the same knowledge base can be re-used for design, diagnosis, operation, maintenance, redesign and instruction. KACTUS supports an integrated approach embracing computer-integrated manufacturing and engineering methods, and knowledge engineering methods by creating an ontological and computational basis for re-use of product knowledge across different applications within technical domains. It achieves this by creating domain ontologies and reusing them for different applications [36]. The main formalism in KACTUS is CML (Conceptual Modelling Language) and it also provides a toolkit that is an interactive environment for browsing, editing and managing (libraries of) ontologies.

7.10. KRAFT (www.csd.abdn.ac.uk/~apreece/Research/KRAFT.html)

KRAFT is a research project on the integration of heterogeneous information using agent architecture [37]. The project is a joint collaboration between the Universities of Aberdeen, Cardiff and Liverpool in conjunction with British Telecom and began in May 1996. KRAFT architecture has been applied in the area of student-admission policies and the design of a router configuration in the telecommunications domain.

8. Closing remark

This paper ends with a summary by Uschold and Gruninger [12] on the frontiers of ontology research:

‘We conclude this paper with a brief discussion of several important issues and opportunities for ontology research. These are:
- development of ontologies as interlingua to support interoperability among tools in some domain;
- development of tools to support ontology design and evaluation;
- development of libraries of ontologies;
- development and integration of new ontologies;
- methodologies for the design and evaluation of ontologies.’

References


A review of ontologies with the Semantic Web in view

of the Eleventh Workshop on Knowledge Acquisition, Modeling and Management, KAW’98, Banff, Canada, 1998.


