Ontology Languages – A Review

V. Maniraj, Dr.R. Sivakumar

Abstract—**Ontologies have been becoming a hot research topic for the application in artificial intelligence, semantic web, Software Engineering, Library Science and information Architecture. Ontology is a formal representation of set of concepts within a domain and relationships between those concepts. It is used to reason about the properties of that domain and may be used to define the domain. An ontology language is a formal language used to encode the ontologies. A number of research languages have been designed and released during the past few years by the research community. They are both proprietary and standard based. In this paper a study has been reported on the different features and issues of these languages. This paper also addresses the challenges for research community in the further development of ontology languages.**

I. Introduction

Ontology engineering (or ontology building) is a subfield of knowledge engineering that studies the methods and methodologies for building ontologies. It studies the ontology development process, the ontology life cycle, the methods and methodologies for building ontologies, and the tool suites and languages that support them.

Ontology engineering aims to make explicit the knowledge contained within software applications, and within enterprises and business procedures for a particular domain. Ontology engineering offers a direction towards solving the interoperability problems brought about by semantic obstacles, such as the obstacles related to the definitions of business terms and software classes. Ontology engineering is a set of tasks related to the development of ontologies for a particular domain.

Most ontologies describe individual (to instances), classes (concepts), attributes and relations. Common components of ontologies include individuals, classes, attributes, relating, functions terms, restriction, rules axioms and events.

Ontologies are used in artificial intelligence, semantic web, software engineering, Biomedical information, library science and information architecture as a form of knowledge representation about the world or some part of it.

An ontology language is a formal language used to encode the ontology. There are a number of such languages for ontologies, both proprietory and standard-based such as common Algebraic specification language, common logic, CycL, DOGMA, Gellish, IDEF5, KIF, RIF, and OWL. Ontology languages can be classified as

V. Maniraj, Assoc. Professor, Department of Computer Science, A.V.V.M. Sri Pushpam College, Poondi, Thanjavur, Tamilnadu, India.
Dr.R. Sivakumar, Assoc. Professor & HOD of Computer Science, A.V.V.M. Sri Pushpam College, Poondi, Thanjavur, Tamilnadu, India.

1) Logical Languages
   - First order predicate logic
   - Rule based logic
   - Description logic
2) Frame based Languages
   - Similar to relational databases
3) Graph based Languages
   - Semantic network
   - Analogy with the web is rationale for the semantic web

II. Background

CycL[1] in computer science and artificial intelligence is an ontology language used by Doug Lenat's Cye artificial intelligence project. Ramanathan V. Guna was instrumental in the design of the language. There is a close variant of cycl known as MELD. Declarative language based on classifical first-order logic, with extensions for model operators and higher order quantification. Knowledge Interchange Format (KIF)[2] is a computer oriented language for interchange of knowledge among disparate program. Gellish[3] is information representation language knowledge base.

DOGMA[4] short for Developing Ontology-grounded methods and Application, is name of research project in progress at Vrije Universiteit Brussel's START Lab, Semanties Technology and applications Research Laboratory.

IDEF5[5] (Integrated Definition for Ontology Description Capture Methods) is a software engineering method to develop and maintain usable, accurate, domain ontologies. This standard is part of the IDEF family of modeling languages in the field of software engineering. The Web Ontology Language (OWL)[6] is a family of knowledge representation languages for authoring ontologies and endorsed by the world wide web consortium. RIF[7] is a standard Rules Languages for the Semantic Web.

III. Types Of Languages

CyCL

CycL is used to represent the knowledge stored in the Cyc Knowledge Base, available from Cycorp. The source code written in CycL released with the OpenCyc system is licensed as open source, to increase its usefulness in supporting the semantic web.

Basic ideas of CycL :

- Naming the constants used to refer to information for represented concepts.
- Grouping the constants together in a generalization / specialization hierarchy (usually called categorization).
- Stating general rules that support inference about the concepts.
- The truth or falsity of a CycL sentence is context – relative; these contexts are represented in CycL as Microtheories.

Constants

The concept names in Cyc are known as constants. Constants start with "#$" and are case-sensitive. There are constants for Individuals, Collections, Truth Functions and Functions.

Specialization and Generalization

The most important predicates are #$isa and #$genls. The first one (#$isa) describes that one item is an instance of some collection (ie: Specialization), the second one (#$genls) that one collection is a subcollection of another one (ie:generalization). Facts about concepts are asserted using certain CycL. Sentences. Predicates are written before their arguments, in parentheses :

Rules

Sentences can also contain variables, strings starting with "?".

Microtheories

The knowledge base is divided into microtheories (Mt), collections of concepts and facts typically pertaining to one particular realm of knowledge. Each microtheory has a name which is regular constant; microtheory constants contain the string "Mt" by convention.

Knowledge Interchange Format (KIF)

It has declarative semantics (i.e. the meaning of expressions in the representation can be understood without appeal to an interpreter for manipulating those expressions); it is logically comprehensive (i.e. it provides for the expression of arbitrary sentences in the first-order predicate calculus); it provides for the representation of knowledge about the representation of knowledge; it provides for the representation of nonmonotonic reasoning rules; and it provides for the definition of objects, functions, and relations.

The ontologies built into KIF are

Sets

A version of non Neumann/Bernays/ Godel set theory.

Sequences

Primitives for representing sequences (i.e., ordered bags or lists)

Numbers and arithmetic

Functions and relations for basic integer and real-number arithmetic

Relations

Functions and relations, value and handling undefined functions, variable arity, relationship to sets.

"KIF Meta"

For representing metalinguistic knowledge about expressions in the KIF language, such as quoted constraint expressions.

Gellish

Data storage and data communication lack a common standard widely applicable data model as well as a common data language and knowledge base with a taxonomy of concepts and a grammar for data exchange messages. The solution to this problem is in the form of the new "open" industry standard Gellish language and knowledge base, as a further development of the standard data model and ontology of two new ISO standards. Gellish is a suitable language for neutral data exchange between systems. The definition of Gellish includes the public domain ("Open data") Gellish knowledge base (STEPlib) with definitions of a large number of concepts and product models. This provides an ontology with standard reference data for customization of systems in a standardized way to be prepared for data harmonization, data integration and data exchange. It illustrates that a single table database or data exchange file format, the Gellish table format, is sufficient to express a wide range of kinds of facts about classes as well as facts about individual objects so that it can replace data models.

DOGMA

DOGMA is an ontology approach and framework that is not restricted to a particular representation language. This approach has some distinguishing characteristics that make it different from traditional ontology approaches such as (i) its groundings in the linguistic representations of knowledge and (ii) the methodological separation of the domain –verses – application conceptualization, which is called the ontology double articulation principle. The idea is to enhance the potential for re-used and design scalability.

Another distinguishing characteristic of DOGMA is the explicit duality (Orthogonal to double articulation) in interpretation between the language level and conceptual level. The goal of this separation is primarily to disambiguate the lexical representation of terms in a lexon (on the language level) into concept definitions (on the conceptual level), which are word senses taken from lexical resources such as WordNet. The meaning of the terms in a lexon is dependent on the context of elicitation.

For example, consider a term "capital". If this term was elicited from a typewriter manual, it has a different meaning (read: concept definition) than when elicited from a book on marketing. The intuition that a context provides here is: a context is an abstract identifier that refers to implicit and tacit assumptions in a domain, and that maps a term to its intended meaning (i.e. concept identifier) within these assumptions.

Ontologies naturally co-evolve with their communities of use. De Leenheer (2007) identified a set of primitive operators for changing ontologies. We make sure these change primitives are conditional, which means that their applicability depends on pre- and post-conditions. By doing so, it guarantees that only valid structures can be built.

IDEF5

In the IDEF5 method, an ontology is constructed by capturing the content of certain assertions about real-world objects, their properties, and their interrelationships and representing that content in an intuitive and natural form. The IDEF5 method has three main components.

- A graphical language to support conceptual ontology analysis
- A structured text language for detailed ontology characterization, and
- A systematic procedure that provides guidelines for effective ontology capture.

In the 1990s IDEF5 is developed under the “Information Integration for Concurrent Engineering” (IICE) project, funded by the US Air Force, and developed by knowledge Based Systems, Inc. (KBSI), Texas.

Ontology development process

The IDEF5 ontology development process consists of the five activities such as Organizing and Scoping, Data Collection, Data Analysis, Initial Ontology Development, Ontology Refinement and Validation.

IDEF5 Building blocks

Kind

Kinds are properties typically expressed by common nouns such as ‘Employee’, ‘machine’, and ‘lathe’.

Individual

The most logically basic kind of real world object. Prominent examples include human persons, concrete physical objects, and certain abstract objects such as programs.

Referent

A construct in the IDEF5 elaboration language used to refer to a kind, object, property, relation, or process kind in another ontology or an IDEF model.

Relation

An abstract, general association or connection that holds between two or more objects.

State

A property, generally indicated by an adjective rather than a common noun, that is characteristic of objects of a certain kind at a certain point within a process.

Process

A real world event or state of affairs involving one or more individuals over some (possibly instantaneous) interval of time.

Diagram types

Various diagram types, or schematics, can be constructed in the IDEF5 Schemetic Language. There are four primary schematic types derived from the basic IDEF5 Schematic Language which can be used to capture ontology information directly in a form that is intuitive to the domain expert. They are classification Schematics, Composition Schematics, Relation Schematics and Object State Schematics.

OWL Ontologies

The data described by an OWL ontology is interpreted as a set of “individuals” and a set of “property assertions” which relate these individuals to each other. An OWL ontology consists of a set of axioms which place constraints on sets of individuals (called “classes”) and the types of relationships permitted between them. These axioms provide semantics by allowing systems to infer additional information based on the data explicitly provided.

OWL is both a syntax for describing and exchanging ontologies, and has a formally defined semantics that gives them meaning.

The W3C-endorsed OWL specification includes the definition of three variants of OWL, with different levels of expressiveness. They are OWL Lite, OWL DL and OWL Full.

RIF (Standard Rules Language for the Semantic Web)

RIF is a Web Rule Language. Rule Interchange Format (RIF) BLD (Basic Logic Dialect)

It horns rules with equality for the web.

It is based on RDF and OWL.

Rules, Ontologies and Data

It uses RDFS/OWL ontologies as data models for rules.

It makes Rules as extension to RDFS or OWL ontologies.

It defines interoperation between RIF, RDF and OWL by connecting their semantics which will be the solution.

The data sets for rules are RDF graphs.

RIF document (Ruleset) R may reference (import) RDF/OWL documents.

RIF – RDF combination is a tuple.

RIF-OWL DL combinations is a tuple.

Simple RDF, and RDFS entailment for combinations may be embedded into RIF.

IV. Discussion

This section presents a comparative study on the previously discussed features of various ontology languages. The CYC project has its own ontology language called CYCL based on first – order predicate calculus with some higher order extension. DOGMA (Developing Ontology Grounded Methods And Applications) adopts the fact oriented modeling approach to provide higher level of semantic stability.

The Gellish Language includes rule for its own extension and thus integrates an ontology with an ontology language. IDEF 5 is a software engineering method to develop and maintain usable, accurate, domain ontologies, KIF is a syntax for first order logic that is based on S-expressions. RIF (Rule Interchange Format) and F-Logic combine ontologies and rules. OWL is a language for making ontological statements. OWL is intented to be used over the world wide web.

Various ontologies have been published so far as follows:

- Basic formal ontology is a formal upper ontology designed to support scientific research.
- BIOPAX is an ontology for the exchange and interoperability of biological pathway data.
- CCO, Cell-Cycle Ontology is an application that represents the cell.
- CCOntology is an e-business ontology to support online customer complaint management.
- CIDOC is an ontology for cultural heritage.
- COSMO is an foundation ontology that is designed to contain representation of all of the primitive concepts needed top logically specify the meanings.
- CYC is an ontology for formal representation of the universe of discourse.
- Disease Ontology is designed to facilitate the mapping of desease.
- DOLCE is an ontology for linguistic.
- GUM (Generalized Upper Model) is an ontology for mediating between client system and national language technology.
- GOLD is an ontology for linguistic description.
- IDEAS GROUPA is an ontology for enterprise architecture.
- Linkbase is an ontology for formal representation of the bio medical domain.
- LPL is an ontology (Lawson Pattern Languages).
- OBO is an ontology for bio medicine.
- Plant Ontology is an ontology for plant structure.
- POPE Purdue is an ontology for pharmaceutical Engineering.
- PRO is protein ontology for protein information.

Types of ontologies with Languages

Basic formal ontology	OWL
BIOPAX	OWL
CCO	RDF
CIDOC	XML & RDF
COSMO	RDF
CYC	CYCL
GUM	OWLDL
GOLD	Owl/rdf
IDEAS	OWL
LPL	OBO

Language standardization it offers has resulted in the single most prominent change in ontology editors. This growth indirect support for RDF and various species of OWL has created some controversy.

The issue arises in consideration of whether RDF is the best base language for implementing ontologies on the Web or elsewhere, and whether it afforts the scalability necessary to implementing ontologies and webs of ontologies, and whether it affords the representational power or expressiveness to build ontologies of the sophistication necessary for demanding application.

Other ontology languages such as SCL, CYCL and LOOM, for example, arguably offer more power of expression and reasoning, but lack intimate support of RDF. The advantage offered by RDF that remains compelling for ontologies seems to be the universal use of URI and XML namespace protocols on the Web. This unifying aspect, for instance, may make it easier to establish, through collaboration and consensus, the utilization vocabularies (as ontologies) needed for far=flung cooperative and integrative applications using the Web.

Wearing the mantle of W3C standardization, OWL enjoys much more attention today than any other ontology language – 1 or but of the Web World. Its detractors tend to single out its limits of expression, its inelegant syntax and. Of course, its reliance onto the RDF model of representation using triples. Some basic language constructs like lists and other collections are deemed cumbersome and in need of extension in new language implementations.

V. Future Work

The world of ontology applications itself is changing in a way that is putting more pressure on ontology language implementations and editing tools to handle new tasks. Some see the gathering demands as an impending crisis for providing editing environments that can accommodate an expanding scope of ontology language responsibilities. Eventually, editing will have to address the ontology language and reasoner functions currently under development including:

- Ontology rule languages like the Semantic Web Rule Languages (SWRL), which combine RuleML with OWL.

- Probabilistic extensions for OWL like those being pursued at the University of Maryland Baltimore Country (UMBC).

- Defeasible or alternative logics to support forms of non-monotonic reasoning in closed – world and open – world ontologies that are being investigated.

- Complete reasoning capabilities of OWL Full and RDF, which are being investigated.

- Enforcing formal ontology principles on the design and implementation of ontologies as imposed by development environments like IODE and others.

- Semantic Web services ontology languages like those present in The Discovery Machine and OPCAT.

- Facility for associating an ontology or parts of an ontology with specific problem-solving methods (PSMs) how that knowledge is organized, such as the URIQA scheme.

- Effective integration of RDF ontologies and XML, schema for full mapping between them using path-query mechanisms analogous to XPath.

- Generation of ontologies specifically suited for use by agents in a multiple – agent system implementation as in the DISCIPLE tool.

- Automatic updating of domain ontologies through built in processes to acquire and analyze source domain information and identify modifications to the ontology as new or modified concepts and relations.

VI. Conclusion

A study on different ontology languages has been reported in this paper. The study included languages such as CycL, DOGMA, IDEF5, OWL, RIF, Gellish and KIF. Various features such a Rules, Relations, Micro theories, kind, State, Constants and process have been taken into account for the discussion. It is understood that the need of such languages is essential to encode ontologies. The future work will be concentrating on developing complete reasoning capabilities for OWL.

References

[1] The Syntax of CycL (http://www.cyc.com/cycdoc)
[2] http://www.ksl.stanford.edu/knowledge-sharing/kif
[3] http://ieeexplore.ieee.org
[4] http://www.starlab.vub.ac.be/website
[5] http://www.idef.com/IDEF5.html
[6] http://www.w3.org/TR/owl-guide
[7] RIF Basic Logic Dialect.H.Boley, M. Kifer
[8] What is an ontology?
[9] (http://www.ksl.stanford.edu./kst/what-is-anontology.html)
[10] John Bateman's Ontology Portal (http://www.fb10.uni-)
[11] Jarrar 2005, Jarrar 2007, De Leenheer etal 2007
[12] http://www.cellcycleontology.org
[13] http://www.linguistics – ontology.org/gold/2008/languages

V. Maniraj, Assoc. Professor, Department of Computer Science, A.V.V.M. Sri Pushpam College, Poondi, Thanjavur, Tamilnadu, India.

Dr.R. Sivakumar, Assoc. Professor & HOD of Computer Science, A.V.V.M. Sri Pushpam College, Poondi, Thanjavur, Tamilnadu, India.