# Ontologies and Knowledge Bases: State-Of-The-Art Report

Federico Peinado  Mateus Mendes

Junio 2006

Informe Técnico TR-9/06

Dpto. Sistemas Informáticos y Programación

---

Facultad de Informática
Universidad Complutense de Madrid

# Ontologies and Knowledge Bases: State-Of-The-Art Report

Federico Peinado Mateus Mendes

May 3, 2005

# Contents

**Abstract**

The aim of this report is to describe the most important ontologies, knowledge bases, and lexical databases available worldwide. Although this is not an *official* taxonomy, and several other ones may be found in the literature, we group the ontologies into this taxonomy: upper ontologies, which include only the most general concepts; general-purpose ontologies, which include the most often used concepts; common-sense ontologies, built with the goal of grabbing common-sense knowledge; multi-source ontologies, which include two or more other ontologies as their source of knowledge, and linguistic resources.

We also review some of the most important languages used to represent knowledge and build ontologies.

# Chapter 1

# General Resources

In these references the reader can find several links to the ontologies described here and many other knowledge resources that are not included in the scope of this review.

- **John Bateman's Ontology Portal**. Up to date list of ontologies.
  www.fb10.uni-bremen.de/anglistik/langpro/webspace/jb/info-pages/ontology/ontology-root.htm
- **Ontology Bibliography**. Links and documents about ontologies and ontological models.
  glotta.ntua.gr/nlp/StateoftheArt/Ontologies
- **Some Ongoing KBS/Ontology Projects and Groups**. Probably the most complete list of ontologies.
  www.cs.utexas.edu/users/mfkb/related.html
- **Linguistic Resources on the Internet**. A topically organized list of resources that may be of interest to the linguist.
  www.sil.org/linguistics/topical.html

# Chapter 2

# Ontologies

Ontologies are reusable components used in the field of Knowledge Engineering to build Knowledge-Based Systems. They represent knowledge in a generic way that allows them to be shared by different groups of people and applications. In Artificial Intelligence applications, ontologies are also useful for automated reasoning and knowledge representation.

As reflected by the different sections of this chapter, there are different ontologies for different purposes. Ontologies can also show different degrees of formalization. Some linguistic resources, also called lexical databases, have not been built with the aim of contributing to AI systems, but become more and more popular in such applications. The reasons for that might be large coverage, ease of use, and free availability.

## 2.1 Upper Ontologies

Upper Ontologies are limited to generic concepts, abstract and philosophical ones. Domain-specific concepts are not included in upper models. Usually more specific ontologies are built on upper ontologies' concepts when developing applications for specific domains.

### 2.1.1 DOLCE

Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) is an ontology developed as part of the OntoWeb project, which aims to develop ontologies for the semantic web. It does not intend to be a "universal" standard ontology, but it is simply used to compare and elucidate relationships between elements in the semantic web. It aims to be *minimal*, containing only the necessary and most used elements, to elucidate about the most important contents other ontologies should describe. Hence, it is being used as a source for other ontologies.

**URL:** www.loa-cnr.it/DOLCE.html

| | |
|---|---|
| **Contents:** | |
| **Main influences:** | WordNet |
| **Main applications:** | Semantic web |
| **Language:** | OWL including KIF as comments |
| **Availability:** | Free download |

### 2.1.2 Generalized Upper Model

The Generalized Upper Model (GUM) [BHR95] is one of the most popular ontologies. It is based on classic ontologies, like Penman Upper Model[1] [BKMW90], Merged Upper Model [HB94] and Ideation Base[2] [HM99]. It is an interface ontology between language and conceptual representation: close enough to linguistic features to be very objective, but general enough to go beyond the concrete details of the syntactic and semantic representations.

| | |
|---|---|
| **URL:** | purl.org/net/gum2 |
| **Main influences:** | Penman UM, Merged UM and Ideation Base |
| **Strong point:** | Interface NL–conceptual representation |
| **Main applications:** | Natural Language Processing |
| **Language:** | Loom, being translated into OWL |
| **Availability:** | Free download |

### 2.1.3 SUMO

SUMO (Suggested Upper Merged Ontology) is being created by the IEEE Standard Upper Ontology Working Group.The goal of this Working Group is to develop a standard upper ontology that will promote data interoperability, information search and retrieval, automated inferencing and natural language processing.SUMO has been translated into various representation formats, but the language of development is a variant of KIF (a version of the first-order predicate calculus): SUO-KIF[3].It also has an open source implementation in OWL.

According to its authors, SUMO and its domain ontologies "form the largest formal public ontology in existence today". It is constituted by a mapping of all the WordNet lexicon and many other specific ontologies (Communications, Countries and Regions, Distributed Computing, Economy, etc.). SUMO's contents are very general and abstract, but it also incorporates MILO (Mid-Level Ontology), which works as a bridge between

[1]Succintly, the Penman Upper Model states that a concept is relevant for the ontolgoy if it has specifiable consequences for the grammatical constructions that can be used to express it.

[2]More on Ideation Base in section 2.2.2.

[3]More about SUO-KIF in section 3.19.

the abstract concepts of SUMO and the fine detail of the various domain ontologies.

Comparing to other general-purpose ontologies, SUMO has the advantages of being very complete, and written in the standard language SUO-KIF.

| | |
|---|---|
| **URL:** | www.ontologyportal.org |
| **Contents:** | 20,000 concepts |
| **Main influences:** | WordNet and many other resources |
| **Strong point:** | Uses standards and aims to become a standard |
| **Main applications:** | |
| **Language:** | SUO-KIF, OWL and other implementations available |
| **Availability:** | Free download |

### 2.1.4 Upper CyC Ontology

Upper CyC Ontology is the upper level of a huge ontology called Cyc, developed by Cycorp, Inc., based in Austin, Texas (see section 2.3.2 for more information on Cyc). The Upper model contains the most general 3000 concepts of the whole ontology.

| | |
|---|---|
| **URL:** | researchcyc.cyc.com |
| **Contents:** | 3,000 concepts |
| **Main influences:** | Part of Cyc |
| **Strong point:** | Deals with Cyc, probably the most complete KB ever built |
| **Main applications:** | Every system |
| **Language:** | CyCL and an OWL version of OpenCyc |
| **Availability:** | Free for research purposes |

## 2.2 General-Purpose Ontologies

The vast majority of the most used ontologies were designed to contain general knowledge about the world. Their concepts may be instances of the ones present in an upper ontology or not. *NOTE:* We are including here many linguistic resources like linguistic-oriented ontologies and knowledge bases.

### 2.2.1 Frame Ontology

The frame ontology defines the terms that capture conventions used in object-centered knowledge representation systems. Since these terms are

built upon the semantics of KIF, one can think of KIF plus the frame-ontology as a specialized representation language. The frame ontology is the conceptual basis for the Ontolingua translators. One purpose of this ontology is to enable people using different representation systems to share ontologies that are organized along object-centered, term-subsumption lines. Translators of ontologies written in KIF using the frame ontology, such as those provided by Ontolingua, allow one to work from a common source format and yet continue to use existing representation systems.

| | |
|---|---|
| **URL:** | www.ksl.stanford.edu/htw/dme/thermal-kb-tour frame-ontology.html |
| **Contents:** | |
| **Main influences:** | |
| **Main applications:** | Interface between KR systems |
| **Language:** | Improved KIF |
| **Availability:** | Free download |

### 2.2.2 Ideation Base

Ideation Base [HM99] is an ontology that provides the ways of representation how we construe our experience through a language.

| | |
|---|---|
| **URL:** | www.brain.riken.go.jp/labs/lbis/publication NI_JASFLWS01.ppt (An introduction to the ontology) |
| **Contents:** | |
| **Main influences:** | |
| **Main applications:** | |
| **Language:** | |
| **Availability:** | |

### 2.2.3 KR Ontology

KR Ontology is based on the book Knowledge Representation by John F. Sowa. The basic categories and distinctions have been derived from a variety of sources in logic, linguistics, philosophy, and artificial intelligence. The two most important influences have been the philosophers Charles Sanders Peirce and Alfred North Whitehead, who were pioneers in symbolic logic. Peirce was also an associate editor of the Century Dictionary, for which he wrote, revised, or edited over 16,000 definitions. It has a top level that can be used as an upper ontology.

| | |
|---|---|
| **URL:** | www.jfsowa.com/ontology/kronto.htm |
| **Contents:** | More than 16,000 concepts |

**Main influences:** Various theories of KR
**Main applications:**
**Language:** Proprietary
**Availability:**

### 2.2.4 Mikrokosmos

The Mikrokosmos Ontology is a part of the Mikrokosmos machine translation project. Mikrokosmos is not committed to any particular theory of ontologies, but is built on more practical considerations. The main principle is a careful distinction between language-specific knowledge, represented in the lexicon, and language-neutral knowledge represented in the ontology. As a consequence, the semantics of words is represented partly in the lexical entries and partly in the ontological concepts. A set of detailed guidelines governs what belongs to a concept and what belongs to a lexical entry. The division of semantics also gives us the answer to how concepts are related to lexical items. In Mikrokosmos one is not forced to have one-to-one mapping between words and concepts. Words with related but not equivalent meanings can map to the same concept, while the differences are captured in the lexical entries.

**URL:** crl.nmsu.edu/mikro
**Contents:**
**Main influences:** None
**Main applications:** Translation
**Language:** Proprietary
**Availability:** Closed source, paid

### 2.2.5 MindNet

MindNet, from Microsoft Research, is a knowledge representation project that uses a broad-coverage parser to build semantic networks from dictionaries, encyclopedias, and free text. MindNets are produced by a fully automatic process that takes the input text, sentence-breaks it, parses each sentence to build a semantic dependency graph (Logical Form), aggregates these individual graphs into a single large graph, and then assigns probabilistic weights to subgraphs based on their frequency in the corpus as a whole. The project also encompasses a number of mechanisms for searching, sorting, and measuring the similarity of paths in a MindNet. The authors believe that automatic procedures such as MindNets provide the only credible prospect for acquiring world knowledge on the scale needed to support common-sense reasoning.

**URL:** research.microsoft.com/nlp/Projects/MindNet.aspx
**Contents:**
**Main influences:**
**Main applications:**
**Language:** Proprietary
**Availability:** Unavailable

### 2.2.6 The Component Library

The Component Library [CP97] is a hierarchically organised "library of formal representations of common actions, entities, and modifiers to enable building knowledge bases efficiently". Its concepts are domain-independent, inspired by existing ontological and lexical resources, such as WordNet, FrameNet and VerbNet. Most components have two descriptions: the *specification*, which defines the component, its relations and properties, and *the set of axioms*, which support automated reasoning.

By searching The Component Library we can find general definitions of some concepts, and relations to other concepts, such as its superclasses and subclasses. It may eventually be used to confirm some WordNet and FrameNet relations.

**URL:** www.cs.utexas.edu/users/mfkb/RKF/clib.html
**Contents:**
**Main influences:** WordNet, FrameNet and VerbNet
**Strong point:**
**Main applications:** Natural Language Processing
**Language:** KM (Knowledge Machine)
**Availability:** free download

## 2.3 Common Sense Ontologies

Sometimes computers seem ingenuous because they lack very basic knowledge that it is obvious for human beings, such as "one lives no more when one dies". This is called *common sense*, and many researchers have long sought to grab this basic knowledge for years.

The interesting point of common sense is that formalize knowledge that is impossible to find in ordinary resources. That knowledge is a specially useful 'raw material' for creative systems and natural language generators.

### 2.3.1 ConceptNet and OMCSNetCPP

ConceptNet and OMCSNetCPP are commonsense ontologies built on Open Mind's knowledge base. Basically, their authors parsed, interpreted and

organized the textual sentences in order to build semantic networks and high level representations of the knowledge.

The system, produced by the MIT Media lab, have Python and Java APIs. OMCSNetCPP is the C++ implementation of ConceptNet.

Using these ontologies one can draw other conclusions than the ones available through the use of Open Cyc, as well as confirm its inferences.

| | |
|---|---|
| **URL:** | web.media.mit.edu/∼hugo/conceptnet |
| | www.eturner.net/omcsnetcpp |
| **Contents:** | 280,000 pieces |
| **Main influences:** | Open Mind |
| **Main applications:** | Natural Language Processing |
| **Language:** | Semantic networks |
| **Availability:** | Free download |

### 2.3.2 Cyc and OpenCyc

Cyc is, according to its authors, "the world's largest and most complete general knowledge base and commonsense reasoning engine". It contains commonsense knowledge, represented in a logical form. OpenCyc is the open-source version of Cyc, and contains most of the knowledge and information of Cyc.

By asking OpenCyc what it knows about a person, for instance, we find out things such as *all people are human* and *a person can have different marital status*, among other conclusions. This ontology can be used to draw conclusions such as: *if we have an entity which is a mother, then it is a female*.

There is yet another version of Cyc, which is free for research purposes: ResearchCyc.

| | |
|---|---|
| **URL:** | opencyc.org, cyc.com, and research.cyc.com |
| **Main influences:** | |
| **Strong point:** | |
| **Main applications:** | |
| **Language:** | CycL and an OWL version of OpenCyc |
| **Availability:** | Free download of OpenCyc |
| | ResearchCyc free upon request |

### 2.3.3 Open Mind Common Sense

Common sense database Open Mind has a goal somehow similar to OpenCyc's. The aproach is different, though. Open Mind can be taught by

anyone, and contains its knowledge in plain English sentences (e.g., *dogs cannot fly*), while Open Cyc's ones are formal representations.

The recopilation of assertions for OMCS-1 has been running from September 2000 to August 2002.

| | |
|---|---|
| **URL:** | www.openmind.org |
| **Contents:** | 456,195 assertions from 9296 different people (OMCS-1) |
| **Main influences:** | |
| **Main applications:** | |
| **Language:** | Plain English |
| **Availability:** | Free download |

### 2.3.4 ThoughtTreasure

ThoughtTreasure is yet another common sense ontology, completed with an inference system and a natural language interface [Mue98]. Although its knowledge base is much smaller than Open Cyc and Open Mind's ones, it is able to reason and make plans based on its knowledge, since it contains planning agents based on finite automata. For each plan it defines goals, which can only be achieved once some subgoals are also achieved. For instance, if someone *wants to watch TV*, and the TV set is unplugged, he/she must plug it first, since *a TV set is an electrical device*, and *electrical devices need to be plugged to work*.

| | |
|---|---|
| **URL:** | www.signiform.com/tt/htm/tt.htm |
| **Contents:** | 27,000 concepts and 51,000 assertions |
| **Main influences:** | |
| **Main applications:** | Natural Language Processing |
| **Language:** | Proprietary + CycL |
| **Availability:** | Free download |

## 2.4 Multi-Source Ontologies

Some Multi-Source ontologies are inspired on existing ones and others use them permanently as a powerful combination of knowledge resources.

### 2.4.1 Ibrow

Ibrow (Intelligent Brokering Service) is a system being developed for the websemantic, to solve problems by querying many different software components distributed over the Internet. The authors state that "a user [will] log[...] on to the IBROW server on the World-Wide-Web and enter[...] the

specification of the knowledge-intensive problem he or she wants to solve. For instance, this could be an engineering design problem. The broker will then examine the available libraries of software components and configure a suitable problem solver for the problem in question."

Ibrow relies on more than 90 shared ontologies, where it searches for solutions for input problems that best match a set of known facts (observables), as in Case Based Reasoning. These ontologies, though, are specific to some fields, and not general ones, as the WordNet or the FrameNet.

| | |
|---|---|
| **URL:** | kmi.open.ac.uk/projects/ibrow |
| **Contents:** | From more than 90 ontologies |
| **Main influences:** | |
| **Main applications:** | Semantic web |
| **Language:** | Proprietary |
| **Availability:** | Part of |

### 2.4.2 Multi-Source Ontology MSO / AnyKB

AnyKB is the ongoing project of creating yet another ontology, by integrating information from existing ones. The author is integrating SUMO, Dolce and WordNet, among other smaller ontologies.

The main difference between this ontology and the others is that it is has only a minimal "core" of primitives, and all the other contents are introduced and checked by the users, through an web interface. MSO is not open source, AnyKB is supposed to be an open-source version of the former.

| | |
|---|---|
| **URL:** | meganesia.int.gu.edu.au/∼phmartin/WebKB/doc/MSO.html |
| **Contents:** | |
| **Main influences:** | WordNet, DOLCE and others |
| **Main applications:** | |
| **Language:** | FS (For Structuration) |
| **Availability:** | In the future |

### 2.4.3 SENSUS and Omega

SENSUS [AAH+01] is a terminology taxonomy, as a framework into which additional knowledge can be placed.

This ontology was built by extending and reorganizing the WordNet. Its top-level nodes include those from the Penman Upper Model (PUM)[4], and lower level ones were reorganized to fit the new taxonomy. SENSUS was

[4]More on the PUM in section 2.1.2.

developed with the initial goal of building a single ontology, out of several existing ones, such as the WordNet and OpenCyc, and later the idea of creating a framework that can be extended with additional knowledge was included. Cross-ontology alignment algorithms were developed, in order to make it possible to transfer knowledge between the ontologies. Many errors and omissions were discovered in various ontologies, leading to extensions in some of them.

The SENSUS project, however, was abandoned, and the Omega ontology was built after it, containing even more information than its predecessor—still from other ontologies, such as the Framenet, PropBank[5] and other ontologies.

| | |
|---|---|
| **URL:** | www.isi.edu/natural-language/projects/ONTOLOGIES.html<br>omega.isi.edu |
| **Contents:** | 70,000 nodes |
| **Main influences:** | Penman Upper Mode and WordNet |
| **Main applications:** | |
| **Language:** | Proprietary |
| **Availability:** | Probably soon |

## 2.5 Language Resources

Some language resources are very complete, including not only data about the language, but also important pieces of knowledge, thus fading the barrier between ontologies and databases of linguistic information.

### 2.5.1 FrameNet

The Berkeley FrameNet project is a lexicon-building effort in which the authors (1) study words, (2) describe the frames or conceptual structures which underlie these, (3) examine sentences using a very large corpus of contemporary English that contains these words, and (4) record the ways in which information from the associated frames is expressed in these sentences.

Each semantic frame is essentially a representation of a situation type in a given domain (eating, writing, etc.), as well as its participants, properties and other conceptual roles which are part of that situation. A frame is usually "evoked" by a verb (e.g. the WRITING frame by the verb 'to write'). Some nouns taking arguments can also be frame evokers. Frame-relevant annotations are limited to the sentence-level, as opposed to "filling in all information about a situation from a multi-sentence text". The aim is

[5]The PropBank project is creating a corpus of text annotated with information about basic semantic propositions. More information at www.cis.upenn.edu/∼ace/.

to document the senses of frame-evoking words, and their respective range of semantic and syntactic combinatory possibilities.

This ontology also maps frame-to-frame relations, like Inheritance, Use, Subframe and Causative-of, among others.

By exploring Framenet, we can discover many characteristics of some given situation (for example, we can state that **writing** is a form of *intentionally creating something*, and uses the frame **communication**, thus needing a *topic*, etc.).

There's also a Spanish Framenet, maintained by Carlos Subirats, available at gemini.uab.es/SFN.

| | |
|---|---|
| **URL:** | framenet.icsi.berkeley.edu |
| **Contents:** | More than 625 semantic frames and 8,900 lexical units |
| **Main influences:** | Fillmore's Case Grammar, theoretical lexicography |
| **Main applications:** | Natural Language Processing |
| **Language:** | Proprietary |
| **Availability:** | Free download upon request |

### 2.5.2 LCS Database

Lexical Conceptual Structure (LCS) is a compositional abstraction with language-independent properties that transcend structural idiosyncrasies. An LCS is a directed graph with a root. Each node is associated with certain information, including a type, a primitive and a field [TH00].

LCS Database contains structures built by hand by its author in 1994, organized into semantic classes.

| | |
|---|---|
| **URL:** | www.umiacs.umd.edu/∼bonnie/ LCS_Database_Documentation.html |
| **Contents:** | |
| **Main influences:** | WordNet and others |
| **Main applications:** | Natural Language Processing |
| **Language:** | |
| **Availability:** | Free download |

### 2.5.3 PropBank

The PropBank project is creating a corpus of text annotated with information about basic semantic propositions. Predicate-argument relations are being added to the syntactic trees of the Penn Treebank.

| | |
|---|---|
| **URL:** | www.cis.upenn.edu/∼ace |
| **Contents:** | |

**Main influences:**
**Main applications:** Natural Language Processing
**Language:**
**Availability:** Possibly on request

### 2.5.4 Roget's Thesaurus

RT is a thesaurus of English concepts, available online for free querying, but not for download.

**URL:** thesaurus.reference.com
**Contents:** 17,000 concepts
**Main influences:**
**Main applications:** Natural Language Processing
**Language:**
**Availability:** Unavailable

### 2.5.5 The CMU Pronouncing Dictionary

It includes a text file about rhyming (over 125,000 words and their transcriptions) which indicates the syllable sounds, weak/strong stress, and an identifier at the end which allows you to query the list to find a set of rhyming words.

**URL:** www.speech.cs.cmu.edu/cgi-bin/cmudict
**Contents:**
**Main influences:**
**Main applications:** Speech, Poetry and Lyrics Generation
**Language:** Proprietary format (simple plain text) and Sphinx
**Availability:** Free download

### 2.5.6 The Penn TreeBank Project

The Penn Treebank Project annotates naturally-occuring text for linguistic structure. It produces skeletal parses showing rough syntactic and semantic information—a bank of linguistic trees.

**URL:** www.cis.upenn.edu/∼treebank
**Contents:**
**Main influences:**
**Main applications:** Natural Language Processing
**Language:**

**Availability:** Available for members of the LDC

### 2.5.7 Unified Verb Index

Unified Verb Index combines information from the VerbNet, PropBank, and FrameNet projects, and it is available on the Internet.

**URL:** www.cs.rochester.edu/∼gildea/Verbs
**Contents:**
**Main influences:**
**Main applications:** Natural Language Processing
**Language:**
**Availability:** Available in HTML

### 2.5.8 VerbNet

VerbNet is a verb lexicon with syntactic and semantic information for English verbs. For each syntactic frame in a verb class, there is a set of semantic predicates and relations associated with it. For instance, searching VerbNet we can state that *to write* is related to other verbs in the same subclass, such as *to paint* and *to draw*, and is also related to a transfer of information.

**URL:** www.cis.upenn.edu/group/verbnet/home.html
**Contents:**
**Main influences:**
**Strong point:** Verbs
**Main applications:** Natural Language Processing
**Language:** Proprietary
**Availability:** Free download

### 2.5.9 Visual Thesaurus

The Visual Thesaurus is a dictionary and thesaurus with an intuitive and interesting visual interface that encourages exploration and learning. Designed for the user to improve vocabulary and understanding of the English language, there are two commercial versions: Desktop Edition and Online Edition.

**URL:** www.visualthesaurus.com
**Contents:** 145,000 English words and 115,000 meanings
**Main influences:**
**Main applications:** Human Learning

**Language:**
**Availability:** Commercial

### 2.5.10 WordNet and EuroWordNet

WordNet [MBF+90] is an English dictionary containing nouns, verbs, adjectives and adverbs, organized into synonym sets, each representing one underlying semantic concept. Different relations link the synonym sets, namely synonymy, hypernymy, holonymy, meronymy, hyponymy and troponymy, among others. For each concept, WordNet contains a gloss (definition), and a set of relations to other words.

This dictionary is widely used for Natural Language Processing, for it is very complete in what concerns to lexical relations, it's free and very easy to use.

WordNet may be used to extract lexical relations for some entities, because as well as linguistic categories, it includes semantic relations like synonyms, antonyms, hypernyms (*concepts of*), hyponyms (*instances of*, *is-a*), holonym (whole of) and meronym (part of).

The main drawbacks of WordNet are: 1) it doesn't contain many other important semantic relations (i.e., there's no relation between table and chair.), 2) its definitions are short textual explanations, which need to be parsed and semantically interpreted for any possible automatic processing[6].

**URL:** wordnet.princeton.edu
**Contents:** 115,424 synsets
**Main influences:**
**Strong point:** Widely used for its versatility and availability
**Main applications:** Natural Language Processing
**Language:** Proprietary
**Availability:** Free download

There is a commercial version of WordNet for European languages called EuroWordNet. EuroWordNet was a two-phased European research project building wordnets for the following European languages: Czech, French, English, Estonian, Spanish, German, Italian, Dutch. Starting from the idea of synsets and relations, the EuroWordNet team built its own resources including

- a proprietary database system;

[6]Extended WordNet, an ongoing project at the University of Texas at Dallas, intends to do some parsing and disambiguation of the WordNet glosses (xwn.hlt.utdallas.edu [March, 25, 2005]).

- an extended set of relations (at least their definitions, not all of them have been implemented for all languages);
- a unified upper-level concept structure, called Inter-Lingual Index (ILI), which comes very close to an intercultural ontology. The ILI is mapped to synsets in the individual languages, which allows for a cross-language comparison of synsets and lexicalisations.

EuroWordNet lexical databases are available from ELRA/ELDA, against a fee. Documentation is freely available online. The EuroWordNet effort was continued and extended by BalkaNet. BalkaNet developed another WordNet browser, VisDic/DEB, which uses EuroWordNet data in XML.

# Chapter 3

# Languages for Ontologies

Every ontology needs a formal paradigm of representation. Many ontology authors wrote their own languages. The most widely used paradigms include predicate logic, description logic and frame systems.

## 3.1 ClearTalk

ClearTalk is a set of conventions for stating information clearly and concisely. It is a compromise between uncontrolled English and excessively formal languages, e.g. those based on logic. It is designed so that the computer may be able to do a significant amount of "intelligent" processing with this information, which would be very difficult or impossible without such conventions.

ClearTalk was developed to enter information into a knowledge management system—namely IKARUS[1], that will be able do some semantic processing on ClearTalk statements. ClearTalk can beviewed as an idea for introducing structure into normally unstructured text.

**URL:** www.csi.uottawa.ca/∼kavanagh/Ikarus/Cleartalk.html
**Main influences:**
**Inference engine:**
**Ontologies:**

[1]Ikarus is a knowledge management system with a World Wide Web interface. More details can be found at www.csi.uottawa.ca/∼kavanagh/Ikarus/IkarusInfo.html (April 16, 2005).

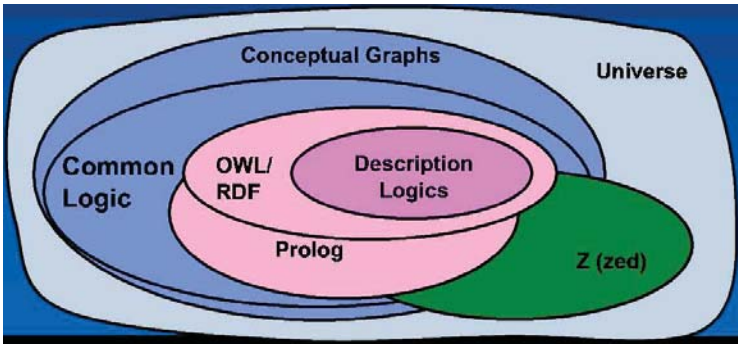


Figure 3.1: Domain coverage of some of the most important languages (Based on the one by Harry Delugach, Presentation to ISO/IEC JTC1 SC32 Open Forum, Berlin, Germany, April 2005 [available at cl.tamu.edu/docs/cl/Berlin_OpenForum_Delugach.pdf].).

## 3.2 Common Logic

Common Logic (CL) is a recent development of previous logic formalisms, namely KIF and Conceptual Graphs, which intends to have a broader coverage of concepts. CL is being proposed as an ISO standard.

**URL:** philebus.tamu.edu/cl
**Main influences:** KIF and Conceptual Graphs
**Inference engine:**
**Ontologies:**

## 3.3 Common Logic Controlled English

Common Logic Controlled English (CLCE) is a formal language with an English-like syntax. Anyone who can read ordinary English can read sentences in CLCE with little or no training. Writing CLCE, however, requires practice in learning to stay within its syntactic and semantic limitations. Formally, CLCE supports full first-order logic with equality supplemented with an ontology for sets, sequences, and integers. The fundamental semantic limitation of CLCE is that the meaning of every CLCE sentence is defined by its translation to FOL; none of the flexibility of ordinary English and none of its metaphorical or metonymic extensions are supported. The primary syntactic restrictions are the use of present tense verbs and singular nouns, variables instead of pronouns, and only a small subset of the many syntactic options permitted in English. Despite these limitations, CLCE can express the kind of English used in software specifications, textbooks of mathematics, and the definitions and axioms of formal ontology.

**URL:** www.jfsowa.com/clce/specs.htm
**Main influences:**
**Inference engine:**
**Ontologies:**

## 3.4 Concept Modeling Language

CML is the modelling language developed for CommonKADS, a methodology developed to support structured knowledge engeneering.

**URL:** www.educery.com/papers/educe/models/cml.htm
www.commonkads.uva.nl
**Main influences:**
**Inference engine:**
**Ontologies:**

## 3.5 CGIF

Conceptual Graphs (CG) are a tool to represent meaning in a form that is logically precise, humanly readable, and computationally tractable. With their direct mapping to language, conceptual graphs serve as an intermediate language for translating computer-oriented formalisms to and from natural languages[2].

CGIF (CG Interchange Format) is intended to be one of the standard notations for exchanging knowledge and is the official textual notation for Conceptual Graphs.

**URL:** www.jfsowa.com/cg
meganesia.int.gu.edu.au/∼phmartin/WebKB/doc/CGIF.html
**Main influences:** Existential graphs (C. S. Peirce) and Semantic Networks
**Inference engine:**
**Ontologies:** WebKB-2

## 3.6 CycL

CycL is a formal language developed to represent common sense in all the Cyc projects. Its syntax was inspired in first-order predicate calculus and

[2]Extracted from http://www.cs.uah.edu/∼delugach/CG/Sowa-intro.html (April 23, 2005).

Lisp. It goes, however, far beyond first order logic. The vocabulary of CycL consists of terms. The set of terms can be divided into constants, non-atomic terms (NATs), variables, and a few other types of objects. Terms are combined into meaningful CycL expressions, which are used to make assertions in the CYC knowledge base.

**URL:** www.cyc.com/cycdoc/ref/cycl-syntax.html
**Main influences:** First-order predicate calculus and Lisp
**Inference engine:**
**Ontologies:** Cyc, OpenCyc, ResearchCyc, ThoughtTreasure

## 3.7 DAML and DAML+OIL

DARPA Agent Markup Language (DAML) was developed as an effort to facilitate the concept of the Semantic Web. Being built after XML and RDF, it is more powerful than these languages for unambiguously representing data and information in texts and ontologies.

The latest release of the language (DAML+OIL) provides a rich set of constructs with which to create ontologies and to markup information so that it is machine readable and understandable.

**URL:** www.daml.org/index.html
**Main influences:** XML, RDF and OIL
**Inference engine:**
**Ontologies:**

## 3.8 EXPRESS

EXPRESS is a language that can be exported, pretty-printed and graphically browsed as EXPRESS-G.

There is a public-domain TNO EXPRESS parser that includes an option to translate EXPRESS to Prolog clauses. The toolkit can import EXPRESS by translating these Prolog clauses to the internal representation. The toolkit also supports loading a STEP Physical File (instances of EXPRESS data models) and linking these instances to the data model.

**URL:** ftp.cme.nist.gov
**Main influences:**
**Inference engine:**
**Ontologies:**

## 3.9 F-Logic

Frame Logic (F-logic) [KW95] is a clean and declarative fashion for most of the structural aspects of object-oriented and frame-based languages. These features include object identity, complex objects, inheritance, polymorphic types, query methods, encapsulation, and others. In a sense, F-logic stands in the same relationship to the object-oriented paradigm as classical predicate calculus stands to relational programming. F-logic has a model-theoretic semantics and a sound and complete resolution-based proof theory. A small number of fundamental concepts that come from object-oriented programming have direct representation in F-logic. Other secondary aspects of this paradigm are easily modeled as well.

**URL:** www.informatik.uni-freiburg.de/∼dbis/Publications/95/flogic-jacm.html
**Main influences:**
**Inference engine:**
**Ontologies:**

## 3.10 Flora-2

FLORA-2 is an object-oriented knowledge base language and application development environment. The programming language of FLORA-2 is a dialect of F-logic with numerous extensions, including meta-programming in the style of HiLog and logical updates in the style of Transaction Logic. FLORA-2 was designed with extensibility and flexibility in mind, and it provides strong support for modular software design through its unique feature of dynamic modules.

**URL:** flora.sourceforge.net
**Main influences:** F-Logic
**Inference engine:**
**Ontologies:**

## 3.11 KIF

Knowledge Interchange Format (KIF) [GF92] is a computer-oriented language for the interchange of knowledge among disparate programs. 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 *non-monotonic reasoning rules*, and it provides for the *definition of objects, functions, and relations*.

KIF was created as part of the Ontolingua[3] project, and was one of the most widely used languages for knowledge representation.

**URL:** www-ksl.stanford.edu/knowledge-sharing/kif
**Main influences:**
**Inference engine:**
**Ontologies:** Frame Ontology

## 3.12 Knowledge Machine

Knowledge Machine (KM) is a powerful, frame-based language with clear first-order logic semantics. It contains sophisticated machinery for reasoning, including selection by description, unification, classification, and reasoning about actions using a situations mechanism. Its origins were the Theo language and the (now obsolete) language KRL.

**URL:** www.cs.utexas.edu/users/mfkb/km.html
**Main influences:** Theo and KRL
**Inference engine:**
**Ontologies:** The Component Library

## 3.13 KQML

Knowledge Query and Manipulation Language (KQML) is a language and protocol for exchanging information and knowledge. It is part of a larger effort, the ARPA Knowledge Sharing Effort, which is aimed at developing techniques and methodology for building large-scale knowledge bases which are sharable and reusable. KQML is both a message format and a message-handling protocol to support run-time knowledge sharing among agents. KQML can be used as a language for an application program to interact with an intelligent system or for two or more intelligent systems to share knowledge in support of cooperative problem solving.

KQML focuses on an extensible set of performatives, which defines the permissible operations that agents may attempt on each other's knowledge and goal stores. The performatives comprise a substrate on which to develop

[3]Ontolingua is a collaborative environment to browse, edit and create ontologies. It is available at www.ksl.stanford.edu/software/ontolingua (April 17, 2005).

higher-level models of inter-agent interaction such as contract nets and negotiation. In addition, KQML provides a basic architecture for knowledge sharing through a special class of agent called communication facilitators which coordinate the interactions of other agents.

**URL:** www.cs.umbc.edu/kqml
**Main influences:**
**Inference engine:**
**Ontologies:**

## 3.14 LOOM

Loom [MB87] is a knowledge representation language developed by researchers in the Artificial Intelligence research group at the University of Southern California's Information Sciences Institute.

The heart of Loom is a knowledge representation system that is used to provide deductive support for the declarative portion of the Loom language. Declarative knowledge in Loom consists of definitions, rules, facts, and default rules. A deductive engine called a classifier utilizes forward-chaining, semantic unification and object-oriented truth maintainance technologies in order to compile the declarative knowledge into a network designed to efficiently support on-line deductive query processing.

**URL:** www.isi.edu/isd/LOOM/LOOM-HOME.html
**Main influences:**
**Inference engine:**
**Ontologies:** GUM

## 3.15 OIL

The Ontology Inference Layer (OIL) is a proposal for a web-based representation and inference layer for ontologies, which combines the widely used modelling primitives from frame-based languages with the formal semantics and reasoning services provided by description logics. It is compatible with RDF Schema (RDFS), and includes a precise semantics for describing term meanings (and thus also for describing implied information).

OIL presents a layered approach to a standard ontology language. Each additional layer adds functionality and complexity to the previous layer. This is done such that agents (humans or machines) who can only process a lower layer can still partially understand ontologies that are expressed in any of the higher layers.

**URL:** www.ontoknowledge.org/oil
**Main influences:** RDFS and Lisp
**Inference engine:**
**Ontologies:**

## 3.16 Ontolingua

Ontolingua is an extension of KIF and FrameOntology (Gruber 93), usually known as a "standard language" for ontology representation some years ago.

The Ontolingua source release previously available has been removed because it is now no longer supported. Now there is an interactive network service version of Ontolingua instead called Stanford KSL Network Services.

**URL:** ontolingua.stanford.edu
**Main influences:** Lisp
**Inference engine:**
**Ontologies:**

## 3.17 OWL

The Ontology Web Language (OWL) can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms, namely in the semantic web environment.

OWL has more facilities for expressing meaning and semantics than XML, RDF and RDF-S—thus OWL goes beyond these languages in its ability to represent machine interpretable content on the Web, providing additional vocabulary. OWL is a revision of the DAML+OIL web ontology language, incorporating lessons learnet from the design and application of DAML+OIL. It has three increasingly-expressive sublanguages: OWL Lite, OWL DL, and OWL Full.

Although OWL is not as powerful as other languanges (KIF, Common Logic, etc.), it is quite standard, and there are many parsers and other tools available.

**URL:** www.w3.org/TR/owl-features
**Main influences:** DAML+OIL, XML and RDF
**Inference engine:**
**Ontologies:** DOLCE, GUM and OpenCyc

## 3.18 Sphinx

Sphinx knowledge base tools is set of tools for creating dictionaries, language models and for conditioning text corpora.

**URL:** www.speech.cs.cmu.edu/tools
**Main influences:**
**Inference engine:**
**Ontologies:** The CMU Pronouncing Dictionary

## 3.19 SUO-KIF

As refered in section 2.1.3, IEEE Standard Upper Ontology Working Group (SUO WG) is working on standard ontologies and languages. SUO-KIF is a KIF based language which will probably be considered the IEEE standard to specify the syntax of a general-purpose knowledge representation language. The primary purpose of this effort is to support the SUO project.

**URL:** suo.ieee.org
**Main influences:** KIF
**Inference engine:**
**Ontologies:** SUMO and DOLCE

## 3.20 Triple

TRIPLE is an RDF query, inference, and transformation language for the Semantic Web. Based on F-Logic, RDF and Horn Logic, this language was developed in a modular and layered architecture, thus showing good compatibility and scalability.

**URL:** triple.semanticweb.org
**Main influences:** RDF, F-Logic, Horn Logic and SiLRi
**Inference engine:**
**Ontologies:**

## 3.21 XML and XMLS, RDF and RDF-S

Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML. Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important

role in the exchange of a wide variety of data on the Web and elsewhere. XML is a standard W3C language for data modeling.

**URL:** www.w3.org/XML
**Main influences:** SGML and HTML
**Inference engine:**
**Ontologies:**

Another proposal of the W3C, Resource Description Framework (RDF) is a meta-data modeling language, developed with the aim of representing knowledge, while CML is intended to represent documents. Its semantics much more powerful than XML for meta-data, and its syntax can still be represented in XML.

**URL:** www.w3.org/RDF
**Main influences:** XML
**Inference engine:**
**Ontologies:**

The importance of ontologies to the Semantic Web has prompted the development of schema extensions to existing Web standard languages: XML has been extended to support defined schemas (XML-Schema, XMLS), while RDF has been extended to support RDF-Schema (RDFS). Although the language primitives provided by these standards are great improvements compared to original XML and RDF, they are still extremely basic when compared with those typically provided by ontology languages developed within the Knowledge Representation (KR) community.

**URL:** www.w3.org/XML/Schema
www.w3.org/RDF/#schemas
**Main influences:** XML and RDF
**Inference engine:**
**Ontologies:**

# Acknowledgements

The authors want to thank their respective research groups and supervisors for supporting the process of writing this report. Special thanks also to Birte Lönneker for her contributions to the document.

# Bibliography


[AAH+01] J. L. Ambite, Y. Arens, E. H. Hovy, A. Philpot, L. Gravano, V. Hatzivassiloglou, and J.L. Klavans. Simplifying data access: The energy data collection project. *IEEE Computer*, 32(2), February 2001.

[BHR95] John A. Bateman, Renate Henschel, and Fabio Rinaldi. Generalized upper model 2.0: documentation. Technical report, GMD/Institut für Integrierte Publikations-und Informationssysteme, Darmstadt, Germany, 1995.

[BKMW90] John A Bateman, Robert T Kasper, Johanna D Moore, and Richard A Whitney. A general organization of knowledge for natural language processing: The penman upper model. Technical report, USC/Information Sciences Institute, 1990.

[CP97] Peter Clark and Bruce Porter. Building concept representations from reusable components. In *Proceedings of the American Association for Artificial Intelligence (AAAI'97)*, 1997.

[GF92] Michael R. Genesereth and Richard E. Fikes. *Knowledge Interchange Format Version 3.0 Reference Manual*. Computer Science Department of the Stanford University, Stanford, California, USA, 1992.

[HB94] Renate Henschel and John A Bateman. The merged upper model: A linguistic ontology for german and english. In *COLING*, Kyoto, Japan, 1994.

[HM99] Michael A K Halliday and Christian M I M Matthiessen. *Construction experience through meaning: A language-based approach to cognition*. Casell, London, 1999.

[KW95] Lausen M. Kifer and J. Wu. Logical foundations of object oriented and frame-based languages. *Journal of the ACM*, 42, 1995.

[MB87] Robert MacGregor and Raymond Bates. The loom knowledge representation language. Technical report, USC Information Sciences Institute, Marina del Rey, CA, 1987.

[MBF+90] George A. Miller, Richard Backwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. Introduction to wordnet: An on-line lexical database. *Journal of Lexicography*, pages 234–244, 1990.

[Mue98] Erik T. Mueller. *Natural language processing with ThoughtTreasure*. Signiform, New York, 1998.

[TH00] David Traum and Nizar Habash. Generation from lexical conceptual structures. In *Workshop on Applied Interlinguas*, Seattle, WA, USA, 2000. ANLP-2000.