ONTOLOGY: A SURVEY

Christophe Roche

University of Savoie
Equipe Condillac - Campus Scientifique
73 376 Le Bourget du Lac cedex – France
christophe.roche@univ-savoie.fr - http://ontology.univ-savoie.fr

Abstract: During the last decade, the explosive growth of information technologies led to a shift in the market and economic view of the society: communication and knowledge sharing became the new economic stakes. But everyone speaks his own language, with his own terms and meanings. Ontologies seem to be one of the most suitable solutions faced with this problem and have become a very popular research topic in knowledge representation. But several problems remain which claim for clarification. The main objective of this survey is to make explicit the main questions about ontology and to draw some guidelines about the possible answers.

Keywords: ontology, terminology, artificial intelligence, semantic web, collaborative engineering, corporate memory, knowledge engineering, knowledge representation.

1. WHY ONTOLOGIES ?

1.1 From the Information Society to the Knowledge Society

The increasingly global economy, the explosive growth of the "world" in the "world wide web", the constantly moving market characterised by short product life cycles and increased demand for flexibility as well as the extensive use of information have led to a new vision of the society: the Information Society. A new socio-economic system which relies more and more on circulation, sharing and exchanging information whose repercussions modify daily life, culture, economy and industry. Collaborative Engineering based on co-operation and collaboration of multi-disciplinary people is a good example. Collaborators, from design to manufacturing, have to share and exchange information: client requirements, business information, simulation results, workshop loads, supplier delays… as they have to co-ordinate their decisions in order to react as quickly as possible to changes. The Semantic Web, a new form of web content which is meaningful to computers, is another and more popular example. The goal is here for software agents to better understand and process the data the web pages contain. Communication, knowledge sharing and exchanging are the new economic stakes of the Information Society.

1.2 The Communication & Knowledge Sharing Problem

The Information Society (virtual enterprise, e-business, etc.) relies on communication between of interacting and heterogeneous actors: people, organisations and even software systems (let us notice that software systems often aggravate the problem of communication by isolating information and using their own protocols). But everybody speaks his own language with his own terms and meanings and no communication is possible, and all the more so no co-operation or collaboration, if one can not understand each other: the Information Society is the new Tower of Babel.

Communication between people, organisations and software systems is difficult due to the fact that each of these actors speaks a different language. To address this problem, we need a common communication language that agents can read and

understand. Using a single, normalised language like KQML (Knowledge Query Manipulation Language) can reduce the gap of misunderstanding by using a same syntax. But, although such languages give some useful indications about the pragmatic content of the message, the semantic problem has still to be addressed. As a matter of fact, two entities, either people or software agent, can communicate only if they agree upon on the meaning of the terms they use, and the problem becomes more complex when one takes into account multilingualism.
Let us notice that communication cannot be reduced to only exchanging data but must take into account the exchange of knowledge. Co-workers in an enterprise work at their own levels, using their own knowledge and engineering models. Furthermore, the software tools extensively used in information societies, requiring specific and dedicated representations, are more concurrent than collaborative. The way to address this problem is to define a shared understanding.

Agreement must be achieved about the shared knowledge used as a communication medium among people and software tools.

Ontology, understood as an agreed vocabulary of common terms and meanings shared by a group of people, is a solution to that problem. The reason for ontologies are so popular is in large part due to what they promise: "*a shared and common understanding of some domain that can be communicated across people and computers*."

2. WHAT IS ONTOLOGY?

It is amusing to notice that when the main goal of ontology is to normalise the meaning of terms, the term "ontology" itself is not clearly defined: "*although* ontology *is currently a fashionable term, no agreement exists on the exact meaning of the term" and "seems to generate a lot of controversy in discussion about AI (artificial intelligence)*" (Gruber 1995). In fact, ontology finds applicability in many domains of application in knowledge and software engineering, and each of them gives its own definition.

2.1 A bit of etymology and philosophy.

The word "ontology" has a very long history in philosophy starting with the Aristotle's works. Defined as "the science of being", it comes from the Greek "ontos" which means *being* and "logos" for both *language* and *reason*. Ontology is then the branch of metaphysics that deals with the nature of being. From the point of view of phenomenology, a more modern philosophy started with the 19th century German philosophers, an ontology is a systematic account of existence. This going back to things (the phenomenological approach) does not mean, in spite of being and existence are different notions, that theses two approaches can not be combined (think about the phenomenological reduction).

To keep in mind the epistemological definition may help us to better understand what ontology should be. As a matter of fact, one can retain three dimensions: *knowledge*, *language* and *logic*. In others words: *language* to speak about the world, *conceptualisation* for understanding the world and *representation* for the manipulation of our understanding.

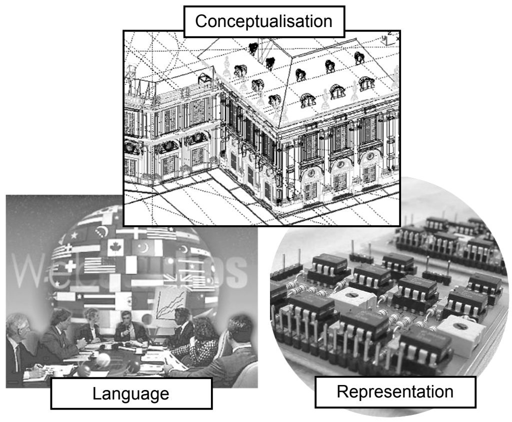

Fig.1 The semiotic triangle.

If these three dimensions are strongly linked, they must not be confused. In particular, it does not mean that the language used for representation is the right language to speak about the world and that the underlying primitives of this representation language can be used as concepts for understanding the world. Logic-based languages, first-order or description logic, are good examples of this problem.

2.2 The knowledge engineering point of view.

The software and knowledge engineering communities borrowed the term of ontology[1] at the beginning of the nineties focusing more on a systematic account of existence rather than a metaphysical approach of the nature of being. As a matter of fact, for artificial intelligence systems, what exists is that which can be represented in a declarative language. Ontology is then an explicit formal specification of how to represent objects, concepts and relationships that are assumed to exist in some area of interest. It is what Gruber called "*a specification of a conceptualisation*" that is a

[1] "The use of the term *ontology* is somewhat unfortunate since it has a definite meaning in the philosophical literature which has little to do with describing the content of information repositories". KACTUS - Esprit Project 8145.

description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents.

To resume the most definitions used in knowledge engineering including semantic web we can say that: "*an ontology is a shared description of concepts and relationships of a domain expressed in a computer readable language*". The Fig.2 is an example from the generic Upper Cyc Ontology.

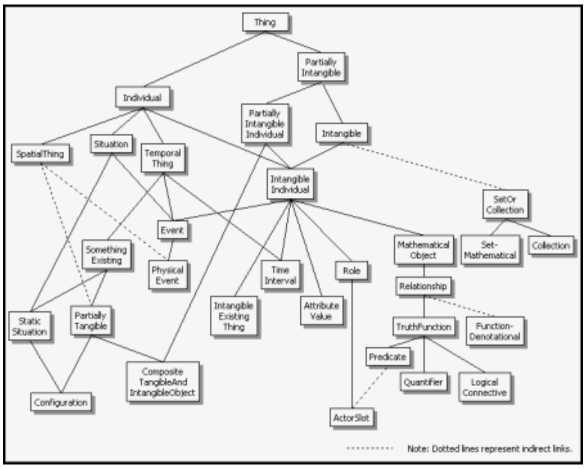

Fig.2 The Upper Cyc Ontology.

3. WHAT IS ONTOLOGY FOR?

"*What is important is what an ontology is for*" (T.R Gruber). As a matter of fact, the ontology's requirements depend on the purpose of the final application. The main objective of an ontology is to enable communication and knowledge sharing between computer systems by capturing a shared understanding of terms that can be used by humans and programs. We can then identify different categories of uses for ontologies:

- communication between people and organisations;
- inter-operability (communication) between systems, including enterprise modelling and multi-agent systems;
- system engineering, for specification, reliability and reusability;
- knowledge management like information retrieval, document management, knowledge base systems;
- natural language treatment for semantic analysis, lexical structure;

and all applications in relation to internet: electronic commerce and semantic web. Let us say few words about this last one which well illustrates the actual ontology trends. Facing the increasing use of the web, more humans and computer programs need to interact with data on websites. A new form of web content defined as an extension of the current one was required. *"The Semantic Web is the representation of data on the World Wide Web"*. (W3C's definition). The semantic web is based on the Resource Description Framework (RDF) for semantics, XML for syntax and URIs for naming.

4. WHAT PROPERTIES?

Ontologies are agreements about shared conceptualizations of a domain (*ontological commitment*) and the reason for why ontologies are so popular is in large part due to what they promise: "a shared and common understanding of some domain that can be communicated across people and computers" (http://www.ontology.org).

This quest for the Holy Grail requires properties which it is difficult (impossible?) to gather into a same context; unless to normalize language and thought. As a matter of fact, in order to be used, re-used and shared, an ontology must be *consensual*, *coherent* and *precise*. Is using a same representation language with a clear syntax and semantic, like logic-oriented languages, a solution? Before to see this representation language problem which should be also an epistemological and linguistic problem, let us say few words about what ontology looks like.

5. WHAT DOES ONTOLOGY LOOK LIKE?

"*An [explicit] ontology may take a variety of forms, but necessarily it will include a vocabulary of terms and some specification of their meaning (i.e. definitions)."* (Ushold and al. 96).

The definition of concepts and properties can take different forms, from informal definitions expressed in natural language to definitions stated in very formal languages as first order logic. It depends on the intended use of the final application, from shared glossaries among users to more formal requirements such as inter-operability between software. Including the type of knowledge, more or less specific or generic, we obtain dimensions along which ontologies vary and can be classified. As the purpose dimension has been presented in §3, let us see the formality and type of knowledge dimensions.

5.1 Formality.

This dimension is based on the degree of formality by which the knowledge is specified. An ontology is said :

- "highly informal" if the definitions are expressed in natural language, e.g. glossaries;
- "semi-informal" if a restricted and structured form of natural language is used in order to reduce ambiguity;

- "semi-formal" if the definitions are expressed in an artificial formally defined language, e.g. frame languages (see §7);
- "rigorously formal" when the definitions are precisely defined with a formal semantics, e.g. logic-based languages.

5.2 Type of knowledge.

According to the type of the knowledge which is conveyed by the ontology, this one will be qualified as being:

- a generic ontology, also called top ontology, specifies general concepts defined independently of a domain of application and which can be used in different application domains. *Time, space, mathematics…* are examples of general concepts;
- a domain ontology is dedicated to a particular domain which remains generic for this domain and which can be used and reused for particular tasks in the domain. *Chemical, medicine (UMLS), enterprise modelling…* are domain ontologies.
- an application ontology gathers knowledge dedicated to a particular task including more specialised knowledge of the experts for the application. In general it is not reusable.
- a meta-ontology, or representation ontology, specified the knowledge representation principles used to define concepts of domain and generic ontologies e.g. what is a class, a relation, a function…

6. A DEFINITION

If we try to sum up the previous paragraphs, we can say that ontology requires a multidisciplinary approach:

- linguistics, as we use words to communicate;
- epistemology, since words refer to knowledge (in general concepts) which represents their meaning;
- logic, in order to guarantee some coherence;
- knowledge representation in order to manipulate our understanding of the domain.

what we can resume by the following definition:

« An ontology is a conceptualisation of a domain to which one or several vocabularies can be associated and which participates to the meaning of terms. Defined for a given objective, an ontology expresses a point of view shared by a community. An ontology is represented in a language (explicit ontology) whose theory (semantics) guarantees the properties of the ontology in terms of consensus, coherence, sharing and reuse. »

7. WHAT METHODOLOGY?

Building ontology raises a lot of questions from methodology to tools to be used. It is a difficult and crucial problem (Uschold *et al.* 95); and although some useful guidelines have been defined the problem remains complex (Uschold *et al.95*) (Gruber 95): "*Ontologies are becoming increasingly popular in practice, but a principled methodology for building them is still lacking.*" (Guarino *et al. 00*).

As software, ontologies have to be specified, implemented and maintained under the control of a project management defined in terms of planning, control and quality. In a way very similar to IEE software development process we can define an ontology-oriented engineering where the ontology's life cycle, including ontology capture and coding phases, is embedded in a dedicated development process and where the used techniques are knowledge-dedicated. The FIPA (Foundation for Intelligent Physical Agents) ontology methodology as well as Methontology, a methodology specified by the University of Madrid, are some examples.

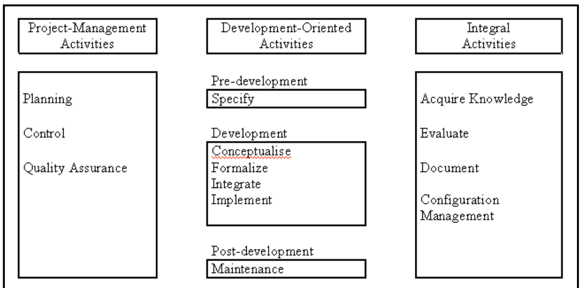

Figure 12: Ontology development process (proposition from [1])

Fig.3 The FIPA Ontology Service Specification.

The development phase of the ontology's life cycle is an acquisition knowledge problem that can be split into interlaced and incremental processes: extracting potential terms, finding objects, concepts and relationships, producing definitions… The Uschold and King's methodology proposes the following steps: identify purpose and scope, building the ontology (i.e. ontology capture, ontology coding and integrating existing ontologies), evaluation and documentation. Let us notice that starting by analysing knowledge sources or finding potential terms using text mining techniques on documents can well kick off the process of this phase.

8. WHAT LANGUAGES?

There are a lot of potential representation languages for ontology definition, from natural language for highly informal ontologies to formal languages for rigorously formal ontologies: *Ontolingua, KIF, Frame Ontology, OKBC, OCML, Flogic, LOOM, SHOE, RDF, RDFS, XOL, OIL, DAML, OWL…*

The computational languages, those which interest us here, can be classified according to different point of views: knowledge representation (historically the first one) logic and more and more the world wide web. But it is important to keep in mind that thought and language are strongly linked. As a matter of fact, one does not define in a same way a concept in natural language, in calculus predicate or in a frame system (recall in mind the Sapir-Whorf hypothesis).

8.1 Logical languages.

The logical languages, first-order logic and description logic, fit very well the ontology's objectives. They rely on a clear syntax and a formal semantics and provide sound inference mechanisms. So, we can hope consensual, coherent, precise and sharable knowledge bases. Furthermore, logical languages are computational and can be used as interchange languages.

Concepts, considered as intensional definitions of sets, as well as relationships are defined as logical well formed formulas. Let us take the following example taken from the Sowa'ontology:

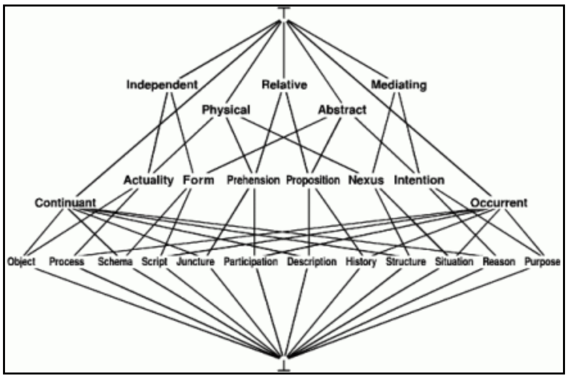

Fig.4 The Sowa's Ontology.

A concept (category) is an unary predicate and a new concept is defined as a conjonction of higher unary predicates. For example:

$$\text{'form(x)'} = \text{independant}(x) \wedge \text{abstract}(x)$$

But, some problems remain, especially epistemological and linguistic problems. A set is not a concept, even if a concept can be understood as the set of individuals subsumed by the concept. And unlike properties which can be true or false, a concept is not a well formed formula, it is not true or false, it is. In fact, if logic is necessary in order to ensure soundness of knowledge bases, it is necessary *a posteriori* but not *a priori*.

8.2 Frame languages.

Frame languages have a very long history in artificial intelligence and knowledge representation (starting with the Minsky's works in 1974). These languages provides most the modeling primitives commonly used in ontology. Frames, i.e. concepts or classes, are defined by attributes (the slots) and structured according to a subclass relationship into graphs or taxonomies. Among the numerous possible frame languages, let us consider two of them as examples.

<u>*Ontolingua*</u>

Ontolingua was developped by the Knowledge Systems Laboratory of Stanford University in 1992. Defined on the top of KIF (Knowledge Interchange Format) it provides forms for defining classes, relations, functions, objects and axioms.

```
(define-class AUTHOR (?author)
"An author is a person who writes things…"
:def (and
        (person ?auhor)
        (= (value-cardinality ?author AUTHOR.NAME) 1)
        (value-type ?author AUTHOR.NAME biblio-name)
        (>= (value-cardinality ?author …)…)
```

<u>*Protégé 2000*</u>

Protégé 2000 is a tool which allows the user to construct domain ontologies. Here is an example written in the frame-based knowledge model of Protégé.

```
(defclass Author "Authors are the people or organizations
which provide articles"
    (is-a USER)
    (role abstract)
    (single-slot name_
        (type STRING)
        (cardinality 0 1)
        (create-accessor read-write)))
```

As logical languages, frame languages raise some problems. A technique, and a frame system is a technique, does not define a knowledge theory. In the same way, subsumption is more than an inheritance relationship and an essential property is more than an attribute.

8.3 Ontology-oriented languages.

Today, an ontology-oriented language must satisfy several requirements: it must provide a human readable form, it must be computational, it must rely on a well defined semantics in order to allow sound inferences and it must be "connected" to the web.

<u>*OWL*</u>

The Web Ontology Language (OWL) is a good example of such languages. It was designed by the W3C Web Ontology Working Group as a revision of the DAML (DARPA Agent Markup Language) and OIL (Ontology Inference Layer) web ontology languages. OWL wants to be a standard ontology language for the web and an interchange format for ontologies. It unifies three useful aspects for ontology: knowledge modeling primitives provided by frame systems; formal semantics and efficient reasoning support from Description Logics; and a syntax compatible with the Web standards. OWL is an extension of RDF (Ressource Description Framework) and RDFS (RDF Schema).

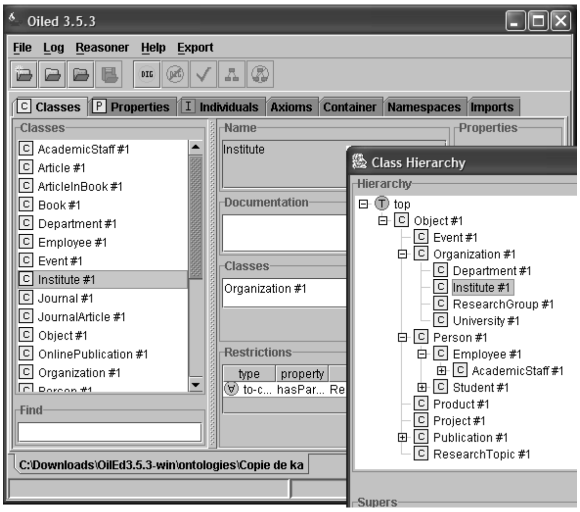

Fig.5 The OilEd ontology editor.

```
<owl:Class rdf:about="…/ka.daml#Institute">
  <rdfs:subClassOf>
    <owl:Class rdf:about="file:…/ka.daml#Organization"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty
        rdf:resource="file:…/ka.daml#hasParts"/>
      <owl:allValuesFrom>
        <owl:Class
          rdf:about="file:F…/ka.daml#ResearchGroup"/>
      </owl:allValuesFrom> …</owl:Class>
```

<u>*LOK*</u>

The LOK language (Language for Ontological Knowledge) is an ontology-oriented language based on the differentia principle (Roche 2001) taken into account epistemological and linguistic principles. The logical formalization is *a posteriori* and not *a priori*. A LOK concept is defined by specific differentiation, i.e. a concept is defined from a previously existing one adding a new differentia. In this approach differentiae and attributes are two different notions. The first one defines and organizes the concepts when the second one describes them. It implies that concept and set are also two different notions.

```
( defineConceptFrom 'Mechanical' 'metal preservation'
    ( leftConcept 'Stamping'
        ( specificDifference 'mechanical deformation)))
```

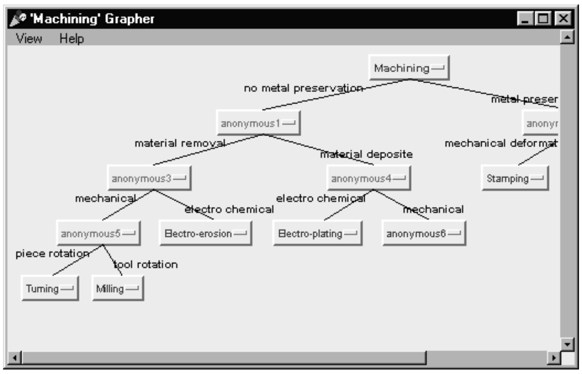

Fig.6 A LOK ontology as a Porphyry's tree.

9. WHAT TOOLS?

Ontologies becoming increasingly important, several tools for building ontologies were developed these last years. How can we choose the most relevant tool for a given application? interface, knowledge representation, interchange format, managing (maintenance, versioning, merging), inference, software architecture are some criteria of evaluation. KSL Ontology Editor, OilEd, OntoEdit, Protégé 2000, VOID (the KACTUS toolkit), WebODE, WebOnto are some examples. Let us see three of them in relation to the previous paragraph.

9.1 OilEd

OilEd (see Fig. 5) was developed in the context of the European IST OntoKnowledge project. It is a simple freeware ontology editor offering enough functionality to allow users to build ontologies using OWL (or underlying web languages like DAML+OIL or RDF).

9.2 Protégé 2000

Protégé 2000, an open source software, is a knowledge-base-editing environment developed by the Stanford Medical Informatics group. Its knowledge model is OKBC compatible (Open Knowledge-Base Connectivity protocol), a common query and construction interface for frame-based systems that facilitates interchange.

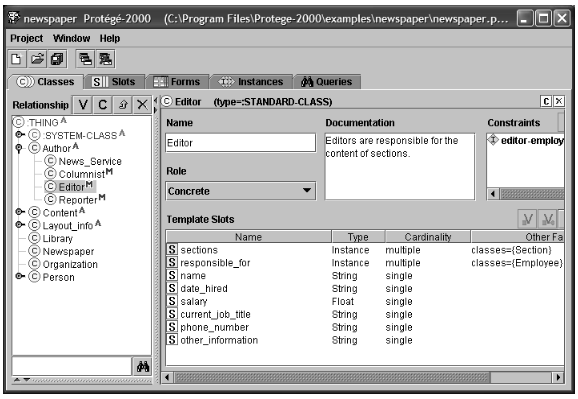

Fig.7 The Protégé 2000 software tool.

9.3 OK Station (Ontological Knowledge Station)

The *OK Station* is a graphical and interactive ontology-design environment. Initially developed at the University of Savoie, it is now a commercial product. Based on epistemological and linguistic principles, its own knowledge representation language (LOK) can be translated in KIF, conceptual graphs and web languages (OWL, OKML).

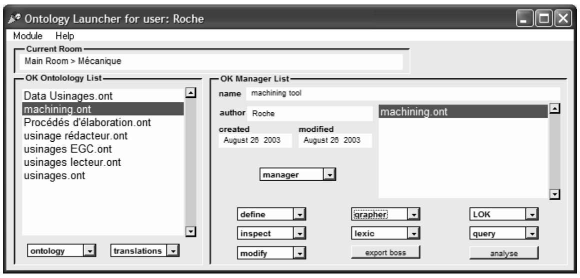

Fig.8 The Ontology launcher of the OK Station.

10. TO CONCLUDE

Ontologies have become increasingly important in a wide range of fields as well research as applications from artificial intelligence to e-commerce. In such a short survey it was neither possible to review all the subjects about ontology, nor to give complete references. Important subjects as merging and updating ontology or interchange format were not presented. Let us conclude with a last question and a final remark.

10.1 What about use, re-use and sharing?

Using different languages with different semantics, as well as not very clear epistemological principles, are barriers to the real use and re-use of ontologies. For example how can we trust in, and then use, an ontology which does not offer "any kind of guarantees" (Mikrokosmos ontology)? Using a same language, even a formal language like first-order logic, for the definition of terms is not a guarantee of such properties. For example, how can we combine the definitions coming from the two enterprise ontologies TOVE (Fox 92) and Enterprise Ontology (Ushold *et al.* 97)?

10.2 A necessary multi-disciplinary approach.

Each type of application has different requirements which imply different approaches of ontology. Web ontologies, engineering ontologies, linguistic ontologies are some examples which reinforce the idea that a general view is illusory.
But even for a same type of application, an ontology will be really used only if every body agrees on it and if some consistency is ensured. If the linguistic dimension is quoted in every approach, it is rarely taken into account except with a simple and naïve meaning theory (the meaning of a term is the concept it denotes) and it is in general limited to the building ontology phase. Some linguistic and terminological theories, e.g. Saussure's structuralism or Wüster's works, should be considered as possible theoretical foundations of ontology. The ontology problem requires a multidisciplinary approach based on sound epistemological, logical and linguistic principles.

REFERENCES

Electronic commerce. http://www.ontology.org

Fox M.S. (1992). The TOVE Project: Towards a Common Sense Model of the Enterprise. In: *Enterprise Integration*, C. Petrie (Ed.), Cambridge MA: MIT Press, 1992

Fernandez, M., Gomez-Perez, A., Juristo, N. (1997). METHONTOLOGY: From Ontological Art Towards Ontological Engineering. In: *Workshop on Ontological Engineering.* Spring Symposium Series. AAAI97 Stanford, USA.

FIPA. http://www.fipa.org/

Gruber T. (1995). Towards principles for the design of ontologies used for knowledge sharing. In: *International Journal of Human Computer Studies*, 43: 907-928.

Guarino N., Welty C. (2000). A Formal Ontology of Properties. In: *Proceedings of 12th Int. Conf. on Knowledge Engineering and Knowledge Management*. Lecture Notes on Computer Science, Springer Verlag, 2000.

KACTUS (1994). Esprit Project 8145. Framework and formalism for expressing ontology. http://www.swi.psy.uva.nl/projects/NewKACTUS/home.html

Chaudhri V., Farquhar A., Fikes R., Karp P Rice J., (1998). OKBC: A Programmatic Foundation for Knowledge Base Interoperability. In: *Proceedings of AAAI-98*, July 26-30 1998, Madison, WI.

OilEd. http://www.ontoknowledge.org/oil/

Ontolingua. http://www.ksl.stanford.edu/software/ontolingua

OntoWeb. Technical roadmap. IST Project IST-2000-29243. http://ontoweb.aifb.uni-karlsruhe.de/

OWL. http://www.w3.org/TR/owl-features/

Protégé 2000. http://protege.stanford.edu/

RDF. http://www.w3.org/RDF/

Roche C. (2001). The 'Specific-Difference' Principle: a Methodology for Building Consensual and Coherent Ontologies. In: *International Conference on Artificial Intelligence*. Las Vegas, USA, June 25-28 2001.

Semantic Web. http://www.w3.org/2001/sw/

Ushold M., Gruninger M. (1996). Ontologies : Principles, Methods and Applications. In: *Knowledge Engineering Review,* Vol. 11, n° 2, June 1996

Uschold M. and King M. (1995). Towards a Methodology for Building Ontologies. Internal report AIAI-TR-183, July 1995.

Uschold M., King M., Moralee S., Zorgios Y. (1997) The Enterprise Ontology. AIAI, The University of Edinburgh, 1997