The Role of Frame-Based Representation on the Semantic Web

Ora Lassila[1] & Deborah McGuinness[2]

[1]Software Technology Laboratory, Nokia Research Center

[2]Knowledge Systems Laboratory, Stanford University

Abstract

A new architecture for the World Wide Web is emerging, known as the *Semantic Web*. In broad terms, it encompasses efforts to populate the Web with content which has formal semantics; this will enable automated agents to reason about Web content, and produce an intelligent response to unforeseen situations.

We believe that to build the Semantic Web, the sharing of ontological information is required. This allows agents to reach *partial shared understanding* and thus interoperate. We are proposing frame-based representation as a suitable paradigm for building ontologies as well as the World Wide Web Consortium's RDF-formalism (and its extensions, such as the DARPA Agent Markup Language) as a manifestation of frame-based representation for the Web.

The paper will discuss required and desirable features of ontological languages, giving examples of the possible usage of frame-based representation and ontologies on the Semantic Web.

Introduction: Frame-based Representation Systems

The term "Semantic Web" encompasses efforts to build a new WWW architecture to support content with formal semantics; that is, content suitable for automated systems to consume, as opposed to content intended for human consumption. The Semantic Web will allow us to use more automated functions (reasoning, information and service discovery, autonomous agents, etc.) on the WWW, easing the workload of humans (we probably need not repeat the well-known example here of using a search engine and getting back thousands of "hits" which then need to pruned by hand). The Semantic Web will also pave the way for true "device independence" and customization of information content for individual users, since the information on the web would now be contained in a "raw form" and any (context-dependent) presentation could be rendered on demand.

The success of the deployment of the Semantic Web will largely depend on whether useful ontologies will emerge, allowing shared agreements about vocabularies for knowledge representation. Considering the WWW community at large (as largely disjoint from the KR community), the acceptance for any particular KR technology or paradigm is of critical importance. It is the authors' belief that the simplicity and ease of understanding of frame-based representation would make it the paradigm of choice for knowledge representation when building the Semantic Web – given that we are really attempting the deployment of "real" KR on a WWW-wide scale. If we offer the WWW community a frame-based representation system which allows people to start modeling and (most importantly) *sharing* their models; some of these models will grow into useful ontologies (partially through a process of "natural selection") and we have taken a step closer to our goal.

The concept of a *frame* was proposed in the 1970's [Minsky 1975], and frame systems subsequently gained ground as basic tools for representing knowledge [Fikes & Kehler 1985, Karp 1992, Chaudhri et. al. 1998]. The fundamental idea of a frame system is rather simple: A frame represents an object or a concept. Attached to the frame is a collection of attributes (slots), potentially filled initially with values. When a frame

is being used the values of slots can be altered to make the frame correspond to the particular situation at hand. According to an interpretation by Minsky, the slots of a frame might represent questions most likely to arise in a hypothetical situation represented by the frame.

Frames are closely related to an earlier structure-based KR technique, called semantic networks [Woods 1975] which, in turn, are based on the idea of human associative memory [Quillian 1967], but may simply be thought of as data structures of nodes - "concepts" - and links - "associations" - between them. If one thinks of frames as concepts, and when other frames are used to fill slots we have an analogous framework.   The notion of semantic networks also led to the predecessor of *description logics* (http://dl.kr.org) as we know them today with the introduction of KL-ONE in Brachman's thesis [Brachman 1977]. This work began with an emphasis on making term definitions in semantic networks more precise. Description Logics provide representation and reasoning languages with precise semantics. They also limit language expressiveness so that they can guarantee tractable inference.

Soon after its inception, the notion of a frame system was criticized as not introducing anything new to the field of KR; for example, Pat Hayes has said, "most of 'frames' is just a new syntax for first-order logic" [Hayes 1979]. Although this statement is easy to accept, it doesn't diminish the value of frame systems as easy-to-understand tools for *simple* KR (starting from what might be called "structural modeling"). An example of a frame-based system that argues both these points is Ontolingua [Farquhar et al. 1997]. It provides a frame-based syntax but then translates all information into KIF, which is just a first order logic encoding of the information (see http://logic.stanford.edu/kif/kif.html).

There also exists a connection between frame systems and object-oriented programming (OOP) [Hynynen & Lassila 1989, Lassila 1990], particularly if we think of the "structural modeling" aspect mentioned above. Basic vocabulary is different, but what the terms denote are approximately the same, see table below:

OOP Systems	Frame Systems	Description Logics
instance	frame, instance, individual	instance, individual
attribute, instance variable	slot	role
value	filler	filler
class, type	frame, schema	class, concept

From the adoption viewpoint, it can be observed that many people understand OOP even if they have never heard of frame systems. We can think of frame systems very pragmatically through a "heuristic" interpretation (they are vehicles for storing knowledge and performing inferences), and depart from Minsky's "metaphysical" interpretation; in comparison to OOP, frame systems add the notion of *reasoning*.

As indicated earlier, the success of the semantic web depends on the emergence of shared ontologies. Why do we consider this important? The fact that we share vocabularies and models allows us to interoperate; given a "base ontology" shared by two agents, each agent can still extend this ontology yet "partial understanding" can be achieved (much like in OOP systems, a base class defines "common" functionality). We believe that this partial understanding will actually advance the web from its current state to what might be called the "Semantic Web".

RDF and What Is Missing From It

The Resource Description Framework (RDF) is a standard for Web metadata developed by the World Wide Web Consortium (W3C) [Lassila & Swick 1999, Lassila 1998]. Expanding from the traditional notion of document metadata (such as something like library catalog information), RDF is suitable for describing any Web resources, and as such provides interoperability between applications that exchange "semantic" information on the Web.

RDF model instances are directed labeled graphs where nodes and arcs are named using URIs Universal Resource Identifiers) [Berners-Lee et al 1998]. Consequently, RDF can describe not just things on the Web (such as pages, parts of pages, collections of pages) but also things not on the Web as long as they can be named using some URI scheme. The RDF basic description model uses object/attribute/value -triples: instead of viewing instances of the model as directed, labeled graphs (in this respect they resemble semantic networks), one can take a more object-oriented view and think of RDF as a frame-based representation system. In comparison, to extend the "translation" from the previous section, at the lowest level, object-oriented programming systems, frame systems, and RDF have much in common but again the terms we use are all different:

OOP Systems	Frame Systems	Description Logics	RDF
instance	frame, instance, individual	instance, individual	resource
attribute, instance variable	slot	role	property
value	filler	filler	property value
class, type	frame, schema	class, concept	class

Descriptions in RDF can be limited to one resource only (e.g., a "library catalog card" which names the author, publisher, etc. of a document); in this case the values of resource attributes (or "properties" as they are called in RDF) are typically strings. Descriptions can also span multiple resources: values of properties can be other resources, and it is therefore possible to describe arbitrary relationships between multiple composite resources with structure. Since properties themselves are also named by URIs and can be described: what are the permitted values of a particular property, which types of resources can it describe, and what is its relationship to other properties. Meaning in RDF comes from specific terms and concepts being defined and then named by URIs. Because URIs can be made unique, two systems can define some concept (say, "person") and can each use a different URI to name it to avoid "clashes"; on the other hand, two systems agreeing on a common concept will use the same URI and effectively "share" semantics.

On top of the basic RDF model is layered an extensible, object-oriented type system (known as RDF Schema) [Brickley & Guha 2000]. The meta-constructs for the type system are terms and concepts named by URIs, so effectively RDF itself is used in representing and defining classes and properties. Class definitions can be derived from multiple superclasses. Property definitions can specify domain and range constraints. One can also think of RDF Schema as a set of ontological modeling primitives on top of RDF. As such, the RDF Schema still needs work. Some of this has recently emerged in the form of the OIL system [Bechhofer et. al. 2000] and work continues within DARPA's DAML program [Hendler & McGuinness 2000]. Additionally there is a new axiomatic semantics for RDF, RDFS, and the language resulting from the DAML program [Fikes & McGuinness 2001].

Ontologies

Ontologies have been around for many years.   Merriam Webster, for example, dates ontology circa 1721 and provides two definitions (1) a branch of metaphysics concerned with the nature and relations of being and (2) a particular theory about the nature of being or the kinds of existents. These definitions provide an abstract philosophical notion of ontology. Ontologies slowly moved into a more mathematical and precise domain and the notion of a formal ontology existed since at least 1900. Smith {Smith 1998] points out that the philosopher Husserl distinguished them from formal logic [Husserl 1900]

Ontologies have been gaining interest and acceptance in broader audiences. Guarino [Guarino 1998] provides a nice collection of fields that embrace ontologies including knowledge engineering, knowledge

representation, qualitative modeling, language engineering, database design, information retrieval and extraction, and knowledge management and organization. That collection put together in early 1998 did not even include as much emphasis from the web as is seen today. We would also include areas of library science [Dublin Core 1999], ontology-enhanced search (e.g., eCyc (http://www.e-Cyc.com/) and FindUR [McGuinness 1998]), and possibly the largest one, e-commerce (e.g., Amazon.com).   Today's use of ontology on the web has a different slant from the previous philosophical notions however. One widely cited definition of an ontology is Gruber's [Gruber 1993 "A specification of a conceptualization". We will use this notion and expand upon it in our use of the term.

People (and computational agents) typically have some notion or conceptualization of the meaning of terms. Software programs sometimes provide a specification of the inputs and outputs of a program, which could be used as a specification of the program. Similarly ontologies can be used to provide a concrete specification of term names and term meanings. Within this line of thought though where an ontology is a specification of the conceptualization of a term, there leaves much room for variation. Web ontologies may be viewed as a spectrum of detail in their specification. One might visualize a simple (linear) spectrum of definitions in Figure 1 [1]below.
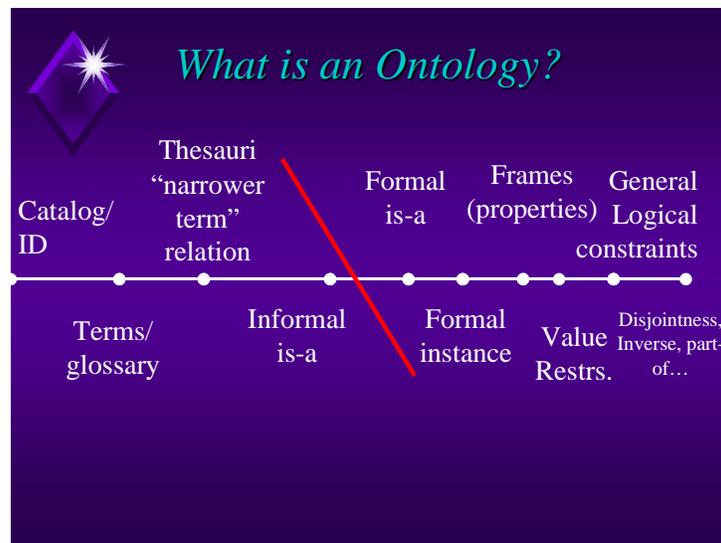

Figure 1: An Ontology Spectrum

One of the simplest notions of a possible ontology may be a controlled vocabulary – i.e., a finite list of terms. Catalogs are an example of this category. Catalogs can provide an unambiguous interpretation of terms – for example, every use of a term, say car – will denote exactly the same identifier – say 25.

Another potential ontology specification is a glossary (a list of terms and meanings). The meanings are specified as natural language statements. This provides more information since humans can read the natural language statements. Typically, interpretations are not unambiguous and thus these specifications are not adequate for computer agents.

Thesauri provide some additional semantics in their relations between terms. They provide information such as synonym relationships. In many cases their relationships may be interpreted unambiguously by

---

[1] This spectrum arose out of a conversation in preparation for an ontology panel at AAAI '99. The panelists (Lehman, McGuinness, Ushold, and Welty), chosen because of their years of experience in ontologies found that they encountered many forms of specifications some people might call ontologies. McGuinness refined the picture.

agents. Typically thesauri do not provide an explicit hierarchy (although with narrower and broader term specifications, one could deduce a hierarchy).

Early web specifications of term hierarchies, such as Yahoo's, provide a general notion of generalization and specialization. Yahoo, for example, provides a small number of top-level categories such as apparel and then dresses as a kind of (women's) apparel. A small number of people consider the previous categories (of catalogues, glossaries, and thesauri) to be ontologies but many prefer to have an explicit hierarchy included before something is considered an ontology. Yahoo does provide an explicit hierarchy. Its hierarchy is not a strict subclass or "isa" hierarchy however. This point was distinguished on the spectrum slide since it seems to capture many of the "naturally occurring ontologies" on the web. In these hierarchies it is typically the case that an instance of a more specific class is also an instance of the more general class but that is not enforced 100% of the time. For example, the general category apparel includes a subcategory women (which should more accurately be titled women's apparel) which then includes subcategories accessories and dresses. While it is the case that every instance of a dress is an instance of apparel (and probably an instance of women's dress), it is not the case that a dress is a woman and it is also not the case that a fragrance (an instance of a women's accessory) is an instance of apparel. This mixing of categories such as accessories in web classification schemes is not unique to Yahoo – it appears in many web classification schemes[2].

The next point on the figure includes strict subclass hierarchies. In these systems if A is a superclass of B, then if an object is an instance of B necessarily implies that the object is an instance of A. Strict subclass hierarchies are necessary for exploitation of inheritance. The next point on the ontology spectrum includes formal instance relationships. Some classification schemes only include class names while others include ground individual content. This point includes instances as well.

The next point includes frames. Here classes include property information. For example, the apparel class may include properties of price and "is-made-from". My specific dress may have a price of $50 and may be made from cotton. Properties become more interesting when they are specified at a general class level and then inherited by subclasses and instances. In a consumer hierarchy, a general category like consumer product might have a price property associated with it. Possibly apparel would be the high level category to have is-made-from associated with it. All subclasses of these categories would inherit these properties.

A more expressive point in the ontology spectrum includes value restrictions. Here we may place restrictions on what can fill a property. For example, price might be restricted to be a number (or a number in a certain range) and is-made-from may be restricted to be a kind of material. Here we can see a possible problem with a classification scheme that does not support strict "isa" or subclass relationships. For example, if fragrance were a subclass of apparel, it would inherit the property is-made-from and the value restriction of material that was stated.

As ontologies need to express more information, their expressive requirements grow. For example, we may want to fill in the value of one property based on a mathematical equation using values from other properties. Some languages allow ontologists to state arbitrary logical statements. Very expressive ontology languages such as that seen in Ontolingua or CycL allow ontologists to specify first order logic constraints between terms.

---

[2] Some prominent hierarchies such as Yahoo have renamed their classes to broad disjunctive categories such as "Apparel, Accessories, and Shoes" presumably in order to provide for more strict subclass relationships. Disjunctive categories make inheritance more problematic however with class-specific properties.

In this paper, we will require the following properties to hold in order to consider something an ontology:

1. Finite controlled (extensible) vocabulary
2. Unambiguous interpretation of classes and term relationships
3. Strict hierarchical subclass relationships between classes

We consider the following properties typical but not mandatory:

4. Property specification on a per-class basis
5. Individual inclusion in the ontology
6. Value restriction specification on a per-class basis

Finally, the following properties may be desirable but not mandatory nor typical:

7. Specification of disjoint classes
8. Specification of arbitrary logical relationships between terms

Uses of Frame-Based Representation on the Semantic Web

What is the purpose of the semantic web? The authors would like to believe that, given the current state and design of the web (i.e., intended for human audiences which makes automating tasks difficult), the semantic web will allow us to use more automated functions (reasoning, information discovery, autonomous agents, etc.) on the web, easing the workload of humans (we probably need not repeat the well-known example here of using a search engine and getting back thousands of "hits" which then need to pruned by hand). The semantic web will also pave the way for true "device independence" and customization of information content for individual users, since the information on the web would now be contained in a "raw form" and any (context-dependent) presentation could be rendered on demand.

As indicated earlier, the success of the semantic web depends on the emergence of shared ontologies. Why do we consider this important? The fact that we share vocabularies and models allows us to interoperate; given a "base ontology" shared by two agents, each agent can still extend this ontology yet "partial understanding" can be achieved (much like in OOP systems, a base class defines "common" functionality). We believe that this partial understanding will actually advance the web from its current state to what might be called the "Semantic Web".

One such potential base ontology candidate for products and services is provided by the joint effort between the United Nations Development Program and Dun & Bradstreet. The two organizations merged their separate commodity classification codes into a single open distributed system called the UNSPSC - Universal Standard Products and Services

Classification (UNSPSC) Code organization (www.unspsc.org). It provides a classification scheme (with associated numbers) that may be an appropriate top five layer for applications in services in products. For example, Category 53 – Apparel and Luggage has a subclass family 5310 of Clothing which in turn contains a subclass class [531020] Dresses and skirts and saris and kimonos, which in turn has a subclass commodity of [53102002] Women's dresses or skirts or saris or kimonos. Many B2B applications today are attempting to extend the UNSPSC for their applications so that they can conveniently interconnect with other web sites also using the coding scheme. While this ontology is not expected to be complete enough for any single application, it is expected to be a reasonably top level or umbrella structure for most services and products applications.

Given the automation aspect, the semantic web represents a radical departure from the current browsing paradigm of the web. Indeed, we can foresee applications and usage situations where browsing is not a

viable alternative. On the other hand, one might argue that browsing still takes place, but by automated agents and not by human users. If we can represent information about web-based services in a formal manner, the discovery of additional functionality needed by agents becomes easier. Generally, the process of matching one agent's description of "missing" functionality with descriptions of services offered by other agents we call "service discovery", and resembles the act of advertising and querying. It is becoming an increasingly important aspect of distributed information systems.

A number of mechanisms for low-level service discovery are emerging (examples include Sun's Jini and Microsoft's Universal Plug and Play [Richard 2000]); these mechanisms attack the problem at a syntactic level, and rely heavily on standardization of a predetermined set of functionality descriptions. Standardization, unfortunately, can only take us half way towards our goal of automated agents on the web, as our ability to anticipate all possible future needs is limited. The semantic web offers a possibility to elevate the mechanisms of service discovery to a "semantic" level. Here a more sophisticated description of functionality is possible, and the shared understanding between the consumer and the producer can be reached via the exchange of ontologies, which provide the necessary vocabulary for discussion.

More importantly, it will also be possible to reach a "partial match" between a request and an advertisement [e.g., Sycara et al 1999], yet be able to take advantage of the service thus discovered. An agent could thus enhance its functionality in the following manner:

1. Exchange of ontologies, partial understanding

2. Partial match of needs with services offered

3. Composition of exact required functionality from partially matched discovered services (using, for example, automated planning or configuration techniques).

This approach takes information system interoperability beyond what mere standardization or simple interface sharing enables, since it would be based on "deeper" descriptions of service functionality and can be performed ad hoc and on demand.

Conclusions

The Semantic Web can be viewed as the next generation of the World Wide Web architecture, an essential part of which is the content producers' ability to the share the ontological information that defines the formal semantics of the Web content. Through shared ontological commitments reasoning agents are able to reach *partial shared understanding* and thus interoperate.

In this article we have proposed frame-based representation as a suitable paradigm for building ontologies, and the World Wide Web Consortium's RDF-formalism (and its extensions, such as the DARPA Agent Markup Language) as a manifestation of frame-based representation for the Web. The emergence of the Semantic Web is largely predicated on the broader "Web community" *accepting and deploying* distributed knowledge representation technologies, and in this context the ease-of-understanding of the KR paradigm is of paramount importance. Frame systems, with their close kinship to more widely understood object-oriented technologies, may prove to be the "right" paradigm.

References

[Bechhofer et al 2000]   Sean Bechhofer, Jeen Broekstra, Stefan Decker, Michael Erdmann, Dieter Fensel, Carole Goble, Frank van Harmelen, Ian Horrocks, Michel Klein, Deborah McGuinness, Enrico Motta, Peter Patel-Schneider, Steffen Staab, and Rudi Studer, "An informal description of Standard Oil and Instance OIL", available on-line as http://www.ontoknowledge.org/oil/downl/oil-whitepaper.pdf

[Berners-Lee et al 1998]   Tim Berners-Lee, Roy Fielding, and Larry Masinter, "Uniform Resource Identifiers (URI): Generic Syntax", Internet Draft Standard RFC 2396, August 1998; available on-line as http://www.isi.edu/in-notes/rfc2396.txt.

[Brachman 1977]   Ronald J. Brachman, *A Structural Paradigm for Representing Knowledge,* Ph.D. thesis, Harvard University, USA, 1977.

[Brachman, Schmolze 1985]   Ronald J. Brachman, James G. Schmolze, `An Overview of the KL-ONE Knowledge Representation System', *Cognitive Science,* Vol 9(2), pp 171-216, 1985.

[Brickley & Guha 2000]   Dan Brickley & R.V.Guha, "Resource Description Framework (RDF) Schema Specification 1.0", W3C Candidate Recommendation 27 March 2000, World Wide Web Consortium, Cambridge (MA); available on-line as http://www.w3.org/TR/rdf-schema/.

[Chaudhri et. al. 1998]   Vinay Chaudhri, Adam Farquhar, Richard Fikes, Peter Karp, and James Rice; "*OKBC: A Programmatic Foundation for Knowledge Base Interoperability"*, AAAI 1998.

[Dublin Core 1999]   "Dublin Core Metadata Element Set, Version 1.1: Reference Description", Dublin Core Metadata Initiative, 1999; available on-line as http://purl.org/dc/documents/rec-dces-19990702.htm

[Farquhar et al 1997]   Adam Farquhar, Richard Fikes, and James Rice; "*The Ontolingua Server: a Tool for Collaborative Ontology Construction*", Intl. Journal of Human-Computer Studies **46**, 1997

[Fikes & Kehler 1985]   Richard Fikes & Tom Kehler, "The Role of Frame-Based Representation in Reasoning", CACM 28(9): 904-920 (1985).

[Fikes & McGuinness 2001]   Richard Fikes & Deborah L. McGuinness, "An Axiomatic Semantics for RDF, RDF Schema, and DAML+OIL", KSL Technical Report KSL-01-01, Stanford University, 2001; available on-line as http://www.ksl.stanford.edu/people/dlm/daml-semantics/abstract-axiomatic-semantics.html.

[Guarino 1998]   Nicola Guarino, "Formal Ontology and Information Systems''. In the Proceedings of Formal Ontology in Information Systems, June 1998. Also in Frontiers in Artificial Intelligence and Applications, IOS-Press, Washington, DC, 1998.

[Gruber 1993]   Tom R. Gruber, "A translation approach to portable ontologies". Knowledge Acquisition, 5(2):199-220, 1993.

[Hayes 1979]   Patrick J. Hayes, "The Logic of Frames", in D. Metzing (ed.), *Frame Conceptions and Text Understanding*, 46-61, Walter de Gruyter and Co., Berlin, 1979.

[Hendler & McGuinness 2000]   James Hendler and Deborah McGuinness. ``The DARPA Agent Markup Language''. In IEEE Intelligent Systems Trends and Controversies, November/December 2000. Available from http://www.ksl.stanford.edu/people/dlm/papers/ieee-daml01-abstract.html .

[Husserl 1900]   Edmund Husserl, Logische Untersuchungen, First edition Halle: Niemeyer, 1900/01.

[Hynynen & Lassila 1989]   Juha Hynynen & Ora Lassila, "On the Use of Object-Oriented Paradigm in a Distributed Problem Solver", AI Communications 2(3/4): 142-151 (1989).

[Karp 1992]   Peter D. Karp, "The design space of frame knowledge representation systems", Technical Report 520, SRI International AI Center; available on line as ftp://www.ai.sri.com/pub/papers/karp-freview.ps.Z

[Lassila 1990]   Ora Lassila, "Frames or Objects, or Both?", Workshop Notes from the 8th National Conference on AI (AAAI-90): Object-Oriented Programming in AI, Boston (MA), 1990; also appeared as Report TKO-B67, Department of Computer Science, Helsinki University of Technology, Espoo (Finland), 1991.

[Lassila 1998]	Ora Lassila, "Web Metadata: A Matter of Semantics", IEEE Internet Computing 2(4): 30-37 (1998).
[Lassila & Swick 1999]	Ora Lassila & Ralph Swick, "Resource Description Framework (RDF) Model and Syntax Specification", W3C Recommendation 22 February 1999, World Wide Web Consortium, Cambridge (MA); available on-line as http://www.w3.org/TR/REC-rdf-syntax/.
[McGuinness 1998]	Deborah L. McGuinness ``Ontological Issues for Knowledge-Enhanced Search''. In the Proceedings of Formal Ontology in Information Systems, June 1998. Also in Frontiers in Artificial Intelligence and Applications, IOS-Press, Washington, DC, 1998.
[Minsky 1975]	Marvin Minsky, "A Framework for Representing Knowledge", in Patrick Henry Winston (ed.), *The Psychology of Computer Vision*, McGraw-Hill, New York, 1975.
[Quillian 1967]	M. Ross Quillian, "Word Concepts: A Theory and Simulation of some Basic Semantic Capabilities", Behavioral Science 12: 410-430 (1967).
[Richard 2000]	Golden G. Richard III, "Service Advertisement and Discovery: Enabling Universal Device Cooperation", IEEE Internet Computing 4(5): (2000).
[Smith 1998]	Barry Smith, "Basic Concepts of Formal Ontologies", in N. Guarino (Ed.) Formal Ontology in Information Systems, IOS Press, 1998.
[Sycara et al 1999]	Katia Sycara, Matthias Klusch, Seth Widoff & Jianguo Lu: "Dynamic Service Matchmaking Among Agents in Open Information Environments", in: A. Ouksel, A. Sheth (eds.): ACM SIGMOD Record (Special Issue on Semantic Interoperability in Global Information Systems), 28(1): 47-53 (March 1999)
[Woods 1975]	William A. Woods, "What's in a Link: Foundations for Semantic Networks", in D.G.Bobrow & A.M.Collins (eds.), *Representation and Understanding: Studies in Cognitive Science*, 35-82, Academic Press, New York, 1975.