The Ontology problem in eCommerce applications

Rasheed M. Al-Zahrani
Information Systems Dept., KSU
PO Box 51178, Riyadh, 11543
rasheed@ccis.ksu.edu.sa

Abstract

Originating in AI semantic networks, ontologies are becoming an essential component of many modern systems. An ontology is a set of specifications, relationships and constraints that describe a certain domain. These specifications capture the concepts pertaining to the domain. Research in this domain is now witnessing intensive efforts due to the growth and success of distributed computing systems in real world applications such as eCommerce, eHealth, eLearning and other eServices. Though at the core of modern distributed technologies, such as multi-agent systems, the ontology issue has sometimes been considered secondary and related issues are underestimated. In this paper, we attempt to address the ontology issue in modern distributed services, and the various problems to be investigated, with special emphasis on eCommerce systems. Our paper illustrates how the semantic-web initiative integrates with ontology. It critically appraises existing solutions, and offers ideas for tackling major ontological issues in eServices.

1. Introduction

Distributed systems is the future computing model. This fact is proved by the success of the client-server model and the recent extensions to that model. Maturity of cheap PC, networking and communication technologies contributed to the wide spread adoption of this model. The advent of Internet and the services developed for it gave solid successful examples of distributed systems, even when the underlying network is a hybrid network like the internet.

In the next two sub-sections, we address two major technologies that will have significant impact on the future of distributed systems.

1.1 Web revolution and eServices

The world wide web (W3) was one of the most successful innovations of information technology. It allowed, for the first time, merger of multimedia objects into a single logical medium that is easily browsed and flexibly extended. Web was a result of many constituent technologies that together allowed interoperability of distributed services seamlessly. Hardware, operating system, communication and application differences were notably avoided and humans are now enjoying a super revolution that yielded more connectivity, cooperation, and exchange of services, information and goods.

WWW users face two problems. The first is the exponential growth of the web and the resulting information explosion. People are nowadays finding it harder and harder to get to the right information despite of the availability of powerful search engines. They need more personalized tools to assist them in surfing the web.

The second problem is that of interoperability in an inconsistent / environment. Modern organizations are currently exploiting the web to achieve more benefits through mergers, partnership and integration with other organization. This requires interaction between the software systems of those organizations. This is a complicated issue, given the fact that, when they are converted into web-enabled applications, such systems have many types of differences that are beyond those solved by the web. Those systems may not be able to cooperate. Visitors of websites having such softwares may encounter problems due to those underlying differences.

On the other hand, the web is rich with sites that offer similar services to users. However, those services, e.g. bookselling, have their own concepts, rules and pricing schemes. It is important for users to be able to compare those services, a non-trivial task in the presence of those many peculiarities. The difficulty here stems from the semantic variations within web services.

For present day web community, utilizing even static web sites that are not connected to any software systems has become uneasy. Finding relevant information in huge sites requires tools for fast discovery according to user interests, enquiries, history, and so on. Computers must do more for people to benefit from web contents.

Semantic web was suggested as a solution for the two problems. Today, WWW is a medium of documents for people rather than of information that can be manipulated automatically. By augmenting web pages with data targeted at computers and by adding documents solely for computers, we transform the web into the Semantic web [LEE2001].

1.2   Agents in eCommerce

Rapidly evolving network and computer technology, coupled with the exponential growth of the services and information available on the Internet, will soon bring us to the point where hundreds of millions of people will have fast, pervasive access to a phenomenal amount of information through desktop machines at work, school and home, through televisions, phones, pagers and car dashboards, from anywhere and everywhere [KOT1999]. In this ocean, software agents will be essential.

Agents are active, autonomous and personalized softwares to which tasks can be delegated. Agents are meant to support humans in conducting repetitive, time-consuming and demanding tasks, such as email filtering, information searching and travel planning. They are usually semi-automatic softwares that are continuously running. Research into this area started since 1994, with Maes' classis paper "Agents that Reduce Work and Information Overload" [MAE1994]. Since then, many journal issues, conferences and research groups were established. Covering this topic in more details is beyond the scope of this paper.

In eCommerce, agents will be (and are actually being) used to filter, discover, negotiate, personalize, take decisions, mediate, and keep their owners up-to-date with the task in hand. Customers need assistance because eCommerce is very competitive, and virtual shopping is an extremely time consuming experience for shoppers due to the international gathering of markets in the Internet. According to many customer buying behavior models, agents are most suited for product brokering, merchant brokering, and negotiating stages of buying [GUT1999]. Those are personalized evaluations that require complex coordination in a timely manner.

Our paper introduces the need for ontology, its relation to agents and eCommerce, and summarizes the major efforts and standards related to the ontology layer in cooperating agents. It also attempts to address the ontology problem and its demanding applications in modern eServices. The paper demonstrates the role of ontologies in multi-agent systems. It also describes the interaction needed between the context layer and the content layer of multi-agents, putting more emphasis on standardizing various levels surrounding an ontology.

This paper consists of six sections. Section 2 discusses the semantics issue and introduces semantic web and its complementary technologies that were developed to support semantic capturing and management in modern distributed information systems. In section 3, we offer an insight into the ontology science and its relevance to solving traditional and modern interoperability problems. More discussion on the interaction between agents and ontologies is given in section 4. Then we briefly highlight the major issues encountered by ontology developers and propose courses of action in section 5. Section 6 concludes the paper.

2.   Semantics

Semantics of a certain domain are the background knowledge that experts of the

domain implicitly use to interpret information and take decisions. Those semantics are sometimes hard to explain and document. It is the inability of computers to understand and utilize those semantics that restricted application of many IT solutions to a variety of traditional and emerging needs in our life. Traditionally, this issue had always complicated the integration of legacy systems. In recent years, information floods caused by the success of WWW is calling for innovative tools that can help people manage their share of this flood. The semantics issue will be the major concern of any such tools.

Current web is designed for human consumption. This is to say that computers of today can manage websites, can publish them can understand their structure, but cannot interpret their content meaning.

Semantic web is not a new web, but rather an extension to the current one in which information is given well defined meaning, better enabling computers and people to work in cooperation [LEE2001]. Computers find the meaning of semantic data by following links to definitions of terms and rules for reasoning on those terms, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users [LEE2001]. Semantic web brings structure to the meaningful content of web pages, which allows computers to reason on those pages. New semantic web pages are added to the web by ordinary people, through special software tools that allow them to define new terms and rules.

Semantic web pages need to support structure, term definitions and reasoning rules in some knowledge representation form. Semantic web builds on two existing technologies: XML and RDF. XML[BOS2002, W3C1998] allows authors to add their own tags and define structure within documents, but says nothing about the meaning of those structures. RDF [LAS1999] specifies meaning for those structures. RDF may use XML tags to annotate information meaning. An RDF assertion consists of a triplet (subject, property, object), which specifies that a resource have as a property another resource.

For example, a school is managed by a head teacher. Both subject and object resources are identified by URIs (Universal Resource Identifiers). The property "is managed by" reflects a relationship between the two URIs. This relationship might be interpreted differently in different web pages. Hence, it must be defined clearly. This is achieved by providing a third URI which points to a meaning for the term used for the relationship. This link points to an Ontology.

RDF is a foundation for processing metadata; it enables interoperability between applications that exchange machine-understandable information on the web [LAS1999]. RDF does not offer reasoning mechanisms. Those can be added on top of it.

Our previous work in merging integrity specification in legacy systems [ALZ98, ALZ96, ALZ95] has given us the chance to realize the complexity of the ontology problem. Although merging two sets of integrity rules might look like a simple task, taking domain semantics into consideration creates a number of challenging alternatives.

In eCommerce, business semantics are the terms each party must understand and use during communications and interactions. For instance, in flight reservation eCommerce, concepts like fare, tax and leg must have specific definitions agreed by different parties involved. Those terms may also have constraints, default values, and relations to other terms. Web agents involved in handling tourist travel planning must be able to recognize and handle those issues.

Although the semantics problem occurs in both traditional and web applications, it has attracted more attention after the emerging of the semantic web technology [LEE2001, LEE1998]. Ontology offers the mechanism for capturing and managing domain semantics.

3.   Whither Ontology

Ontology is a set of terms, relationships between those terms, and inference rules to reason on such terms and relationships [LEE2001]. It defines the common vocabulary needed by a certain community to share information relevant to that community. It is a set of term definitions, their properties, constraints, relationships to other terms and reasoning rules. [Noy2001] defines ontology as

formal explicit descriptions of concepts in a domain of discourse. [DIN2002] describes ontology as a reference model of applications domains with the purpose of improving information consistency and reusability, systems interoperability and knowledge sharing.

Developing ontologies allows those specialized communities to share a common understanding of the structure of information, to reuse knowledge, to make explicit domain assumptions, to separate domain knowledge from the operational knowledge, and to analyze domain knowledge [Noy2001]. Ontology is gaining popularity and is touted as an emerging technology that has a huge potential to improve information organization, management and understanding [DIN2002]. Ontology plays the strategic role for agent communication. Ontology mapping is the capable way to break the bottleneck of B2B marketplace [FEN2002]. An ontology may improve the accuracy of searching and enable the development of powerful applications that tackle complicated questions whose answers do not reside on a single web page [LEE2001].

An ontology works like a reference engine for queries related to concepts. It receives terms and responds with their related terms, constraints, rules, etc.
Research in this domain is not new as it builds on a long history of research in knowledge engineering. What is new is the linkage of this technology to semantic web, and the attempts to use it for interoperability of information systems, in and outside the web boundaries.

Many research groups around the globe are tackling the various issues related to ontology. [DIN2002] summarizes some of those efforts. University of Texas has a dedicated site for this domain (http://www.cs.utexas.edu/users/mfkb/related.html). A survey of those efforts is outside the scope of this paper. [PER2002] offers a comprehensive comparative study of the ontology tools developed by research and commercial organizations. Tens of such tools were built to create, evolve, map, evaluate and query ontologies.

4. Multi-Agent systems interaction with Ontology

In the WWW realm ontology construction has accelerated, and large ontologies are currently available in different forms, including taxonomies (e.g. on Yahoo! And Amazon.com). The wide acceptance of eServices has created extensive demand for Agents that may co-operate through standard ontologies. Consequently, ontologies have received good attention from the intelligent agents research community [NWA1999, PER2002].

Software agents developed on semantic web will be able to exchange knowledge and cooperate in fulfilling user requirements. This inter-communication will be facilitated by ontologies, such as Cyc [LEN1995]. Cooperating agents must be able to determine the proper ontologies through the context of the task in hand. Agent communication languages, e.g. ACL , KIF, DAML and KQML [NOY2001], do not have means for defining task context of a certain query. The presence of ontologies helps in addressing this problem but does not solve it fully. There is a need for inter-ontology communication as well. This will be further illustrated in the next section.

Effectiveness of agents increases exponentially as more machine-readable web-content and automated services become available [LEE2001]. Agents need to be able to exchange proofs of information. These are explanations of how results are deduced. They also will need digital signatures to verify that their information is provided by trusted sources [LEE2001].

5. Ontology issues

Semantic web, augmented with ontologies, is a promising technology for next generation multi-agent systems that serve humans in all aspects of life. AI has the basic components needed for knowledge engineering, but those components need to be linked into a single global system. A powerful language is needed to express semantic web knowledge, and to allow existing knowledge representation systems to be exported onto the web. When this is achieved, it will be possible to make inferences, choose courses of action and answer questions. This is the challenge and

the task of the semantic web community nowadays [LEE2001].

One major draw back of ontologies will be the performance overhead that occurs due to the need for tracing URIs and capturing information over remote sites. Although this is not needed in all page accesses, the growth of semantic web applications is based on this roaming through web links [LEE2001].

Another problem stems from the difficulty of capturing all the concepts of a domain and organizing them in a semantically useful manner. Different ontology designs have different levels of flexibility usability and richness. There is a need to sketch standardized methodologies and specifications to reduce design differences.

Integrating two different ontologies designed differently is another challenging issue, as this requires intelligence and domain expertise to recognize the best ways of merging schemas. Our work in [ALZ1995] demonstrates that this is not an easy task. Our ontologies must be self described, and dynamically designed in order that they maybe merged or integrated. Human intervention in this process has to be eliminated.

Ontology evolution is needed to accommodate future modification on the ontology. This process may lead to changes on the design of the ontology, which must be implemented carefully to avoid disrupting an existing version of an ontology. Although this might sound similar to the popular schema evolution issue, it is quite different since an ontology may contain rules and assertions that are not usually part of a schema.

Creating ontologies is conducted by establishing a repository of terms with inherent relations constraints and rules. The current trend is towards developing editors for creating domain-specific ontologies and translators for mapping concepts between ontologies. CLASSIC, Protégé, Chimaera and ontolingua are only few examples [PER2002] . Efforts are underway to build domain ontologies, e.g. CYC, SNOMED, UNSPSC, RosettaNet, DMOZ, etc; see [NOY2001]. There is a need for international efforts to unify ontology designs and standards. Without this, multi-

agent systems will always be limited by discrepancies of dispersed ontologies.

Despite the existence of preliminary and general standards and languages for agent knowledge specification and exchange, such as (e.g. FIPA ACL, KQML, RDF and DAML by DARPA and W3C), this work is still in its infancy, and more research is needed. Standards are especially essential for particular domains as a first step, before better understanding is reached and general standards become feasible.

6.   Conclusion and future work

Handling the ontology issue requires defining a unified knowledge model that includes the structure, the associations the restrictions and also the reasoning assertions. This model should be supported by flexible tools to create expand and use ontologies. Tools must not only allow this but in addition should facilitate connectivity and integration with similar ontologies. In order for this to happen, ontology models should be standardized and tools architecture and interfaces have to be fully published.

We believe that exploiting database, knowledgebase and AI technologies is the way forward to tackle the semantics issues in modern applications. Interdisciplinary efforts on this arena will bring about innovative methods and technologies that will benefit the next generations of web users.

Our future research tackles such issues as ontology methodologies, agent interoperability using ontologies and ontology mapping and evolution. Due to ontology's tight coupling with linguistic characteristics, our emphasis will be on techniques that are more relevant to Arabic ontologies.

References

[ALZ1998]   R. Alzahrani, et al, PROMAQ: An Object-Oriented Multidatabase Tool Utilizing Semantics, Symposium of Computer networks and Distributed databases, Riyadh, 1998.

[ALZ1996] R. Alzahrani, Semantic Object-Oriented Multidatabase Access, Ph.D. Thesis, University, of Wales, Cardiff, 1996.

[ALZ1995] R. Alzahrani, et al, Integrity Merging in an Object-Oriented Federated Database Environment, BNCOD95, Manchester, 1995.

[BOS2002] J Bosak, T Bray, XML and the Second-Generation Web, Scientific American Special Online Issue, 2002.

[BRA1998] T Bray, et al, Extensible Markup Language (XML) 1.0, W3C, 1998.

[DIN2002] Y Ding, S Foo. Ontology Research and Development: Part 1 – A Review of Ontology Generation. *Journal of Information Science* 28(2), 2002.

[DUI1999] A Duineveld, et al, Wonder Tools? A comparative study of ontological engineering tools, http://www.swi.psy.uva.nl/wondertools.

[FEN2002] D Fensel, et al, On-To-Knowledge in a Nutshell. To appear: *Special Issue of IEEE Computer on Web Intelligence (WI), 2002.*

[GUT1999] R Guttman, et al, Agent-mediated Electronic commerce: A Survey, http://agents.umbc.edu/introduction/hn-dn-ker99.html.

[KOT1999] D Kotz, R Gray, Mobile agents and the future of the Internet, ACM Operating Systems Review 33(3), 1999.

[LAS1999] O Lasilla, R Swick, Resource Description Framework (RDF) Model and Syntax Specification, http://www.w3.org/TR/1999/REC-rdf-syntax-19990222.

[LEE2001] T Berners-Lee, et al, The Semantic Web, Scientific American Special Online Issue, 2001.

[LEE1998] T Berners-Lee, Semantic Web Road Map, http://www.w3.org/DesignIssues/Semantic.html, 1998.

[LEN1995] D Lenat, CYC: A Large Scale Investment in Knowledge Infrastructure, CACM 38(11), 1995.

[MAE1994] P Maes, Agents that Reduce Work and Information Overload, Communications of the ACM 37(7), 1994.

[Noy2001] N Noy,D McGuinness, Ontology Development 101: A Guide to Creating Your First Ontology, http://protege.stanford.edu/publications/ontology_development/ontology101.html, 2001.

[NWA1999] H Nwana, D Ndumu, A Perspective on Software Agents Research, http://agents.umbc.edu/introduction/hn-dn-ker99.html, 1999.

[PER2002] A Perez, et al, A Survey on ontology tools, OntoWeb, http://babage.dia.fi.upm.es/ontoweb/wp1/OntoRoadMap/documents/D13_v1_0.pdf, 2002.