INTEGRATING A KNOWLEDGE MANAGEMENT SYSTEM WITH DATABASE

Sujoy P. Paul

Process and Information System Engineering, School of Engineering, University of Surrey, Guildford, GU2 7XH, UK

Phone: 44 1483 686568; Fax: 44 1483 686291; email: s.paul@surrey.ac.uk

INTEGRATING A KNOWLEDGE MANAGEMENT SYSTEM WITH DATABASE

Sujoy P. Paul

Process and Information System Engineering, School of Engineering, University of Surrey, Guildford, GU2 7XH, UK

Abstract

The edifice of a knowledge management system resides on the building blocks of ontology. This paper describes the usage of ontology and its efficient usage in Internet marketing by storing knowledge in a database that is independent of the database vendor and consequently retrieves instances in XML format across the Internet. It describes the various concepts, relations, axioms, instances involved in building ontology and subsequently exporting the ontology as a knowledge base into a database.

Keywords: knowledge management, ontology, concepts, database, e-marketing, E-R diagram

Introduction

The spectacular growth of the World Wide Web (WWW) in the recent past has ushered in an era in which the development of the Semantic Web is being regarded as the key to harness the plethora of information available into tangible knowledge.

Tim Berners-Lee (Berners-Lee, 1999) envisioned the idea of the Semantic Web in which automated agents would scour for intelligent information, whereby software processes would interpret machine-understandable data linked to the web. The manifestation of the idea is inexplicably linked to the annotation of the web resources with metadata. Standard languages, like the RDF, RDFS and DAML+OIL (The DARPA Agent Markup Language Homepage, 2003), for such annotation are being developed under the auspices of World Wide Web Consortium (W3C). However, automated processes would not be able to make use of the said languages unless there is a common vocabulary. It is in light of this context, that ontologies have come to be regarded as the building blocks of the semantic web that can be communicated across people and machines.

Ontology can be described as an explicit specification of a shared conceptualization (Gruber, 1995). This means that the concepts and relations along with the axioms can be shared for a particular domain of knowledge. Knowledge engineers in confluence with domain experts develop ontologies by tapping into the tacit knowledge that is inherent in each of the experts. However, the marketing paradigm of ontology not only encompasses concepts, attributes and rules that constitute the schema but also entails instances or data.

This has heralded the use of knowledge management as a tool in the efficient and intelligent usage of Internet marketing since, as envisaged by Hai Zhuge (Zhuge, 2002), it enables the uniform sharing of knowledge management across the Internet by developing a semantic web-based knowledge grid. Using the resource space model and the semantic web that permits the machines to be interoperable, participants of the market are allowed to store their knowledge and retrieve the desired knowledge from the repositories across the Internet.

Abramson and Hollingshead (2001) has classified Internet users as Surfers and Shoppers and elucidated the pros and cons of marketing in the Internet while maintaining that product, price, promotion and place as the 4 traditional principles of marketing are still applicable to e-marketing. Although the prices of goods on the

Internet are low since the cost of maintaining a web site, unlike a retail store, reduces the overhead costs, thereby allowing the development of virtual stores as in the case of shops like Amazon.com, the benefits can only be accrued if the primary concerns of product quality and transactional security of customers are addressed.

Customer satisfaction of product quality can be addressed by communicating a fair return policy and providing assurances of consumer protection made by credit card companies (Abramson and Hollingshead, 2001). The semantics of communication can be alleviated by developing a common vocabulary, and hence the development of ontology for the specific business domain, that traverses the industry.

The vast number of mission critical databases like sales automation, financial, payroll along with semi-structured and un-structured data in emails, spreadsheets and knowledge nuggets in companies make it ever more imperative to seamlessly integrate these datasources that will provide the users the one-stop interface. However, applications although developed with the ability to interconnect with different datasources do not necessarily eliminate the syntactic incompatibilities of data. Maier **et al** (Maier, 2003) contented that since ontologies have the ability to provide a conceptual basis for communicating knowledge, the goal of integration is to consolidate the distributed data free from redundancy that will provide intelligent feedback and hence added value to both the buyer and the seller. The use of ontologies as the primary strategy facilitates the understanding, communication, change and measurement in the development of e-business (Osterwalder and Pigneur, 2002).

While N-tier architecture using EJB (Enterprise Java Beans) in conjunction with network protocols can process the business logic and transfer data it does not provide the semantic sophistication of ontologies provided by languages such as

RDFS (Staab, 2000) and hence does not have the capability to leverage the meaning of the data. It is this ability to provide a common vocabulary drawn from varied data sources that make ontology an attractive paradigm in interacting between user interface and data sources.

This paper shows the export of ontologies to a relational database and the data is then formatted in XML for users to access the data virtually in the web and its usefulness in Internet marketing.

Development of Ontology

The ontology developed is a representation of the patient management system using OntoEdit (v2.64) provided by Ontoprise GmbH. This is shown in Figure 1. Health care givers will have the ability to monitor the progress of the patient online anytime and at anyplace without the need to be present with the patient enhances efficiency and reduces the cost. The ontology currently has 32 concepts and 94 relations and the ontology is expanding. Concepts in ontology are referred to real-world entities that encompass the conceptual description of a domain. The subsumption relation of concepts are similar to is-a or inheritance relationship of classes in the object-oriented programming paradigm (Booch, 1994). The concepts have been classified according to the kinds of patients, the different type of symptoms, the existing healthcare givers, the prevalent types of drugs and the medications, the kinds of hospitals and other people employed in the health

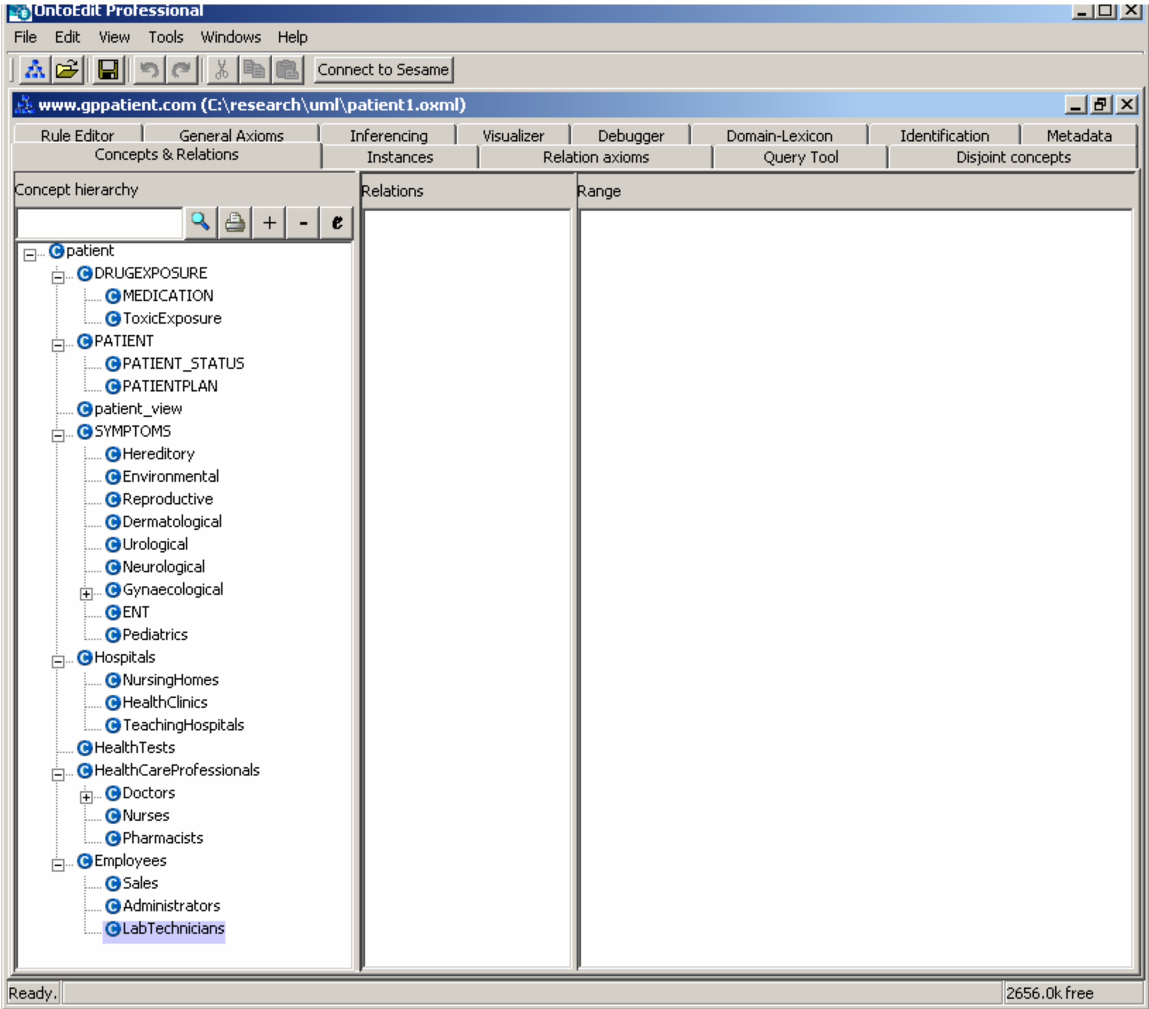

Figure 1. The Ontology of a Patient Management System

sector. The generalization/specialization relationship developed assumes that the criteria developed in a parent concept are automatically inherited by the child concept along with its own unique criteria. Similarly, it is also interpreted that the instances of a child will also be instances of the parent. Those concepts that are not interlinked are classified as disjoint. Such constraints are used to check the consistency of the ontology. As such, there cannot be instances of one concept that exists in another. This reduces the semantic heterogeneity and makes the knowledge base more precise and compact. Concepts like Consultants and Nurses, as shown in the following figure, are disjoint concepts.

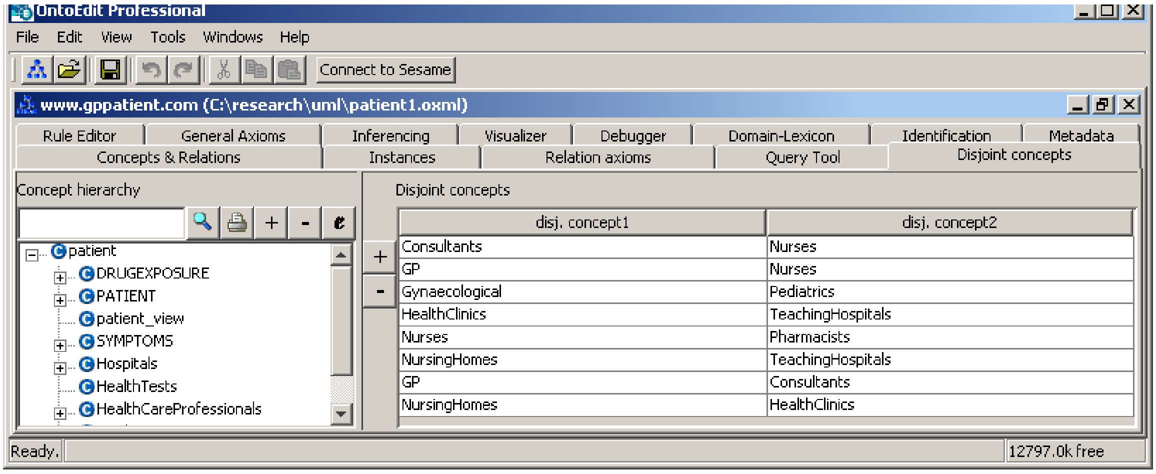

Figure 2. Examples of disjoint concepts in the ontology

Adding similar terms to the lexicon of the domain can expand the ontology further. The addition of synonym gives added meaning to the established concepts and where necessary filtering out erroneous synonyms (Ontoedit manual, 2003). This would reduce the ambiguity of terms and eliminate any problems when integrating with external databases where such redundancy of data is abundant. Terms like SkinDisease and SkinCare is regarded as analogous to Dermatological. While such terms and concepts can be stored in different languages, homonyms or a single concept having several meanings or criteria is not available.

The attributes or relations can be regarded as part-whole (part-of) of the entities. In case of the concept ToxicExposure, a child of DrugExposure, it inherits all the parent's relations and adds the following which are unique:

(a) carbon-monoxide poisoning is part-of toxic exposure

(b) exposure to toxic chemicals like solvent is also a part-of toxic exposure

(c) asbestos is also toxic to humans

The relations in the knowledge base can be used to generate rules that are valid in the domain. Rules based on symmetric, transitive and inverse are developed. So according to an inverse relationship if you have DiabetesMellitus then you need to take insulin and conversely if you take insulin you have DiabetesMellitus. A symmetric relation, rel, is a binary relation that can be expressed as:

(rel inst1 inst2) <=> (rel inst2 inst1) for all inst1 and inst2

This is represented in the ontology as (rel Urine urologist) ⇔ (rel urologist Urine)

A transitive relation is represented as (rel inst1 inst2) and (rel inst2 inst3) ⇔ (rel inst1 inst3).

Business Logic

The consistency and the purpose of the knowledge base can be enhanced by implementing business logic in the knowledge base. This is facilitated by the use of F-Logic which acts as a bridge between the model and the logic since it covers ontological data and the rules (Maier, 2003). This is made possible by the fact that every object in the ontology can be written in F-Logic atoms or molecules that in turn can be embedded into logical rules (F-Logic Tutorial, 2003). Here we illustrate 3 business rules that have been implemented in the knowledge base

(1) Any patient that has the symptoms of flu would have both fever and sworethroat, as demonstrated with the use of in-built function “concat”. This is represented as follows:

```
FORALL X,Y,Z1,Z2 (X:symptoms[flu->Y:patient])
<- ( (X[fever->Z1]
   and (Y[sworethroat ->Z2]
   and concat (“flu”,Z2,Z1))) ).
```

(2) Anyone who is pregnant, then the gender is a woman:

FORALL X,Y (X:obstetrics[pregnant ->>Y])

<- (Y:patient[gender ->woman]).

(3) In this scenario if you need to have any insurance plan regarding healthcare then you need to be a minimum of 18 years of age:

FORALL X(message (“Warning: The patient’s age needs to be 18 or above

To have a patient’s insurance plan”))

<- (X:Patient(minimum(age,18))).

F-Logic can also be used to query the instances and the schema as mentioned before. If we want to retrieve the list of symptoms in the knowledge base then we execute the following:

FORALL X <-X:”www.gppatient.com”#SYMPTOMS

(where www.gppatient.com is the namespace used at the time of the creation of the ontology)

Some of the output generated is shown below:

“www.gppatient.com”#PD

“www.gppatient.com”#”Alzheimer’s Disease”

…….

Figure 3. Output of the query generated by F-Logic

Integration of Ontology to Database

The ontology developed can now be integrated into a new database. This is the knowledge, and hence the vocabulary of the data, that is implemented into the database and not just the data. As such, by acting as the central factory which interconnects different data sources and rules it expedites the integration of the semantics of data and reduces the redundancy of the data. It is this knowledge integrator that, in fashioning an e-business strategy, will help in the confluence of information technology and product innovation that builds market share (Malhotra, 2000).

Since today's best practices can end up as in-built system rigidities for corporations in the future, ontologies can help eradicate the future problems of legacy data by exporting database schema specific to the RDBMS since they are independent of database vendors. While databases like SQLServer, Oracle and IBM DB2 are built on the basic model of relational tables, each of them have their own extensions of SQL(T-SQL in case of SQLServer, PL/SQL in case of Oracle) and particular functions. The automatic generation of the SQL code by OntoEdit specific to the RDBMS allows for the rapid integration of the knowledge to the database. The use of ontology as a formal approach to model such business processes helps both the managers and the IT staff to effectively communicate and hence share the business logic between themselves (Fensel, 2001). Added to this, the semantic heterogeneity that is prevalent in biological (Schweigert, 1995) databases can be removed by building the ontology first although the syntactic heterogeneity (Verschelde, 2003) for each of the RDBMS remains. The consecutive E-R diagram is shown in Figure 4.

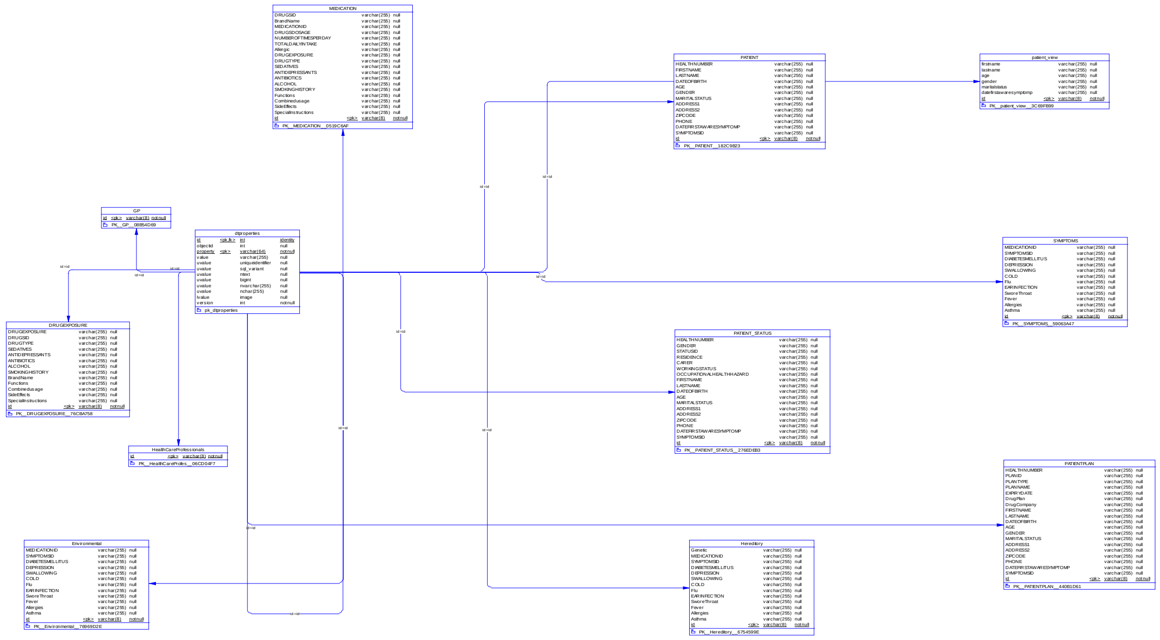

Figure 4. E-R diagram of the knowledge base generated

In this case, the concepts are interpreted as tables while the attributes or relations are regarded as the columns of the table. Unlike an E-R model, which is not an object model, ontologies can read taxonomies while like the E-R model (Batini, 1991), ontologies have the capability to distinguish schema from facts. This enables the facts to be stored in the database. In addition, the ontology does not change should the database needs to be altered thereby saving time and reducing maintenance costs. As such, since the data in the database is concurrently available it is automatically accessible to the ontology. This facilitates the seamless integration of data in the database to the instances of knowledgebase.

MSSQLServer 2000 allows data to be retrieved by HTTP over the Internet. This is made possible by writing data in XML format. Although data can be replicated across heterogeneous datasources(Paul, 2003), XML permits the exchange of a wide variety of data on the Web and elsewhere(www.w3.org/xml,2003). The facts stored in the database can then be retrieved from anywhere, anytime and anyplace. A virtual directory has been created using the IIS (Internet Information Services) Virtual Directory Management for SQLServer as shown below:

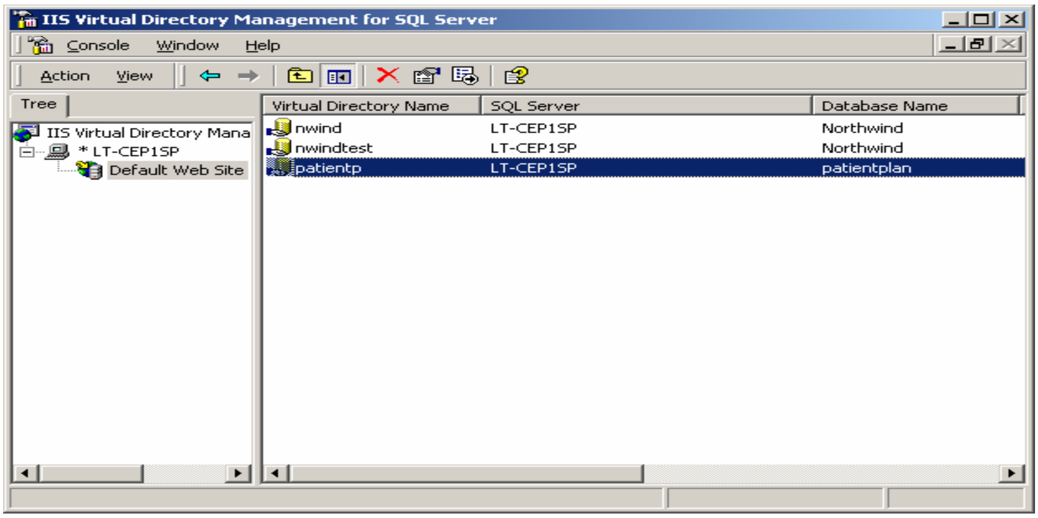

Figure 5. Virtual directory of the patientplan database

The virtual directory patientp is mapped to the patientplan database that stores the knowledge that we have created. It is then possible to retrieve the instances of the knowledge in XML format. The T-SQL (Transact SQL of SQLServer) statement that enables the retrieval of all the instances in the table patient can be written in XML format as:

Select * from patient For XML auto

The queries when specified directly in the URL returns the following result set as shown in Figure 6:

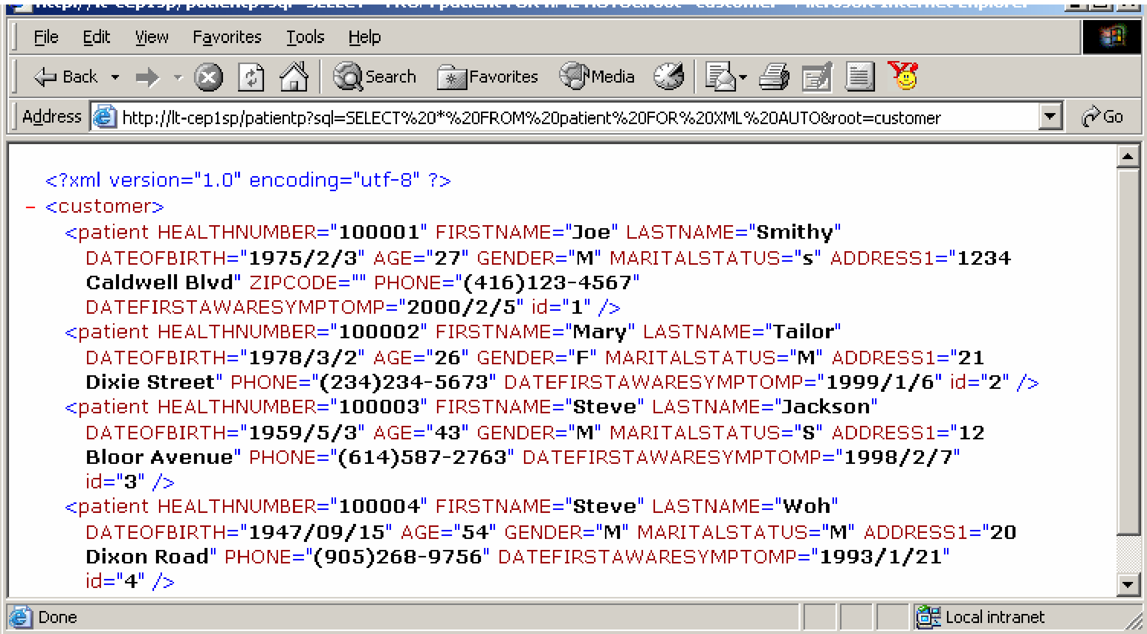

Figure 6. Reading of instances in XML format for the patient management system

In this case, the root parameter is specified in the Internet Explorer (v.6.0.26) browser to demonstrate the top level element root is a customer:

http://lt-cep1sp/patientp?sql=SELECT FROM patient FOR XML AUTO&root=customer

Conclusion

The ontology of the patient management system developed with OntoEdit has shown to integrate seamlessly with the database. The knowledge base was expanded by the addition of symmetric, inverse and transitive relations and rules were derived by writing F-logic rules. The instances of the knowledge base were then read in XML format and accessed over the Internet.

References

Abramson, J., Hollingshead, C. (1998), "Marketing on the Internet – Providing Customer Satisfaction", Journal of Internet Marketing, April 1998, vol. 1, No.1, (Available http://www.arraydev.com/commerce/JIM/9802-01.htm)

Batini, C., Ceri, S.(1991),Conceptual Database Design: An Entity-Relationship Approach, Addison Wesley, 1991

Berners-L, T.(1999), Weaving the Web, Orion Business Books, London, 1999

Gruber, T.H. (1995),"Towards principles for the design of ontologies used for knowledge sharing, International Jounal of Human-Computer Studies, 43 (5/6), 1995

Booch, G.(1994), Object-oriented analysis and design: with applications (2nd ed.), Benjamin/Cummings, Redwood City, California, 1994

F-Logic Tutorial (2003),"How to Write F-Logic Programs", Ontoprise GmbH, (Available http://www.ontoprise.de/documents/tutorial_flogic.pdf)

Fensel, D. (2001), Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce. Heidelberg, Springer-Verlag

Maier, A., Schmurr,H-P., Sure, Y.(2003),"Ontology-based Information Integration in the Automative Industry", ISWC2003, Florida, USA.

Malhotra, Y.(2000), "Knowledge Management for E-Business Performance: Advancing Information Strategy to "Internet Time"", Information Strategy, The Executive's Journal, vol. 16(4), Summer 2000, pp. 5-16.

Ontoedit manual (2003), Ontoprise GmBH

Osterwalder, A., Pigneur, Y.(2002), "An e-Business Model Ontology for Modeling e-Business", 15th Bled Electronic Commerce Conference, June 2002,(Available http://inforge.unil.ch/aosterwa/Documents/eBusinessModels/Publications/bled02.htm)

Paul,S.(2003),”Replicating SQL Server 2000 across Heterogeneous Databases”, www.sqlteam.com, 2003

Schweigert, S., Herde, P. Sibbald, R., “The Difficulties of Handling Biological Data”, CABIOS, 1995, vol11, no.4, 1995

Staab, S., Erdmann, M., Maedche, A., Decker, S.(2000),“An Extensible Approach for Modelling Ontologies in RDF(S)”, ECDL 2000 Workshop on the Semantic Web, Lisbon, Portugal, 2000

The DARPA Agent Markup Language Homepage (2003), www.daml.org, 2003

Extensible Markup Language Architecture domain (2003), www.w3.org/xml,2003

Verschelde,J-L., Santos, C., Deray, T., Smith, B., Ceusters, W., “Ontology-assisted database integration to support natural language processing and biomedical data-mining”, Symposium on Integrative Bioinformatics, Bielfeld, 2003

Zhuge, H (2002), “Knowledge Grid: Enabling Intelligent Internet Marketing”, Journal of Internet Marketing, September 2002, vol.3, No.1,(Available http://www.arraydev.com/commerce/jim/articles.htm)