



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

A Framework for Mining Ontological Information Nets

¹Saqib Saeed and ²Christoph Kunz

¹Bahria University Islamabad, Pakistan

²Fraunhofer Institute for Industrial Engineering Stuttgart, Germany

Abstract: This study describes a framework for mining ontological data. In order to visualize the complete list of relationships from a semantic network it is mandatory to explore the hidden relationships. It is the biggest problem that semantic networks don't have a mechanism to find hidden relationships. In this study a framework was presented to extract the hidden relationships from semantic networks. This research was implemented on Matrix Browser which is GUI for information visualization.

Key words: Semantic web, ontology, matrix browser, information visualization

INTRODUCTION

The field of information visualization has gained more importance in current era than ever before. The reason is that in old ages there was not much information present in electronic format. There were no efficient systems and storage media present to store data. But with the advent of efficient systems problem of storing data has been reduced. Secondly with the passage of time the amount of electronically stored data has increased. It is getting more and more difficult to extend usable information from this huge mass of data. This problem triggered the need for better visualization techniques so that the results, which are presented by the information, should be easily understood and there would not be difficulties in deducing the conclusions from that information^[1].

Secondly the concept of Semantic web is gaining tremendous popularity. The Semantic web is an extension of current web in which information is given well-defined meaning, better enabling computers and people to work in co-operation^[2]. The Semantic web means that websites are designed in a way that the information present on the websites can be processed by the computer systems. Mostly the current websites are implemented in HTML and there is not much possibility of processing a site with the help of computer programs. One can use it only for the viewing purposes but it is impossible to generate any result on the basis of the relationships present in the contents of websites with the help of computer systems. Therefore the need was felt to change the whole World Wide Web into Semantic Web.

PROBLEM DEFINITION

There are two standards for describing Ontologies in computer readable format. First is DAML+OIL and the

second is XML Topic Maps^[3]. The main problem in its way to implementation was that the ontological data has the explicit relationships defined by using the language constructs. But when we want to use information visualization on this data there should be some mechanism by the help of which we can identify and extract the hidden relationships present in the network. The typical examples could be relationships which could be extracted due to transitive property or hierarchical relationships.

In order to make a system where we can apply information visualization techniques on ontological data the plan to enhance the functionality of the Matrix Browser^[3] was formulated. The Matrix Browser is a GUI which shows networked information spaces on the basis of adjacency matrix approach. In adjacency matrix approach, nodes are defined at horizontal and vertical axis (Fig. 1). The relationship between different nodes is

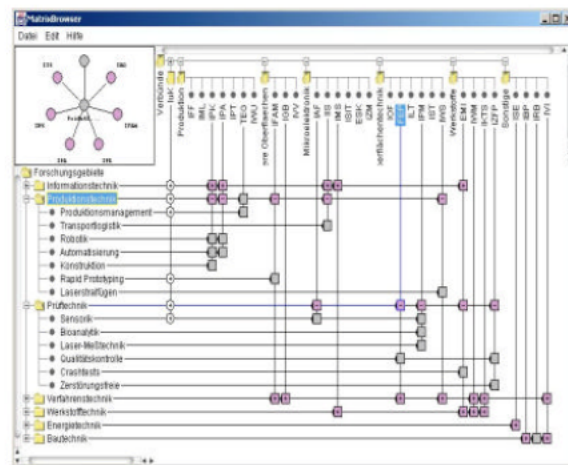


Fig. 1: Matrix Browser snapshot

determined by corresponding cells where respective rows and columns intersect each other. The direction of the relationship is defined by arrow like shape on the square box representing the relationship^[4].

SOLUTION

The problem of finding hidden relationships was solved by introducing an inference engine as a middle layer. All the data which is being extracted from ontologies is being transformed into logic based framework and then by using the capabilities of the inference engine all the hidden relationships are extracted. The mapping of DAML + OIL data to Flora-2 can be understood by the help of an example. Suppose we want to declare a class named Book In DAML + OIL it would be like^[5].

```
<daml: Class rdf: id = Book>
<rdfs: label>Book</rdfs: label>
</daml: Class>
```

Whereas the same class can be defined in Flora-2 as Book.

The main architecture of the complete system is described in Fig. 2.

The DAML+OIL data is being transferred in Flora-2 file format. In order to perform the transition Protégé API is being used. Protégé 2000 is a tool developed at the Stanford University under a research group named as Stanford Medical Informatics. It is an ontology editor and a knowledge-base editor that helps in creating the different ontologies.

The data is transformed into an intermediate state which is readable to Protégé 2000 editor and then this data is transferred into Flora-2 file format. Flora-2 is an object-oriented language used in the field of artificial intelligence

to maintain the deductive databases. It is an extension of F-logic language and can be run under windows and UNIX. By using the artificial intelligence capabilities of F-Logic all the hidden relationships can be extracted. The data present in Flora-2 is in form of classes, objects and different relationships between classes and objects^[6]. The relationships can be inherited or defined in objects and classes. The nodes are interconnected with each other. By using the interprolog API the complete data is forwarded to the Hierarchy Extractor component^[7]. This component refines the data and makes the trees from the data by finding the root nodes. The hierarchies are then constructed from these trees as trees may be overlapping or circular.

CONCLUSIONS

This framework is being presented in order to extract and visualize a knowledge base. With the help of this framework hidden relationships (e.g., due to transitive properties, subclass etc.) in a semantic network can be extracted and visualized. This is being done on a visualization tool called Matrix Browser.

At present the matrix browser is only capable of just querying knowledge base once. This research can be enhanced by implementing a framework by which user can define semantic queries from inside the visualization and sending the queries to knowledge base for information retrieval, then retrieved data is displayed in the matrix browser.

The import of XML topic maps can also be made possible. The DAML + OIL and XML topic maps are two standards for describing the ontological information. Both standards are used by different group of people. So in this way both standards can be handled^[8].

The matrix browser can be developed into a web service. The matrix browser can work on client side. The inference engine can be encapsulated in J2EE application server. The inference engine and the DAML+OIL file can serve as two tiers at server.

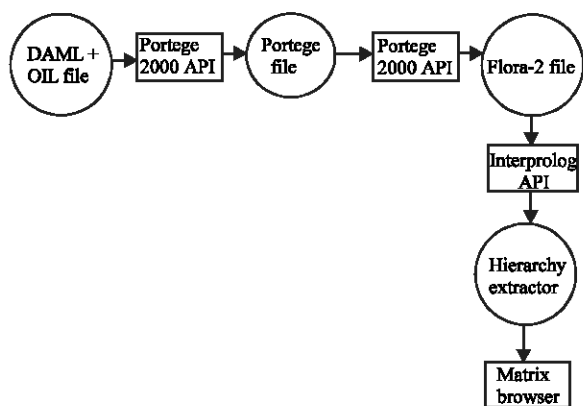


Fig. 2: System architecture

REFERENCES

1. Card, S.K. and J.D. Mackinlay and B. Shneiderman, 1999. Readings in Information Visualization Using Vision to Think. Morgan Kaufman Publishers, pp: 231-235.
2. Berners-Lee, T., J. Handler and Ora Lassila, 2001. The Semantic Web. Scientific American, pp: 34-43.
3. Garshol, L.M., 2005. Topic Maps, RDF, DAML, OIL-A Comparison. [Online]: <http://www.ontopia.net/topicmaps/materials/tmrdfoidaml.html>.

4. Ziegler, J., C. Kunz and V. Botsch, 2002. Matrix browser-visualizing and exploring large networked information spaces. Conference on Human factors in Computing Systems, Minneapolis, Minnesota, USA, pp: 602-603.
5. Saeed, S., 2003. Java based Visualization and Query of XML Concept Nets. University of Applied Sciences Stuttgart, Germany, pp: 20-50.
6. Saeed, S. and C. Kunz, 2004. Loading arbitrary knowledge bases in matrix browser. IEEE International Multi Topic Conference Lahore, Pakistan, pp: 631-634.
7. Calejo, 2005. Interprolog: A declarative Java-Prolog Interface. [Online]: <http://declarativa.com/interprolog/default.htm>.
8. Ouellet, R. and U. Ogbuji, 2005. Introduction to DAML. [Online]: <http://www.xml.com/pub/a/2002/03/13/daml.html>.