European Research Studies,
Volume XII, Issue (4), 2009

Modeling the Data Entities used in the Financial Diagnosis through Onthologies

Ph.D Obancea Georgiana Daniela*, Ph.D Professor Ilie Tamas*

Abstract:

The field of economy and finance is a conceptually rich domain where information is complex, huge in volume and a highly valuable business product by itself.

The correct perception of the financial condition of a company is dependent upon the quality of its information processing system and also an important role has the analyst's expertise. Monitoring and correcting inappropriate conditions become critical tasks, particularly to small and medium companies, where the presence of an expert is not affordable. A system capable of spotting the condition and suggesting alternative actions to control anomalies brings important advantages to the company. Effective financial diagnosis requires accurate, timely and relevant information.

The ideal system will be one able not only to analyze the financial problems, but also to suggest the solutions. In order to accomplish these tasks, the artificial intelligence technologies offer a valid solution. In this research, we propose some possibilities to identify and integrate the data structures involved in the financial diagnosis, through the use of onthologies.

1. Introduction

In order to identify and design a system for the financial diagnosis, we will follow an approach based on onthologies. This brings with it the following advantages:

- The use of onthologies allows a better representation of the involved data structures
- A higher level of abstractization, obtained by separating the data model from the structures used to store the data
- Extensibility and reusability
- The representation of hierarchies

**Academy of Economic Studies Bucharest, Romania*
georgianaobancea@yahoo.com

The onthologies offer few levels of abstractization. The top-level onthologies define the common terms for all the applications and domains. The mid level onthologies define the specific terms of a domain, and the low level onthologies define the terms specific to the applications.

Also, in order to ensure the integrated access to heterogeneous data sources, we need to define a global model of the domain in order to define the specific concepts. This way, each of the individual sources will be connected to this global model.

2. Financial Diagnosis Theory

The financial diagnosis is the classification task performed when a subject makes a judgment of the financial situation of the firm based upon information from the financial statement. This task is performed in several contexts such as bankruptcy prediction, loan decision context.

By analyzing the existing literature we can notice several approaches. The traditional studies in accounting and finance focused on a predictive approach, using on a large scale statistical instruments and other models to predict the environmental output.

In order to design and implement an information system for the financial diagnosis, the first step is to make a representation of the domain and identify the structures involved. This approach in representing the data structures opens wide perspectives to the knowledge based information systems. Considering the area of financial diagnosis, we need to integrate data coming from various sources:

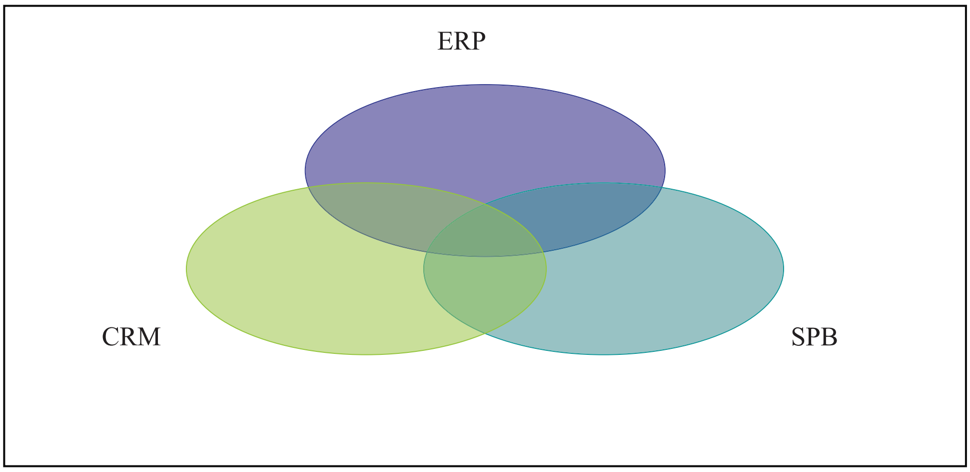

Figure 1: The data sources used in the financial diagnosis

Until recently, the effort to identify and treat financial and operational issues relied on being able to isolate exceptions and identify trends within a nearly unmanageable mass of data.
During the problem definition and data preparation task, the data that are relevant to the problem should be collected from the available sources. The data from different resources should be transformed and merged. This step may be not easy and take much time in practice, since the data are usually not in the same format; some data are even not in the electronic format.

The widespread adoption of web related technologies has increased the need to immediately solve the problem of integrating heterogeneous information sources. There are a large number of contributions to the problem of accessing heterogeneous information sources. These contributions typically address special sub problems of interest, like data warehousing textual and semi structured data access. The recent researches focus on the problem of semantic integration by some means of knowledge representation.

3. Representation of Onthologies in Owl

Onthology is a description of entities of a problem domain and their semantic relations. The onthologies are very useful for defining data structures and also the links between them. For onthology development, languages and technologies are dedicated: Resource Description Framework RDF and schema RDFS, Web Onthology language OWL which consists of three sublanguages: OWL Lite, OWL Description logic and OWL Full.

Onthologies expressed in a standard language, as OWL, ensures interoperability at a high level, taking into account the rich semantic support and the possibilities of inheritance offered.

OWL is based on Description Logics-a family of knowledge representation formalism based on first-order logic. It is a family of three onthology languages: OWL-Lite, OWL-DL and OWL –Full. OWL can model only domains where objects are connected in a tree-like manner, but many real world applications (E.G financial applications) require modeling general relational structures.

The basic modeling constructs of OWL are roles, which correspond to unary and binary predicates. However, many relationships encountered in practice are of arity larger than two. For example, the assets have at the beginning of the financial year a certain value, and this can be represented through the following predicate: Value (Asset, Year, Asset Value)

Prototypes of OWL reasoners, such as RACER have been implemented and applied in research projects, but the practical experience showed several drawbacks of OWL. For example OWL does not allow for integrity constraints. The Rule Interchange Format (RIF) workgroup of the W3C is working on standardizing such a language.

In 2004, the SWRL (Semantic Web Rule language) was proposed. But this is a simple extension of OWL and still cannot allow expressing integrity constraints. The recent direction of the researches in this area is to use OWL and rules side-by-side. The Web Service Modeling Language (WSML) and F-Logic had been proposed as suitable onthology languages based on logic programming.

Rule based formalism based on logic programming has been proposed as a possible solution; so adding a rule layer on top of OWL might be the solution.

Rules provide the mechanism to perform automated reasoning, with the available theory and technology. This has been identified as a design issue for the Semantic WEB, as stated by Tim Berners Lee, director of W3C:

"For the Semantic WEB to function computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning. The challenge of the Semantic WEB, therefore is to provide a language that expresses both data and rules for reasoning about the data and that allows rules from any existing knowledge representation system to be exported onto the WEB "(Tim Berners, 2001).

4. The RDF Support

Onthology is the basis for constructing common understanding though explicitly defined relations.

RDFS and DAML+OIL languages can be used to define the onthology. DAML is a language created by DARPA as onthology and inference language based upon RDF. DAML takes RDF Schema a step further, by giving us more in depth properties and classes.

RDF provides interoperability between applications that exchange machine-understandable information on the WEB. RDF uses XML to exchange descriptions of WEB resources and emphasizes facilities to enable automated processing. In RDF the information maps *directly* and *unambiguously* to a model, a model that is decentralized, and for which there are many generic parsers already available. By this, we will know which bits of data are the semantics of the application, and which bits are just parts of the syntax of the application.

The RDF descriptions provide a simple onthology system to support exchange of knowledge and semantic information on the WEB. RDF schema provides the vocabulary to describe RDF documents. RDF schema can be used to define properties and types of the WEB resources. In a similar fashion to XML schema that gives specific constraints on the structure of an XML document, RDF Schema provides information about the interpretation of the RDF statements. RDF uses triples written as XML tags to express the information as a graph. These triples consist of a subject, property and object, which are like the subject, verb and direct object of a sentence. (Some sources call these the subject, predicate and object)

> The Quick ratio has as Numerator the Assets easily convertible into cash.
> The Quick ratio has as denominator the Current liabilities

For the information above the translation in terms of RDF terms will be:

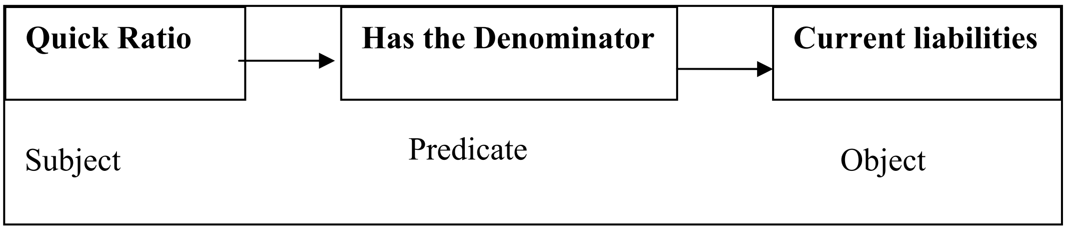

Figure 2.Example of an RDF triple

We present bellow an RDF graph designed for the area of financial analyses and diagnosis.

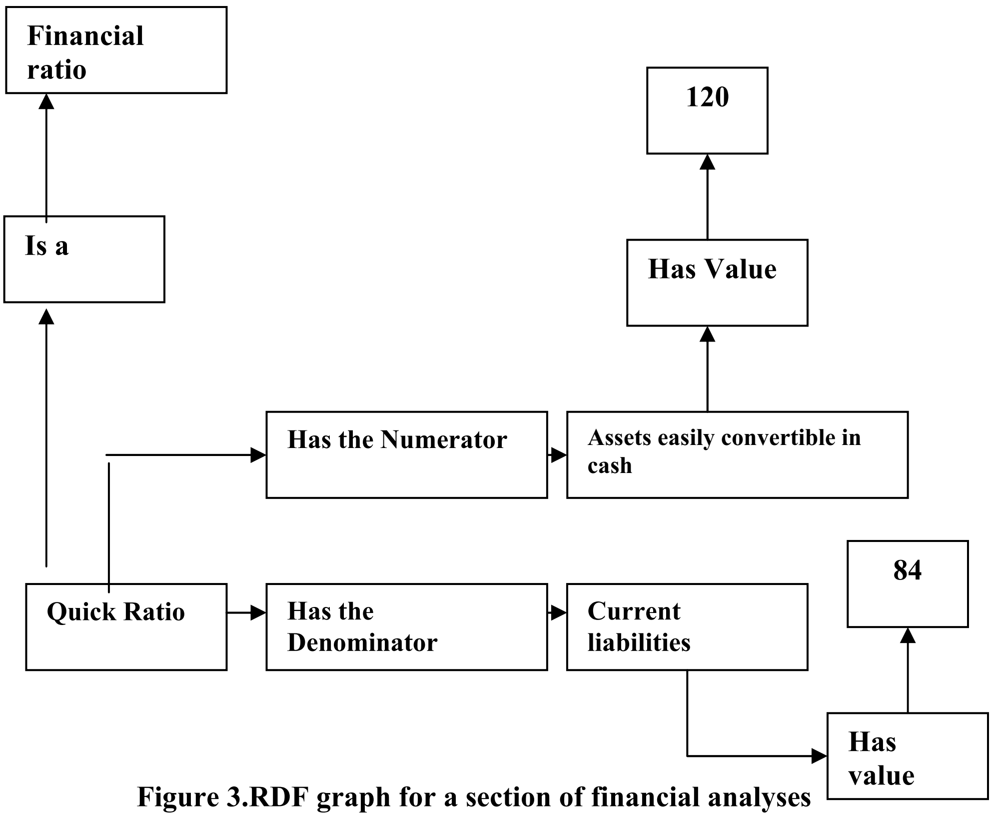

Figure 3.RDF graph for a section of financial analyses

RDF already exists on the Web - for example, it is part of RSS feed creation. Even with the framework that XML and RDF provide, a computer still needs a very direct, specific way of understanding who or what these resources are. To do this, RDF uses uniform resource identifiers (URIs) to direct the computer to a document or object that represents the resource. A URI can point to anything on the Web and may also point to objects that are not part of the web, like appliances in computerized homes. Mailto, ftp and telnet addresses are some other examples of URIs.

One of the long-term goals of the Semantic Web is to allow agents, software applications and web applications to access and use metadata. A key tool for doing this is simple protocol and RDF Query Language (SPARQL). SPARQL's purpose is to extract information from RDF graphs. It can look for data and limit and sort the results. One of the advantages of the RDF structure is that these queries can be very precise and get very accurate results.

5. Transposing the Onthological Model into a Relational Model

An important issue arisen from practical needs is the possibility to store the onthological information and to process this information by various applications.

There are three possible approaches to store the onthological data and execute queries on it. One is to build a dedicated database system taking into account the specific of the onthologies. The second solution is to use an object oriented database system exploiting the modeling capabilities of OODBMS.

Given the popularity of SQL, the third approach is to use a relational database system. The Relational Database model has proved its capabilities to manage large amounts of data. In this approach, the OWL classes are mapped to tables of a relational schema, and also the OWL data types and constraints are mapped to the relational database model.

As OWL is built based on XML, it is worth to try transformations from XML to the relational model, but we need to consider the semantic support offered by the OWL and this requires specific rules to apply in order to map the OWL specific constraints and its inbuilt inference support. Onthology helps business develop information systems that are reusable, maintainable and able to share information with other systems. XML can be used as the underlying technology for OWL since it is quite mature. Considering this, storing XML documents in the relational database is not fully applicable, as we need to preserve all the constraint information.

In the existing literature, we can find some algorithms for transposing the OWL data into a relational database. The most known algorithm is OWL2DB, developed by the Department of Computer science from Massachusetts. The proposed algorithm parses the OWL document, determines the root class and the dependent classes depth, and maps them to relational database tables. Also, separate tables are created for disjoint classes. Once the information on table names and attributes is collected, the document is parsed again to collect the values of the instances to populate the database.

6. Conclusions

In the competitive and dynamic actual market scenario, decision-making has become complex and latency may occur in many processes. In addition, the amount of data to be analyzed has increased substantially.

Computers are fundamentally suited for performing mechanical computations, using fixed programmed rules. This allows artificial machines to perform simple monotonous tasks efficiently and reliably, which humans are not so suited to perform. It is true that for more complex problems, things get more difficult, but the recent applications in the field of artificial intelligence showed their ability to analyze and provide solutions even to these complex problems.

In this research we analyzed some ways to identify and represent the data structures involved in the financial diagnosis area, with a strong focus on onthologies and their representation through RDF schema.

RDF and RDFs provide means to describe both descriptive and semantic metadata. On top of RDFs, using its primitives like subClassOf and subPropertyOf, onthology languages like OIL are built.

Unfortunately, the development of a single, universal onthology for subject identification -an onthology that everyone in the world will use to identify all subjects can be just a dream, not reality. In order to achieve the goal of Web-scale semantic integration without compromising the semantic authority of each information contributor, we have to take another step back. What is required is a basis for diverse onthology definitions such that the semantic integration of diverse information expressed in terms of those onthologies is facilitated.

7. Bibliography

1. A. Ioniță, M. Foca, M.Vişan, 2004, "An Integrated Decision Support System for the Evaluation of Risks and Disasters Management for local/regional level", 9th GI & GIS Workshop, La Coruna, Spain, June 25-28.
2. De Bruinjn, 2005, "The Web Service Modeling Language WSML", version 0.2 available at http://www.wsmo.org/TR/d16/d16.1/v0.2.
3. Goedfroy, W., Meersman, R.,De Troyer, O., 1998, "UR-WSDM: Adding User Requirement Granularity to Model Web Based Information Systems", Proceedings of 1st Workshop on Hypermedia Development, Pittsburgh, USA.

4. M. Kerrigan, 2005, "The Web Services Modeling Toolkit (WSMT)", available at http://www.wsmo.org/TR/d9/d9.1/v0.2/.
5. Riichiro Mizoguchi, 2004, "Tutorial on ontological engineering", New Generation Computing Journal, v.22 n.2, p.198-220.
6. T. Berners Lee, J Hendler, and O. Lassila, 2001, available online at http://www.sciam.com/2001/0501issue/0501berners.