A Semantic Web Based Ontology in the Financial Domain

S. Banerjee

Abstract—The paper describes design of an ontology in the financial domain for mutual funds. The design of this ontology consists of four steps, namely, specification, knowledge acquisition, implementation and semantic query. Specification includes a description of the taxonomy and different types mutual funds and their scope. Knowledge acquisition involves the information extraction from heterogeneous resources. Implementation describes the conceptualization and encoding of this data. Finally, semantic query permits complex queries to integrated data, mapping of these database entities to ontological concepts.

Keywords—Ontology, Semantic Web, Mutual Funds.

I. INTRODUCTION

DATA, information and knowledge management are key activities in modern finance. The fields of Economics and Finance are conceptually rich domains where information is complex, huge in volume and obtained from a variety of heterogeneous sources. Currently, most financial investment domain information produced by information providers is primarily textual, and , therefore it cannot be interpreted and processed by computers. This results in problems of Web content management. One approach to overcome this problem is the application of Semantic Web Technologies [1] . These technologies provide an explicit and formal representation of the semantics underlying information sources and by exploiting such formal semantics.

Ontologies have been proposed as a backbone technology for the Semantic Web and, more generally, for the management of formalized knowledge in the context of distributed system [2,3] They enrich content with machine-processable semantics, which can be communicated and processed by software programs. The main principle in this vision is to make information understandable for computers, thus enabling new, more powerful information processing and management capabilities, and reducing the involved costs.

Ontology is a vocabulary that is used for the description and presentation of a domain and also the meaning of the vocabulary in that domain. Definition of ontology can be obtained from several aspects like taxonomy [4], vocabulary [5], topic maps [6], conceptual model [7] and complex knowledge and logic theory [8]. In our proposed Ontology, we use an XBRL based ontology [9]. XBRL uses XML, XML Schema and Xlink.

S. Banerjee is with the Department of Natural Sciences, School of Management, West Bengal University of Technology, Sector I Salt Lake, Kolkata 700064. email: bsreep@sify.com

The design of our Ontology comprises of four steps, namely, specification, knowledge acquisition, implementation and semantic query. Specification involved a description of the taxonomy and different types of mutual funds available for various purposes. Knowledge acquisition consists of extracting data from the Internet web pages, several books, online libraries, free text as well as discussions with domain experts. Implementation involved data conceptualization and encoding. The relevant taxonomy, as well as different categories and their salient features are described using XML & RDF. Any relevant data and concepts available in other ontologies can be integrated as the project progresses. A query answering service is included. Attempts are in progress to recast the encoding in OWL.

The paper is organized as follows. Section II gives a brief review of some other ontologies. Our design is detailed in Section III & we conclude with a discussion of future work.

II. RELATED WORK

Earlier, the use of Ontologies was mainly in the domain of academia. Recently, ontologies are gaining acceptance in Biological Sciences as well as in Economics and Finance.

The Semantic Web vision" [10, 11] has prompted numerous efforts at the construction of ontologies in the Biological Sciences. Examples include the Fungal Web Ontology [12], Imagestore Ontology [13], Gene Ontology [14], Unified Medical Language System [15], Framework for Protein Classification [16] and Glycol Ontology developed at the University of Georgia, USA.

Since our focus is in the financial domain, a brief description of the work on ontologies in the financial domain is in order. Castells et. al. [17] have developed Semantic Web Technologies in Economic and Financial Information Management in the TIF domain. TIF or Tecnología, Información y Finanzas is part of a corporation that generates high-quality economic information (equity research notes, newsletters, analysis, sector reports, recommendations), and provides technology solutions for information consumers to access, manage, integrate and publish this information in web portals and company intranets. The platform includes an ontology-driven knowledge base, where information products are enriched with semantic descriptions. The platform provides means for content provision, access, and administration of this knowledge repository. The information access facilities include semantic-based search, exploration and visualization facilities. The advantages of the search, visualization, and management modules do not lie only in their application to the

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:7, No:6, 2013

particular case at hand. Besides improving the end-user experience, they provide important advantages for developers, as flexible, general-purpose modules, portable to other ontologies, easy to configure, supporting a variety of options and power vs. simplicity levels. The authors propose to extend their work to incorporate relations between different content classes.

Lavbic and Bajec [18] have employed Semantic Web in financial Instruments trading. They have applied the Rapid Ontology Developer approach to the development of the Financial Instruments and Trading Strategies (FITS) ontology. They hope to extend it to cover a wider number of possible use cases.

In [19], Zhang et. al. have described their efforts to develop an ontology supported web service for Enterprise Application Integration (EAI). A deep analysis of application pattern and characteristic of web services in EAI, and on the basis of the analysis, extends current semantic matching algorithm of web services to support service state was made. This paper introduces an ontology supported semi-dynamic semantic web service composition method. Their method combines the advantages of both static and dynamic web service composition, not only support complex business process, but also keep the flexibility of service matching and improves the maintainability. This is being extended into multi-ontology application systems.

An ontology for financial Investments has been developed in the Algeria Case study by Saleh and Mohamed [20]. They hope to extend their work to develop an expert system integrated with this ontology to provide financial intelligent assistance to various users, Shan [21] has detailed an ontology for demonstrating the domain knowledge about news in financial markets.

The ontology model comprises two components. One is represented using OWL DL (which is a sub-language of Web Ontology Language), which provides a hierarchical framework for the domain knowledge, including primary classes of news, classes of financial markets participants, classes of financial instruments, and relations between these classes. This component is a specification of domain-specific vocabulary terms. The other component is a causal map, used to demonstrate the impact of different classes of news events on financial instruments. It is of either a direct or an indirect "cause-effect" form, which can be written as rules using OWL rules language.

Furletti [22] describes a method for extracting new implicit knowledge directly from ontology by using an inductive/deductive approach. This methodology is applied in the context of MUSING, a European project in the field of business intelligence.

III. PROPOSED ONTOLOGY

Our proposed ontology describes three orthogonal aspects of Mutual Funds, namely, the taxonomy, structure and different functionalities of the different funds. The ontology is specified in four steps which are described as follows.

A. Specification

The purpose, scope and degree of granularity are crucial for the ontology development. The scope of this work extended to the establishment of an instantiated knowledge base describing the taxonomies, parts used and the functionalities of the medicinal plants and to make this semantically available to Bio-medical application and researches. Concepts correspond to classes which are interpreted as sets of objects. Roles correspond to relations that are interpreted as binary relations on objects[13]. We have include three types of DLs to define concepts and roles. This three types of DLs are i) Taxonomy DL, ii) Investment_typeDL, iii) Functionality DL. For example, the concept & roles of can be defined as follows:

For i) Taxonomy DL,
Mutual_Fund$\equiv$ Investment $\cap$ $\exists$ Investment_type
Mutual _Fund$\equiv$ Investment$\cap$ $\exists$ Risk, Lower Risk
Mutual_Fund $\equiv$ Investment $\cap$ $\exists$ Return, Higher Return

For ii) Investment_type DL,
Mutual_Fund $\equiv$ Investment_type$\cap$ $\exists$ structure.openended

For iii) Functionality DL,
Mutual_Fund$\equiv$ Investment_type$\cap$ $\exists$ equity.open-ended.high-return.

B. Knowledge Acquisition

This step involves extracting data from different resources. All this data are being collected from various resources and knowledge is being created using this data. Examples include web resources [9, 25].

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:7, No:6, 2013

```
<?xmlversion=1.0?>
<document>
   <taxonomy>
        <investment_type>mutual_fund</investment_type>
            <structure>open-ended or close-ended or
interval</structure>
                <risk>lower</risk>
                <return>high</return>
                    </investment_type>
                    <investment_type>
fixed_deposit</investment_type>
                <structure>close-ended</structure>
                <risk>lower</risk>
                <return>lower</return>
        </investment_type>
        <investment_type>Equity</investment_type>
            <structure>open-ended or closed-
ended</structure>
                <risk>high</risk>
                <return>high</return>
        </investment_type>
   </taxonomy>
</document>
```

Fig. 1 Example of Mutual Fund using XML schema

C. Implementation

Firstly, we have represented the taxonomy, structure, and different functionalities of the mutual funds using XML which represents the vocabulary for schemas for defining and documenting object classes. An example is given in Fig. 1.

Then we have represented it using RDF. As we know that RDF (Resource Description Framework) was the first language specified by the W3C for representing semantic information about arbitrary resources [23].It uses an RDF representation of VCARDS. RDF is the best thought of in the form of node and arcs diagrams [24]. The framework for Mutual Funds is given in Fig. 2.
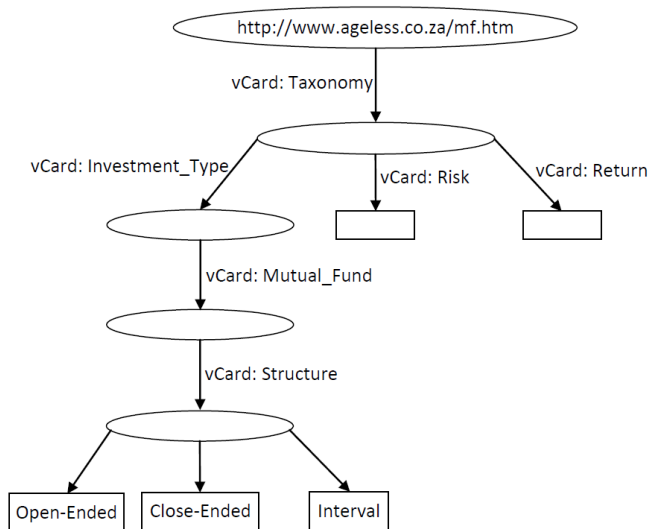

Fig. 2 An RDF model based framework for Mutual Fund

The figure describes the resources that we have taken into consideration and is depicted as ellipses .A property is represented by an arc, labeled with the name of a property. An internet resource is defined as any resource with a Uniform Resource Identifier (URI). This includes the Uniform Resource Locators (URL) that identify entire Web sites as well as specific Web pages. As with today's HTML META tags, the RDF description statements, encased as part of an Extensible Mark-up Language (XML) section, could be included within a Web page (that is, a Hypertext Mark-up Language - HTML - file) or could be in separate files [23].

In Jena [23], there are several methods for reading and writing RDF as XML form to the standard output stream model. The model is written in XML form to a file. It will take an OutputStream argument

The output of the model for mutual fund should look something like the Fig. 3.

Fig. 3 Output of RDF model for Mutual Funds

D. Semantic Query

The semantic query is expected to include many kinds of query-answering services with access to numerous types of information represented in widely different formats. To permit complex queries to integrated data, mapping of these database entities to ontological concepts is necessary [3]. In the domain of interest the different types of mutual funds (see Fig. 4) have different attributes like structure and functionality. For instance, a diversified equity fund (ELSS) (Fig. 4) can be open ended and have high risk and high return. We have designed the to semantic queries using OWL-QL.

OWL Query Language (OWL-QL) [26] is a formal language and protocol for a querying agent and an answering agent to use in conducting a query-answering dialog using knowledge represented in the Ontology Web Language (OWL) [27]. An OWL-QL query-answering dialog is initiated by a client sending a query to an OWL-QL server. An OWL-QL query is an object necessarily containing a query pattern that specifies a collection of OWL sentences in which some URI references are considered to be variables. For example, a client could ask "Which mutual fund is equity linked and open ended ?" with a query pattern shown in Fig. 5.

IV. SCOPE FOR FUTURE WORK

The mutual fund ontology project, which is in its infancy, aims to integrate mutual fund information using resources from

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:7, No:6, 2013

diverse sources of knowledge acquisition. Attempts will also be made to integrate the different aspects namely taxonomy, structure and functionality to improve the Semantic Query system by being able to access several fund schemes that have a common medical functionality (e.g. high return).

It is hoped that this financial ontology will be of immense use to computerize mutual fund related information
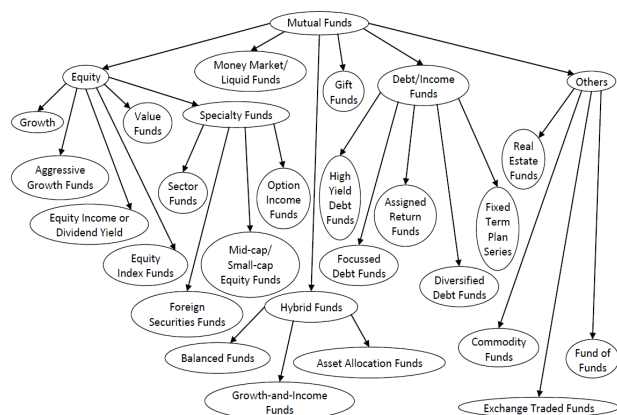

Fig. 4 Broad Mutual Fund Types

```
Query:("Which fund is  equity linked and  open-ended?")
Query pattern{is?p?c)(type(?c structure)
Must-bind variable list(?p)
May-bind variables list()
Answer KB pattern
Answer("ELSS is equity-linked and open-ended")
Answer pattern instance{(is open-ended  "ELSS")}
Query: ...
Server: ...
```

Fig. 5 A simple OWL-QL query and answer

REFERENCES

[1] Tim Berners-Lee, James Hendler and Ora Lassila, "The semantic web",scientific American, may 2001, p. 29-37.
[2] Thomas Gruber: Towards principles for the design of ontologies used for knowledge sharing, international international journal of human-computer studies, 43 (5-6) 907-928 (1993
[3] Thomas Gruber, Ontolingua: a mechanism to support portable ontologies, technical report, knowledge systems laboratory, Stanford University, 1991.
[4] Corcho, O., M. Fernandez-Lopez, and Gomez-Perez A.,Methodologies, tools and languages for building ontologies. Where is their meeting point? Data & Knowledge Engineering, 2003. 46(1): p. 41-64.
[5] Bechhofer, S. and Goble, C. Thesaurus construction through knowledge representation. Data & Knowledge Engineering, 2001. 37(1): p. 25-45.
[6] Park, J. and Hunting S, Xml topic maps: Creating and using topic maps for the web. 2002, Boston, USA: Addison Wesley. 640.
[7] Mylopoulos, J. Information modeling in the time of the revolution , Information Systems, 1998. 23(3-4): p.127-155.
[8] Waterson, A. and Preece A., Verifying ontological commitment in knowledge- based systems, Knowledge-based systems, 1999. 12(1-2): p. 45-54.
[9] Extensible Business Reporting Language, XBRL http://www.xbrl.org, 1998.
[10] D. Caragea, J. Pathak, j. Bao, A. Silvescu, C. Andorf, D. Dobbs and V. Honovar, Information integration and knowledge acquisition from

semantically heterogeneous biological data sources, B. Ludascher and L. Raschid (eds.) LNBI 3615, pp 175-190 (2005).
[11] D. Caragea, J. Pathak and V. Honavar, Learning classifiers from heterogeneous data, Proceed. Intl. Conf. Ontologies, Databases and Applications of Semantics for large scale information systems, 2004.
[12] A. Shaban-Nejad, C.J.O. Baker, G. Butler and V. Haarslev, The FungalWeb Ontology,, Genome Quebec, Canada (2005)
[13] C. Catton, S. Sparks and D.M. Sotton, The ImageStore Ontology and BioImage database:Semantic Web tools for Biological Research Images, ORIEL Project, IST-2001-32688.
[14] M. Ashburner, C. Ball, J. Blake, D. Botstein, J. Butler, J. Cherry, A. Davis, K. Dolinski, S. Dwight, J. Eppig, M. Harris, D.Hill, L.Issell-Tarver, A. Kasarkis, S. Lewis, J. Matese, J. Richardson, M.Ringwald, G. Rubin, and [14] R.Sherlock, Gene ontology:tool for unification in Biology, Nature Genetics, 25(1) 25-29, 2000.
[15] www.nlm.nih.gov/research/umls
[16] A. Kumar and B. Smith, A framework for protein classification,Proceed. German Conference on Bioinformatics, Vol. II, 55-57 (2003).
[17] Pablo Castells, Borja Foncillas, Rubén Lara, Mariano Rico1, Juan Luis Alonso, Semantic Web Technologies for Economic and Financial Information Management, The Semantic Web: Research and Applications, Lecture Notes In Computer Science 3053 p. 473-487 (2004).
[18] Dejan Lavbič and Marko Bajec, Employing Semantic Web technologies in financial instruments trading, International Journal on New Computer Architectures and Their Applications 2(1): 167-182 (2004)
[19] Kangkang Zhang, Ruzhi Xu, Yan Zhang, Ying Sai, Xiaobin Wang, An Ontology Supported Semantic Web Service composition Method in Enterprise Application Integration, International Multi-symposiums on Computer and Computational Sciences, 222-227 (2008).
[20] Maache Salah and Touahria Mohamed, Developing Ontology for Financial Investment "Algeria Case Study" International Journal of Computer Applications (0975 – 8887) 24(.4), p.1-6 (June 2011).
[21] Wang Shan Shan, Knowledge driven data mining for causal relationships between news and financial instruments, Ph.D. Thesis, City University of Hong Kong (2009)
[22] Barbara Furletti, Ontology-Driven Knowledge Discovery, Ph.D. Thesis, IMT Institute for Advanced Studies, Lucca Italy, (2009).
[23] www.w3.org/TR/1998/NOTE-XML-data/
[24] www.w3.org/TR/rdf-primer/
[25] www.sebi.gov.in/xbrl/XBRLMF.html
[26] Richard Fikes, Patrick Hayes, and Ian Horrocks, OWL-QL – A Language for Deductive Query Answering on the Semantic WebURI : http://ksl.stanford.edu/KSL_Abstracts/KSL-03-14.html
[27] www.w3.org/TR/owl-guide/