IJITKM Volume 6 • Number 2 • July-December 2013 pp. 119-127

# ONTOLOGIES FOR THE SEMANTIC WEB

S. Banerjee[1*], and Sima Ghosh[2]

***ABSTRACT:*** The Semantic Web is an extension of the Web, intended to provide information with a well defined meaning. In turn, the development of the Semantic Web depends on formal ontologies to obtain structured data for comprehensive and transportable machine understanding. Domain specific concepts and relationships between these concepts can be integrated in an ontology in a machine understandable manner. Thus, ontologies are useful in the Artificial Intelligence (AI) domain as these systems enable the aggregation of gargantuan proportions of data for intelligent or statistical analysis and subsequent decision making, using the knowledge acquired from such analysis. Construction of formal ontologies is thus a key ingredient for the success of the Semantic Web. Ontologies can be developed from several complementary disciplines to incorporate both structured and unstructured data from heterogeneous sources. These ontologies can then be integrated into a semi-automatic, co-operative framework within the Semantic Web. In this paper, we discuss some of popular Semantic Web Technologies that are used to develop these ontologies. We also survey some of the ontologies developed for two data rich disciplines, namely, Bio-medicine and Finance. Finally, we provide a description of two ontologies that we have developed in these domains.

***Keywords:*** Semantic Web, Ontology, Biology and Medicine, Finance, Medplant, Mutual Funds

## 1. INTRODUCTION

In the present age, computers are undergoing a rapid metamorphosis from single isolated devices to gateways to a worldwide network of information exchange. Thus, computers, play a key role in the exchange of data, information, and knowledge, some of the key tenets of information technology. Coupled to this is the enormous amount of accumulation of data on the Web, rendering the process of information retrieval very tedious and difficult. The idea of the Semantic Web was conceived by Tim Berners-Lee [1] as a panacea to this problem. The Semantic Web introduces a machine understandable layer of meta-data thus enabling computers to be able to manipulate contents automatically, without human intervention. The success of the Semantic Web relies heavily on the easy creation, integration and use of semantic data. This in turn depends on building an ontology. According to Wikipedia, an "Ontology is a formal representation of the knowledge by a set of concepts within a domain and the relationships between these concepts". The Semantic Web's success hinges largely on the proliferation of Ontologies [2]. The vision of Ontology learning should include a number of complementary disciplines within its ambit and integrate structured, unstructured and semi-structured data from heterogeneous sources to support semi-automatic, cooperative ontology engineering. In this paper, we survey several ontologies in two data-rich fields, namely, Bio-informatics and Finance and we also present two examples of ontologies in these fields that we have attempted to develop, based on some Semantic Web Technologies. The next section gives a brief description of some of the widely used Semantic Web technologies. This is followed by sections reporting surveys on Ontologies in Bio-Informatics and Finance. Two examples of Ontologies that we developed in Medical Informatics and Finance have been described in sections that follow. Finally, some concluding remarks are given.

[1,2] School of Engineering Technology, West Bengal University of Technology, BF-142, Salt Lake Sector I, Kolkata 700064, INDIA
*E-mail: bsreeparna1@rediffmail.com*
[*] *Corresponding Author: E-mail:bsreep@sify.com*

## 2. SEMANTIC WEB TECHNOLOGIES

The Semantic Web consists of media objects like Web pages, images or audio clips, as well as people, places and events. These resources are related through a variety of methods including the hyperlink for the conventional media objects like text and graphics. Different technologies associated with Semantic Web technologies include: XML, RDF [3] , Metadata Ontology; Metadata storage and database storage technologies; and Information and Knowledge Management. Here metadata refers to data about data.

Semantic Web technologies involves publishing in languages specifically designed for data. These include XML (eXtensible Markup Language), OWL (Ontology Web Language) [4] and RDF (Resource Development Framework). Whereas HTML (Hyper Text Markup Language) describes documents and links between them, RDF, OWL and XML can describe arbitrary things such as people, organizations, etc.,.

XML is a markup language defining rules for encoding documents in a format that can be read both by humans as well as machines. It is made up of user defined tags for elements, their attributes and the values of these attributes.

Document Type Definitions (DTD) and XML schemas specify the names of the elements and their attributes and their uses in documents. XML Schemas are successors of DTDs and provide a richer grammar for describing the structure of elements. RDF is a model for representing data about Web resources and is based on the subject, verb and object triple which embodies a statement. RDF Schema is atype system for RDF which provides a means to define domain specific properties and classes of resources. RDF can be extended using Darpa Agent Markup Language (DAML) along with Ontology Interchange Language (OIL) [5] to create a language for describing ontologies. RDF can be written in XML syntax. XML and RDF are two complementary technologies for developing an intelligent Web.

OWL is a standard for writing semantics. There are three sub-languages of OWL, namely, OWL Lite for users primarily needing a classification hierarchy and simple constraints.; OWL DL for maximum amount of expressiveness while retaining computational completeness and OWL Full for maximum expressiveness, syntactic freedom of RDF and no computational guarantee.

XBRL (eXtensible Business Reporting Language) [6] is an open technology standard for financial and economic information. It is essentially based on XML. XBRL incorporates the semantic meaning required for business reporting. Which technology to use i.e. whether to use XML, RDF/OWL or XBRL depends on the application at hand. Depending on the proposed ontology, a suitable Semantic Web technology can be selected.

Ontologies can be highlighted from several sources such as taxonomy which is knowledge with minimum hierarchical structure, vocabulary with words and synonyms, topic maps with the traversing through a large amount of data, conceptual models which embody complex knowledge and logic theory with very complex and consistent theory. Various applications of ontology include natural language processing, knowledge management, digital libraries, information extraction, intelligent search engines, business process modeling, etc.,.

## 3. ONTOLOGIES IN THE BIO-MEDICAL DOMAIN

Bio informatics and Medical Informatics are data rich sciences because of the diverse nature of information that they provide. However, most of this data is generated in contextual form such as natural language and images, which are human readable but not accessible to computer analysis. Such heterogeneous and incompatible databases must be bridged and normalized to enable the content to be more accessible and computer comprehensible. The role of ontology is to impart semantic clarity to the information as well as render it accessible to the computer.

Since biological data sources are developed by different persons and groups, their ontological presentations differ. Differences include assumptions concerning objects that exist in the world, properties or attributes of objects, relationships between objects, the possible values of the attributes and their intended meaning, as well as the granularity or level of abstraction at which objects and properties can be described [7]. Thus semantic differences are an inevitable feature in designing a database. The panacea lies in solving a data integration problem. Solving this information integration problem forms the basis of the Semantic Web.

This "Semantic Web vision" [8] has prompted numerous efforts at the construction of ontologies in the Biological Sciences. Examples include the Fungal Web Ontology [9], Imagestore Ontology [10], Gene Ontology [11] , Unified Medical Language System [12], Framework for Protein Classification [13] and Glyco Ontology developed at the University of Georgia, USA. Some of these ontologies will be discussed briefly. Data sources created for use in one context can be used interchangeably, especially in collaborative scientific discovery applications involving data-driven construction of classifiers from semantically disparate sources [14].

In this section we briefly review three representative ontologies. Image Store ontology provides an example of the usefulness of an Ontology for image databases in the domain of biology.

### 3.1. Fungal Web

The Fungal Web project[15] is affiliated with the Fungal Genomics project at Concordia which is a large scale, gene discovery program on evolutionarily diverse fungal species. This project plans to discover over 70,000 new fungal genes and develop high throughput methods to characterize the function of the enzymes produced by these genes. Integration of the tools developed within the Fungal Web project to the Fungal Genomic bioinformatics is a common goal of the two projects.

The project is comprehensive in that it is vertically integrated[16]. In essence it is several major subprojects each carried out on several fungal species. The bioinformatics platform supports data collection and analysis across the entire project. This project is called Fungal Web. The vision of the semantic Web is to extend the World Wide Web from a collection of data and documents that are often hard to find and use into a collection of knowledge that is very convenient to use. In this project the range of tasks in genomics and the kinds of intelligent reasoning required for those tasks in order to gain a better understanding of the knowledge that must be captured in formally described ontologies were explored. The Fungal Web Ontology, written in OWL, covers enzyme classification, enzyme functional parameters, enzyme applications, and fungal taxonomy gathered from distributed resources. Ongoing research involves improvement of querying capability through the provision of an integrated global view and schema of all concepts/ instances, tools for user formulated querying using ontologically defined terms, system reformulation of queries to produce sub-queries for component resources[3].

### 3.2. Image Store Ontology

The BioImage Database[10] is an ontology-driven image database for multidimensional images of biological specimens. It is built using Jena and other Open Source components around an Image Store Ontology written in OWL-DL that describes all aspects of the image. Manual metadata entry is simplified with the help of the dynamic creation from the underlying ontology of simple Web form user submission interfaces . When metadata already exist in digital form, automated metadata entry is also possible. Accurate relevant images can be retrieved with an advanced search interface. During this retrieval process, web service interactions with third-party services can permit the textual descriptions of the image to be marked up on the fly with definitions of key terms, including the disambiguation of gene names. Image data is integrated with the bioinformatics databases by the mechanism offered by the semantic web technologies, providing a mechanism for helping researchers navigate through what is becoming an intricate and confusing information landscape. The BioImage Database provides a tool to facilitate this linking. A second potentially useful outcome of this project has been the development of a mechanism for creating ontology-driven html input forms for the collection of instance data.

### 3.3. Gene Ontology

Gene Ontology Consortium [17] which is defining ontologies for describing Cellular Components, Biological Process, and Molecular Function. The cellular component ontology[18] describes locations, at the levels of sub-cellular structures and macromolecular complexes. The Biological Process ontology includes standard definitions and term relationships in the biological process. A biological process is series of events accomplished by one or more ordered assemblies of molecular functions. Molecular function ontology describes standard definitions and term relationships in the molecular function ontology .It defines the activities, such as catalytic or binding activities, that occur at the molecular level.

There are three separate aspects to this effort: first, the development and maintenance of the ontologies themselves; second, the annotation of gene products, which entails making associations between the ontologies and the genes and gene products in the collaborating databases; and third, development of tools that facilitate the creation, maintenance and use of ontologies.

Thus Gene Ontology (GO)describes gene products from three orthogonal perspective: molecular function, biological process and cellular component. These vocabularies and their relationships are represented in the form of directed acyclic graphs (DAGs). Thus, the taxonomy in the GO may be referred to as a network in which each term or concept may represent a "child node" of one or more "parent nodes". There are two types of child-to-parent relationships: "is a" and "part of". The first type is defined when a child is an instance of a parent. The second type is used to describe the situation where a child node is a component of a parent.

## 4. ONTOLOGIES IN THE FINANCIAL DOMAIN

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. Hence, a brief description of the work on ontologies in the financial domain is in order.

### 4.1. TIF

Castells et. al. [19] 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 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.

### 4.2. FITS

Lavbic and Bajec [20] have employed Semantic Web in financial Instruments trading. They have applied the Rapid

Ontology Developer (ROD) approach to the development of the Financial Instruments and Trading Strategies (FITS) ontology. Their goal was to develop ontology by constructing schematic ontology including axiomatic information to fully support trading by employing reasoning. Their main objective was to combine dynamic (Web) sources with minimal input from the user. ROD involves capturing concepts, mutual relations and deriving expressions between concepts and relationships and incorporating them into a schematic model which in turn binds with existing instances of that vocabulary. These include data from relational databases, text files, other ontologies etc.,. After ROD, they use the Financial Instruments (FI) ontology which includes basic concepts, stock exchange market and trading day and analysis. The technology used is OWL-DL. They hope to extend it to cover a wider number of possible use cases.

### 4.3. EAI

In [21], 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.

### 4.4. Algeria Case Study

An ontology for financial Investments has been developed in the Algeria Case study by Saleh and Mohamed [22]. This work focusses on the collection, organization, representation and formalization of knowledge in finance, specially in the domain of financial investment. Their knowledge sources include banks and other financial institutions in Algeria. The technology they have used Protege 3.4.4. version of OWL. They hope to extend their work to develop an expert system integrated with this ontology to provide financial intelligent assistance to various users,.

### 4.5. Financial Market Ontology

Shan [23] 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.

### 4.6. MUSING

Furletti [24] describes a method for extracting new implicit knowledge directly from an ontology by using an inductive/ deductive approach. This methodology is applied in the context of MUSING (Multi-industry Semantic based next generation business INtelliGence), an European project in the field of business intelligence. In this approach, the Ontology schema (also referred to as T-box) is the formalism for Knowledge Representation (KR). The deductive process draws inference from the ontology structure which includes the concepts and properties, by using graph theory and applying link analysis techniques and producing rules schemas with only the important concepts involved. The inductive process utilizes the ontology instances (i.e. data, also referred to as A-Box) for enriching implications (i.e. rules schemas) and building the final rules. Their proposed method gives a Bayesian interpretation to the relationship to the relationships that already exist in an ontology.

MUSING aims at developing new generation Business Intelligence Tools by integrating the Semantic Web and Human Language Technologies. Furletti [22] tests the methodology of extracting new implicit knowledge using the above mentioned inductive/deductive approach on MUSING. For handling ontologies the Protege and Jena APIs (the latest stable release of OWL 1.0) have been used and PATTERNIST, developed at their lab, has been used as a pattern discovery algorithm in this project.

## 5. MEDPLANT: A MEDICINAL PLANT ONTOLOGY

Our proposed ontology MedPlant [36] describes three orthogonal aspects of medicinal plants, namely, the botanical taxonomy, structural parts of the plant and different medicinal uses of the plant. The ontology is specified in four steps which are described as follows .

### 5.1. 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[25]. We have include three types of DLs to define concepts and roles. This three types of DLs are (*i*) Taxonomy DL, (*ii*) Plant DL, (*iii*) Usage DL. For example, the concept & roles of Neem can be defined as follows:

For (*i*) Taxonomy DL,

$\text{Neem} \equiv \text{Plant} \cap \exists\ \text{family.Meliaceae}$

$\text{Neem} \equiv \text{Plant} \cap \exists\ \text{genus.Azadirachta}$

$\text{Neem} \equiv \text{Plant} \cap \exists\ \text{species.Indica}$

For (*ii*) Plant DL,

$\text{Neem} \equiv \text{Plant} \cap \exists\ \text{parts\_used.Leaves}$

For (*iii*) Usage DL,

$\text{Neem} \equiv \text{Plant} \cap \exists\ \text{prevents.Intestinal\_worms}$

## 5.2. Knowledge Acquisition

This step involves extracting data from web resources[26] & some books [27-29]. We have collected the information regarding medicinal plants from these various sources These web resources includes some information about medicinal herbs[30], as well as some data regarding medicinal trees[31] which are being used since ancient times[32].

All this data are being collected from various resources & knowledge is being created using this data.

## 5.3. Implementation

Firstly, we have represented the taxonomy, structural used parts, different functionalities of the plants using XML which represents the vocabulary for schemas [32] 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[33].It uses an RDF representation of VCARDS. RDF is the best thought of in the form of node and arcs diagrams. The framework for Neem is given in Fig. 2. 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 Markup Language (XML) section, could be included within a Web page (that is, a Hypertext Markup Language - HTML - file) or could be in separate files.

```
<?xml version="1.0" ?>
<document>
<medicinal_plant>
        <taxonomy>
                <general_name>neem</general_name>
                <family>meliaceae</family>
                <botanical_name>
                        <genus>azadirachta</genus>
                        <species>indica</species>
                </botanical_name>
        </taxonomy>
        <structure>
                <bark>
                        <functionalities>
                                <dental>prevent tooth problems</dental>
                                <digestive>prevent nausea and vomiting</digestive>
                        </functionalities>
                </bark>
                <leaves>
                        <functionalities>
                                <blood>prevent malaria</blood >
                                <lungs>prevent tuberculosis</lungs>
                                <muscular_joints>prevent arthritis</muscular_joints>>
                                <intestinal>prevent intestinal worms</intestinal>
                                <dermal>prevent skin disease</dermal>
                        </functionalities>
                </leaves>
                <berries>
                        <functionalities>
                                <mascular_joints>prevent painful joints and muscles</mascular_joints>
                                <scalp>prevent dandruff</scalp>
                                <antiseptic>antiseptic</antiseptic >
                        </functionalities>
                </berries>
                <flowers>
                        <functionalities>
                                <dermal>prevent skin problems</dermal>
                                <gastric>prevent ulcers</gastric>
                        </functionalities>
                </flowers>
        </structure>
</medicinal_plant>
</document>
```

**Figure 1: Example of Neem Represented Using XML**

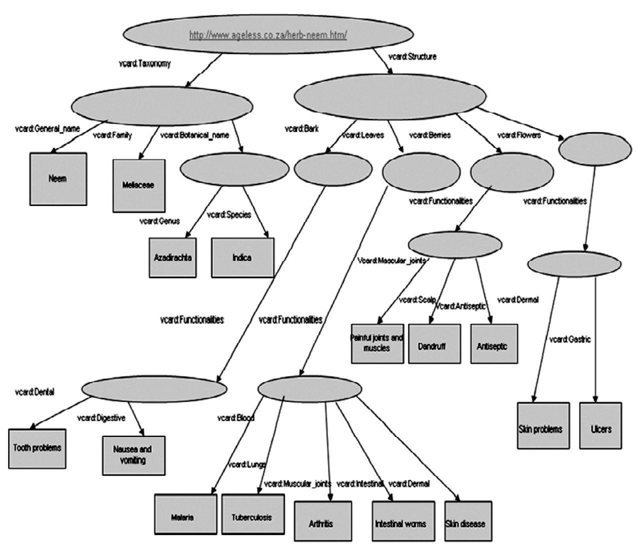


**Figure 2: An RDF Model for Neem (Azadirachta Indica)**

In Jena[34], 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 neem should look something like the Fig. 3.

```
<rdf:RDF
 xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
 xmlns:vcard='http://www.w3.org/2001/vcard-rdf/3.0#'
>
  <rdf:Description rdf:about='http://www.ageless.co.za/herb-neem.htm'>
    <vcard:Taxonomy rdf:nodeID="A1"/>
    <vcard:Stucture rdf:nodeID="A2"/>
  </rdf:Description>
  <rdf:Description rdf:nodeID="A1">
    <vcard:General_name>Neem</vcard:General_name>
    <vcard:Family>Meliaceae</vcard:Family>
    <vcard:Botanical_name rdf:nodeID="A1.1"/>
  </rdf:Description>
  <rdf:Description rdf:nodeID="A1.1">
     <vcard:Genus>Azadirachta</vcard:Genus>
      <vcard:Species>Indica</vcard:Species>
 </rdf:Description>
 <rdf:Description rdf:nodeID="A2">
    <vcard:Bark rdf:nodeID="A2.1"/>
    <vcard:Leaves rdf:nodeID="A2.2"/>
    <vcard:Berries rdf:nodeID="A2.3"/>
    <vcard:Flowers rdf:nodeID="A2.4"/>
 </rdf:Description>
 <rdf:Description rdf:nodeID="A2.1">
  <vcard:Functionalities rdf:nodeID="A2.1.1"/>
 </rdf:Description>
 <rdf:Description rdf:nodeID="A2.1.1">
   <vcard:Dental>Tooth problems</vcard:Dental>
   <vcard:Digestive>Nausea and vomiting</vcard:Digestive>
  </rdf:Description>
  <rdf:Description rdf:nodeID="A2.2">
   <vcard:Functionalities rdf:nodeID="A2.2.1"/>
 </rdf:Description>
 <rdf:Description rdf:nodeID="A2.2.1">
   <vcard:Blood>Malaria</vcard:Blood>
   <vcard:Lungs>Tuberculosis</vcard:Lungs>
   <vcard:Muscular_joints>Arthritis</vcard:Muscular_joints>
   <vcard:Intestinal>Intestinal worms</vcard:Intestinal>
   <vcard:Dermal>Skin disease</vcard:Dermal>
  </rdf:Description>
  <rdf:Description rdf:nodeID="A2.3">
   <vcard:Functionalities rdf:nodeID="A2.3.1"/>
 </rdf:Description>
 <rdf:Description rdf:nodeID="A2.3.1">
   <vcard:Muscular_joints>Painful joints and muscles</vcard:Muscular_joints>
   <vcard:Scalp>Dandruff</vcard:Scalp>
   <vcard:Anteseptic>Antiseptic</vcard:Antiseptic>
  </rdf:Description>
  <rdf:Description rdf:nodeID="A2.4">
   <vcard:Functionalities rdf:nodeID="A2.4.1"/>
 </rdf:Description>
 <rdf:Description rdf:nodeID="A2.4.1">
   <vcard:Gastric>Painful joints and muscles</vcard:Gastric>
   <vcard:Dermal>Dandruff</vcard:Dermal>
 </rdf:Description>
</rdf:RDF>
```

**Figure 3: Output of the RDF model for Neem (Azadirachta Indica)**

## 5.4. 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 biomedical data sources these entities are medicinal plants' name, taxonomy, parts used and functionalities of these plants. We have designed the to semantic queries using OWL-QL. OWL Query Language (OWL-QL)[35] is a formal language and protocol for a querying agent and an answering agent to use in conducting a query-answering dialogue using knowledge represented in the Ontology Web Language (OWL) [4]. An OWL-QL query-answering dialogue 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 plant prevents gastric ailments ?" with a query having the query pattern shown in Figure 4.

```
Query: ("Which plant prevents gastric ailments?"
Query Pattern: {(prevents ?p ?c) (type ?c Ailments) (regarding ?c Gastric)}
Must-Bind Variables List: (?p)
May-Bind Variables List: ()
Don't-Bind Variables List: ()
Answer Pattern: {(prevents ?p "gastric ailments")}
Answer KB Pattern: ...
Answer: ("Neem prevents gastric ailments?")
Answer Pattern Instance: {(prevents Neem "gastric ailments")}
Query: ...
Server: ...
```

**Figure 4: A Simple OWL-QL Query and Answer**

## 5.5. Scope for Future Work

The MedPlant ontology project, which is in its infancy, aims to integrate medicinal plants using resources from diverse sources of knowledge acquisition. Attempts will also be made to integrate the different aspects namely taxonomy, plant structure and medicinal uses to improve the Semantic Query system by being able to access several plants that have a common medical functionality (e.g. cure digestive ailments).

It is hoped that this MedPlant ontology will be of immense use to practitioners of Alternative Medicine.

# 6. ONTOLOGY IN THE FINANCIAL DOMAIN

Our proposed ontology [37] 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.

## 6.1. Specification

The scope of this work extended to the establishment of an instantiated knowledge base describing the taxonomies,types used and the functionalities of thedifferent types of mutual funds to make this semantically available to stock market and mutual fund investment. 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,

$\text{Mutual\_Fund} \equiv \text{Investment} \cap \exists\ \text{Investment\_type}$

$\text{Mutual \_Fund} \equiv \text{Investment} \cap \exists\ \text{Risk, Lower Risk}$

$\text{Mutual\_Fund} \equiv \text{Investment} \cap \exists\ \text{Return, Higher Return}$

For (*ii*) Investment_type DL,

$\text{Mutual\_Fund Investment\_type} \cap \exists\ \text{structure. openended}$

For (*iii*) Functionality DL,

$\text{Mutual\_Fund} \equiv \text{Investment\_type} \cap \exists\ \text{equity.open-ended.high-return.}$

## 6.2. 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, stock markets, mutual fund brochures, etc.,.

## 6.3. 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. 5.

```
<?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>
```

**Figure 5: Example of Mutual Fund using XML Schema**

Then we have represented it using RDF. The framework for Mutual Funds is given in Figure 6.

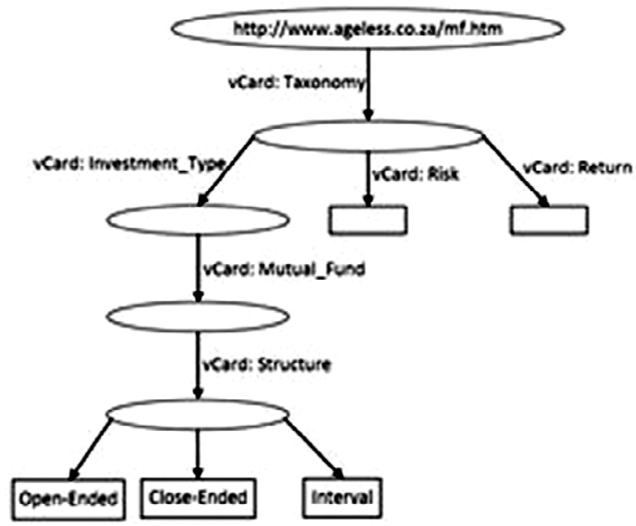


**Figure 6: An RDF Model for Mutual Fund**

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 Markup Language (XML) section, could be included within a Web page (that is, a Hypertext Markup Language - HTML - file) or could be in separate files[32].

The output of the model for mutual fund should look something like the Figure 7.

```
<rdf:RDF
xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
xmlns:vcard=http://www.w3.org/2001/vcard-rdf3.0#>
<rdf:Description rdf:about='http//www.ageless.co.za/mf.htm'>
<vcard:Taxonomy rdf:nodeID="A1"/>
</rdf:Description>
<rdf:Description rdf:node ID="A1">
<vcard:Investment_type>Mutual_fund</vcard:Investment_type>
<vcard:Risk>Lower_Risk</vard:Risk>
<vcard:Return>Higher_Return</vcard:Return>
</rdf:Description>
<rdf:Description>
<vcard:Structure rdf:nodeID="A1.2">
<vcard:Open-ended rdf:nodeID="A1.2.1">
<vcard:Close-ended rdf:nodeID="A1.2.2">
<vcard:Interval rdf:nodeID="A1.2.3">
</rdf:Description>
```

**Figure 7: Output of RDF Model for Mutual Funds**

### 6.4. Semantic Query

Semantic queries 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 Figure 4) have different attributes like structure and functionality. For instance, a diversified equity fund (ELSS) (Figure 8) can be open ended and have high risk and high return We have designed the to semantic queries using OWL-QL. For an example in this financial domain, a client could ask "Which mutual fund is equity linked and open ended ?" with a query pattern shown in Figure 9.

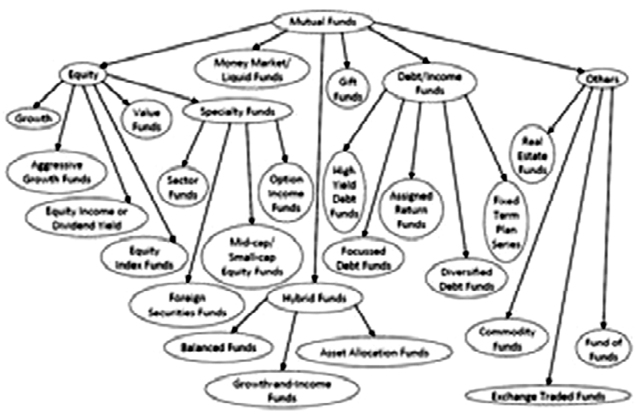


**Figure 8: Broad Mutual Fund Types**

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: ...

**Figure 9: A Simple OWL-QL Query and Answer**

### 6.5. Scope for Future Work

The mutual fund ontology project,like MEDPLANT, aims to integrate mutual fund information using resources from 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 functionality (e.g. high return).

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

## 7. CONCLUDING REMARKS

The Semantic Web, perceived to be the successor of the Web, holds great promise in providing structured information in a machine understandable fashion, along with a set of inference rules that will enable computers to perform automated reasoning. The success of the semantic Web hinges design of formal ontologies, which provide domain dependent concepts along with the relationships between the concepts. Intelligent decision support systems (IDSS) as well as intelligent and statistical data analysis relies on Artificial Intelligence based rules for their efficient functioning. It is hoped that the design of formal ontologies will greatly aid the intelligent data analysis and decision support processes, in the future. Several ontologies in two data rich domains like Biology, Medicine and Finance have been surveyed in this paper and two ontologies that we have developed have also been presented. However, these two ontologies are not exhaustive and more information can be input to develop a more comprehensive and exhaustive system. Furthermore, there also exists a scope for developing many more ontologies in these data rich domains.

## REFERENCES

[1] Berners-Lee T., Hendler J. and Lassila O., *The Semantic Web, Scientific American*, May, pp. 29-37 (2001).

[2] Maedche, A. & Staab, S. Ontology Learning for the Semantic Web, *IEEE Intel. Syst.*, **6**(2), pp. 72-79. (2001).

[3] Decker, S. The Semantic Web Roles of XML and RDF. *IEEE Internet Comp.*, 4(5), pp. 63-74., (2000).

[4] Smith, M. K; Welty, C. and McGuinness, D. L. OWL Web Ontology Language Guide, W3C, 7-15. (2008).

[5] Horrocks, I. DAML+ OIL: A Reasonable Web Ontology Language. In Lecture Notes in Computer Science (LNCS), Springer-Verlag, 2287. pp. 2-13, (2002).

[6] Donnelley R. R. XBRL Reference Guide (2006).

[7] Gruber T, Ontolingua: A Mechanism to Support Portable Ontologies.

[8] Caragea D Pathak J, Bao J,. Silvescu A, Andorf C, Dobbs D 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).

[9] Shaban-Nejad A, Baker C. J. O., Butler G and Haarslev V, The FungalWeb Ontology,, Genome Quebec, Canada (2005).

[10] Catton C, Sparks C and Sotton D M, The Image Store Ontology and BioImage Database: Semantic Web tools for Biological Research Images, ORIEL Project, IST-2001-32688.

[11] Ashburner M, Ball C, Blake J, Botstein D., Butler J, Cherry J, Davis A, Dolinski K, Dwight S, . Eppig J, Harris M,Hill D, Issell-Tarver L, Kasarkis A, Lewis S, Matese, Richardson J, .Ringwald M, Rubin G, and. Sherlock R, Gene Ontology: Tool for Unification in Biology, Nature Genetics, **25**(1) pp. 25-29, 2000.

[12] www.nlm.nih.gov/research/umls

[13] Kumar A. and Smith B., A Framework for Protein Classification, *Proceed. German Conference on Bioinformatics*, **II,** pp. 55-57 (2003).

[14] Caragea D, Pathak J and Honavar V, Learning Classifiers from Heterogeneous Data, *Proceed. Intl. Conf. Ontologies, Databases and Applications of Semantics for Large Scale Information Systems*, (2004).

[15] Baker C.J.O, Butler G, Haarslev V, Ontologies, Semantic Web and Intelligent Systems for Genomics. URI : http://users.encs.concordia.ca/~haarslev/publications/SWIG2004-FungalWeb.pdf (2004).

[16] Butler G, Workflow Scenarios for a Semantic Web for Fungal Genomics. URI: http://www.nettab.org/2005/docs/NETTAB2005_ButlerOral.pdf (2005).

[17] Shun-Liang Cao, Lei Qin Wei-Zhong He, Yang Zhong, Yang Yong Zhu, and Yi-Xue Li, Semantic Search Among Heterogeneous Biological Databases Based on Gene Ontology, Acta Biochimica et Biophysica Sinica, **36**(5): pp. 365-370, (2004).

[18] www.geneontology.org/

[19] Castells P, Foncillas B, Lara R, Rico M, Alonso L J, Semantic Web Technologies for Economic and Financial Information Management, *The Semantic Web: Research and Applications, Lecture Notes In Computer Science* 3053 pp. 473-487 (2004).

[20] Lavbi D and Bajec M, Employing Semantic Web Technologies in Financial Instruments Trading, *International Journal on New Computer Architectures and Their Applications* **2**(1): pp. 167-182 (2004).

[21] 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*, pp. 222-227 (2008).

[22] Maache Salah and Touahria Mohamed, Developing Ontology for Financial Investment "Algeria Case Study", *International Journal of Computer Applications* (0975-8887) **24**(4), pp. 1-6 (June 2011).

[23] Wang Shan Shan, Knowledge Driven Data Mining for Causal Relationships Between news and Financial Instruments, Ph.D. Thesis, City University of Hong Kong (2009).

[24] Barbara Furletti, Ontology-Driven Knowledge Discovery, Ph.D. Thesis, IMT Institute for Advanced Studies, Lucca Italy, (2009).

[25] Horrock I, Sattler U, Description Logics - Basics, Applications, and More. http://www.cs.man.ac.uk/~horrocks/Slides/IJCAR-tutorial/Print/p1-introduction.pdf

[25] Natural Elixir - List of Medicinal Plants for What Ails You. URI: http://www.naturalelixir.com/

[26] Pranbigyan Katha (Life Science), Hazra P. G., Jaydurga Library, 1st Ed., 2004.

[27] Madhyamik Jibbidya (Biology), Chakraborty A, Chaya Prakashani, 9th Ed., 2000.

[28] Pran Bigyan (Life Science), Pal R. N, Bhattacharaya B , Allied Publishing Syndicate, New Ed., 2003.

[29] Ageless - Herbal Encyclopedia. URI: http://www.ageless.co.za/

[30] Orissa Forest Development Corporation - List of Important Medicinal Plants & their uses. URI:

http://www.orissafdc.com/

[31] Belinda Hawkins, Plants for life: Medicina plant conservation and botanic gardens. URI: http://www.bgci.org/files/Worldwide/Publication

[32] www.w3.org/TR/1998/NOTE-XML-data/

[33] www.w3.org/TR/rdf-primer/

[34] Brian McBride, An Introduction to RDF and the Jena RDF API ,v 1.23, 2007 URI : http://jena.sourceforge.net/tutorial/RDF_API

[35] 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

[36] MEDPLANT: A Medicinal Plant Ontology: S. Banerjee and Sima Ghosh, Proceed. International Conference on Web Intelligent Systems (ICWIS 2009), Chennai, India (2009).

[37] A Semantic Web Based Ontology in the Financial Domain, S. Banerjee, accepted, International Conference on Innovation and Information Management, (ICIIM 13) Paris, France (2013).