Ontology Driven Framework to Resolve Information Interoperability in Multidisciplinary Applications and Semantic Information Retrieval

Suresh Babu Golla*, Dr. Sunitha Abburu**
*(Research scholar, Computer Science, Bharathiar University, Coimbatore Email: suresh_babu_golla@yahoo.com)
** (Principal Investigator, Department Of Science & Technology, Delhi Email: drsunithaabburu@yahoo.com)

ABSTRACT
Information interoperability and semantic information retrieval are the major challenges in the era of globalization. Interoperability enables data sharing, information exchange and coordinate actions among multidisciplinary applications. Ontology is a prevalent semantic technology that provides support to address the information interoperability issues. The objective of the current research is twofold: 1) address information interoperability in multidisciplinary applications and data integration 2) ontology concept based semantic information retrieval on the integrated data. The current work provides a generic framework that facilitates information interoperability by addressing semantic and syntactic heterogeneity in multidisciplinary applications data. The framework integrates and exchange data from multiple domains. The performance and efficiency of the framework are evaluated through a real time and multidisciplinary application to test the effect of information interoperability on semantic information retrieval. A query set that covers multiple domains is prepared to test the performance of ontology concept based semantic information retrieval. The evaluation test uses popular relevancy measures that are precision, recall, and F-measure.

Keywords - Ontology, RDF, Information Interoperability, Multidisciplinary Applications, Semantic Information Retrieval

I. INTRODUCTION
Multidisciplinary applications such as disaster management, public health, smart city, tourism etc. require data from diverse sources such as satellite and sensor observations, GIS, crowdsourcer information, social media etc. During and after natural disasters, the world wide web becomes a better platform for government, non-government organizations, volunteers and disaster management communities to post and exchange disaster information. The disaster information such as disaster incident, aids required and available are uploaded into social and disaster management websites. Though the data is available, disaster management communities and applications could not utilize the data due to syntactic, schematic and semantic heterogeneity [1]. The heterogeneity caused mainly due to different disaster management stakeholders uses different modes such as mobile applications, web interfaces that produce diverse information structure and vocabulary. Satellite and sensor observations are providing more useful information such as real-time weather data that plays a crucial role in many multidisciplinary applications. Meteorology domain contains information of various weather parameters such as air temperature, surface temperature, solar radiation, and precipitation etc. Meteorology satellite observation systems consist various sensors and configurations that stores data in heterogeneous file formats (e.g. NetCDF, HDF5, CSV and GRIB etc.) and vocabulary [2]. The capabilities of the existing approaches are not effective in knowledge sharing and distributed problem-solving because of heterogeneous data and lack of a common understanding across disciplines [3]. For example, to understand the impact of climate on the disaster management and effective decision making, there is a need of integrating data from disaster management and meteorology disciplines. However, as discussed, the data from these sources are heterogeneous in:

- File format
- Syntax
- Schema
- Semantics

The heterogeneity in the data limits:
- Coordination, communication of information across the domains
- Provide poor performance for the semantic queries such as “fire disasters during sunny weather state”

The other major issue with multidisciplinary applications is querying cross-linked and integrated data. Many of the current information systems are following data driven
approaches that facilitates answers to the syntactic queries. The data driven approach queries are structural and not effective to access inter linked and integrated data from multidisciplinary application domains. Semantic and conceptual query systems facilitate improved query performance.

To address heterogeneity in the data and achieve semantic information retrieval, there is a strong need of powerful framework that achieves information interoperability and builds multidisciplinary application domain knowledge base. In computer science, information interoperability is the ability to meaningfully exchange information among various applications [4]. Information interoperability increases the quality of the information and enables information exchange, share, reusability, and utilization among multiple domains [5]. Information interoperability among the multidisciplinary application domains can be achieved through bringing data into:

- Common structure
- Common vocabulary
- Machine processable format

Ontology is a formal, explicit specification of a shared conceptualization [6]. Ontology is a highly recommended semantic technology and significant solution to understand heterogeneous data, improve interpersonal communication and interoperability of information systems [7]. Figure 1. shows the role of interoperability in multidisciplinary applications. Ontology plays a vital role to:

- Address heterogeneity in the data
- Achieve information interoperability
- Integrating data from multidisciplinary domains
- Establish semantic relations between instances of same and multiple domains
- Improves query performance through semantic information retrieval

Figure 1. Information interoperability and multidisciplinary applications

The current research work focuses towards design, implementation and test a generic ontology driven framework that improves semantic information retrieval through achieving information interoperability in multidisciplinary application domains. The framework incorporates semantic standards: OWL, RDF, SPARQL and several techniques: ontology mapping, transformation and knowledge management. It builds a knowledge base through mapping multidisciplinary application domain data into Resource Description Framework (RDF) [8], which facilitates information interoperability and semantic retrieval.

Rest of the paper is organized as follows. Section 2 discuss the recent ontology driven approaches on interoperability and semantic information retrieval. Section 3 describes an ontology driven framework, section 4 illustrates implementation and results, section 5 describes evaluation results and section 6 concludes.

II. BACKGROUND WORK

Information heterogeneity and semantic information retrieval are the major concerns in the era of information globalization. Ontology is a powerful semantic technology that provides means to solve the interoperability issues and semantic query. This section discusses the current state of the art on ontology driven interoperability and semantic information retrieval methods in various application domains.
María et al., [9] proposed a semantic web based framework to achieve semantic interoperability in clinical data, archetypes, and terminologies and to manage archetypes and EHR data from different standards. María et al., [10] presented an ontology based approach for transformation and integration of biomedical data sources through addressing heterogeneity. Stefan et al., [11] proposed an ontology based solution to attain semantic interoperability in healthcare through semantic annotations to the clinical information entities. Rodrigo et al., [12] discussed an approach to address interoperability issues in the OntoAgroHidro engineering using ontology that consists knowledge impacts of agricultural activities and climatic changes on water resources. Alfredo [13] discussed semantic based software architecture to merge semantic web and Internet of Things (IoT) through address information interoperability issues.

Malik et al., [14] proposed an approach to address key data interoperability issues in the aerospace domain. Igor et al., [15] discussed a document management system that gives knowledge base describing document metadata using ontologies. The detailed study concludes that majority of the works are very specific to a domain, data formats. This raises demand for an effective and generic ontology driven approach for information interoperability that supports multidisciplinary applications and effective semantic retrieval. The current research provides an ontology driven framework that provides solutions to address interoperability issues in multidisciplinary applications domains and semantic information retrieval.

III. ONTOLOGY DRIVEN FRAMEWORK

The current research provides a generic, flexible and easily extendable framework to address information interoperability, information integration and semantic information retrieval across multidisciplinary applications. To illustrate the framework, the research considers data from two diverse and most popular application domains that are disaster management and meteorology.

Figure 2. Information interoperability and semantic information retrieval framework

www.ijaert.org
To achieve interoperability in these domains, the framework takes ontologies that provide description of concepts, properties, axioms of the domains. Figure 2 shows conceptual workflow of the framework. Integration of weather and disaster information place a vital role for effective decision making and provides better disaster management services such as rescue, relief. Further, the integrated information can also be used to derive hidden insights and study the impact of climate and environment change on natural disasters.

The two major components of the framework are:
- Information interoperability using domain ontologies
- Semantic information retrieval from multidisciplinary domains

A. Information Interoperability using Domain Ontologies

Ontologies provide domain knowledge through a detailed description of domain concepts, concept hierarchies, the relationship between concepts and axioms. Figure 3 shows concepts of disaster domain ontology. Figure 4 shows description of a weather concept. Domain ontologies can be used to achieve information interoperability addressing syntactic, schematic and semantic heterogeneity in the information of multidisciplinary domains. The current research achieves interoperability through representing data into an ontological model. That represents data as instances of the domain ontology concepts. RDF provides technology support to fit heterogeneous data into an ontological model. RDF facilitates applications to model data into a common structure, machine processable format and builds a knowledge base. The knowledge base consists domain ontologies (T-BOX), instance data (A-BOX), axioms and inferred data.

RDF stores data in triple format [8], the triple consists three elements <subject> <predicate> <object>. Subject represents a domain resource such as concept or entity, object represents either a resource or literal and predicate represents the semantic relation between subject and object elements. The framework achieves information interoperability and semantic information retrieval through:
- Read/Extract data from individual domains
- Link instances of diverse domain ontology concepts through appropriate semantic relationships
- Ontology based knowledge management

1) Data extraction

Disaster data extraction: The framework uses a web information extraction framework [16] to extract disaster information published on the web. It takes website URL as input and crawls through all the web pages of the website and extracts disaster information. The extracted information is stored in a CSV file.

Meteorology data extraction: As discussed in section 3.1, the meteorology data from satellites is being distributed in heterogeneous scientific data formats such as NetCDF, OpenDAP, HDF5, GRIB, and NEXRAD etc. There are different libraries to read data from different scientific file formats. The framework develops methods using multiple programming libraries that support data extraction from scientific files based on file format.

2) Mapping data to domain ontology concepts

The data mapping phase identifies relevant domain ontology elements such as concepts, data type properties and object properties to represent the extracted data as instances and relationship between the instances.

Mapping disaster data: This section implements semantic similarity technique to determine more relevant disaster domain ontology concepts for the extracted disaster information. The semantic similarly technique measures the similarity between disaster data elements and ontology elements using WordNet semantic dictionary. It compares disaster event name in the extracted data with ontology concept and other elements with data type properties.
Mapping meteorology data: The weather ontology [17] provides concept descriptions as shown in the figure 4. The framework uses the weather concept descriptions, to identify concepts for the meteorology data extracted from scientific file formats. This section implements an algorithm that organizes the extracted weather data into RDF triples based on closely related ontology concepts and data type properties. In RDF representation Uniform Resource Identifiers (URIs) plays a crucial role. URIs are means of identifying ontology resources such as concepts, properties, and instances. The mapping process uses unique URIs for instance creation.

To link the instances, the framework has designed and implemented an algorithm, shown in table 1. For each concept, the algorithm executes queries on T-Box and A-Box data. From T-Box, it gets object properties and range concepts and A-Box data give instances to be linked.

<table>
<thead>
<tr>
<th>TABLE I. INSTANCE LINKING ALGORITHM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm: Instance Linking</td>
</tr>
<tr>
<td>Input: Ontology concept (Cd) /* Concept whose instances needs to be linked */</td>
</tr>
<tr>
<td>Output: Triples /* Links instances of given concepts with the other concepts */</td>
</tr>
<tr>
<td>Method:</td>
</tr>
<tr>
<td>Begin</td>
</tr>
<tr>
<td>OP ← {opi/object property to which Cd is Domain} /* Gets from T-Box data */</td>
</tr>
</tbody>
</table>
| | Equivalent To +
| | Wind
| | and (hasSpeed some float[> 32.6f])
| | SubClass Of +
| | General class axiom +
| | SubClass Of (Anonymous Ancestor)
| | (hasDirection exactly 1 float[< 360.0f])
| | and (hasSpeed exactly 1 float) |
3) Link instances of diverse domain ontology concepts

The current research defines few spatial and semantic relations in the domain ontologies that can be used to link the instances of concepts of multidisciplinary domains. The object properties that defined in the current work are:

- `airTemperatureAtDisasterEventLocation`
- `solarRadiationAtDisasterEventLocation`
- `precipitationAtDisasterEventLocation`
- `cloudCoverAtDisasterEventLocation`
- `airTemperatureAtReliefShelterLocation`
- `solarRadiationAtReliefShelterLocation`
- `precipitationAtReliefShelterLocation`
- `cloudCoverAtReliefShelterLocation`

This section plays a vital role in linking data from multidisciplinary applications. The current research performs geospatial operations such as within and distance, between meteorological domain instances and disaster event, relief center locations. The meteorology domain instances that satisfy the geospatial operations are linked with disaster domain instances using the object properties.

4) Ontology based knowledge management

The result of section 3.1 is interoperable and integrated information encoded in RDF structure. To do multidisciplinary analysis and semantic information retrieval, there is a strong need of knowledge management system. Three types strategies for knowledge management are discussed in [18]. Strategy 1 uses a single table to store domain ontologies and stance data, strategy 2 uses two tables to separate T-Box and A-Box data. Strategy 3 separates domain ontologies, instance data and inferred data for easy maintenance of knowledge base and reduce information retrieval time complexity. The knowledge management phase uses strategy 3.

B. Semantic Information Retrieval from Multidisciplinary Domains

The existing meteorology data dissemination systems provide effective satellite data retrieval based on the criteria satellite, sensor, date, time, latitude, longitude, etc. Many disaster management systems are following data driven approaches that allow structural queries [19]. This section designs and implements user-friendly query interface to execute ontology concept based semantic queries on the multidisciplinary application knowledge base. Sunitha et al., [20] developed a query interface that allows semantic retrieval on the knowledge base of a domain. The framework enhances the semantic query interface to execute ontology based semantic queries on the knowledge base that includes multidisciplinary domain ontologies, instance data, inferred data. The enhanced method allows SPARQL queries that include concepts, data and object properties from multiple domains.

IV. IMPLEMENTATION AND RESULTS

The current method is implemented in Java, using semantic technology APIs Jena [21] to process RDF data and implement SPARQL query execution, WS4J [22] to compute semantic similarity, scientific libraries to read data from scientific files. Oracle semantic store [23] is used to store and manage knowledge base. Weather, disaster domain ontologies in OWL format are used to address the interoperability issues and semantic information retrieval from the knowledge base that integrates meteorology and disaster domains.

Natural disaster data is taken from international disaster database [24], which is listed in National Disaster Management Authority (NDMA), Govt. of India. Disasters list data is extracted from the web site and transformed into RDF through mapping to disaster domain ontology, RDF triples are shown in table 2. Temperature, solar radiation, precipitation and cloud cover data products of Kalpana1 satellite are taken in HDF5 file format and transformed into RDF, sample RDF triples are shown in table 3.

The methodology brings disaster and weather data into a common structure and machine understandable format using ontology driven semantic technologies that enable
multidisciplinary application interoperability and ontology concept based semantic information retrieval.

V. EVALUATION

This section describes the experiment conducted and the evaluation results. The current research improves information retrieval performance through facilitating interoperability, multidisciplinary information integration and ontology concept based semantic information retrieval. To evaluate the effectiveness of the semantic information retrieval, the current research prepares a query set as shown in table 4, executes the queries in:

- Existing querying system, the search engines of:
  - MSDAC data access
  - The world disaster database
- Current Approach

### TABLE II. RDF TRIPLES GENERATED FROM DISASTER DATA

<table>
<thead>
<tr>
<th>Triples</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.disaster.org#2004_0659">http://www.disaster.org#2004_0659</a> <a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a> <a href="http://www.disaster.org#Tsunami">http://www.disaster.org#Tsunami</a></td>
</tr>
<tr>
<td><a href="http://www.disaster.org#2004_0659">http://www.disaster.org#2004_0659</a> <a href="http://www.disaster.org#start_date">http://www.disaster.org#start_date</a> “2014-12-26”</td>
</tr>
<tr>
<td><a href="http://www.disaster.org#2004_0659">http://www.disaster.org#2004_0659</a> <a href="http://www.disaster.org#end_date">http://www.disaster.org#end_date</a> “2014-12-26”</td>
</tr>
<tr>
<td><a href="http://www.disaster.org#2004_0659">http://www.disaster.org#2004_0659</a> <a href="http://www.disaster.org#country">http://www.disaster.org#country</a> “India”</td>
</tr>
</tbody>
</table>

### TABLE III. RDF TRIPLES GENERATED FROM METEOREOLOGY DATA

<table>
<thead>
<tr>
<th>Triples</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.weather.org/meteorology#2014-12-26T07:30_8.48_79.22">http://www.weather.org/meteorology#2014-12-26T07:30_8.48_79.22</a></td>
</tr>
<tr>
<td><a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a> <a href="http://www.weather.org/meteorology#CatastrophicHurricane">http://www.weather.org/meteorology#CatastrophicHurricane</a></td>
</tr>
<tr>
<td><a href="http://www.weather.org/meteorology#hasLatitude">http://www.weather.org/meteorology#hasLatitude</a> “8.48”</td>
</tr>
<tr>
<td><a href="http://www.weather.org/meteorology#hasLongitude">http://www.weather.org/meteorology#hasLongitude</a> “79.22”</td>
</tr>
</tbody>
</table>

### TABLE IV. QUERY SET

<table>
<thead>
<tr>
<th>Query Number</th>
<th>Query</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Show moderate precipitation</td>
<td>Meteorology</td>
</tr>
<tr>
<td>Q2</td>
<td>Locate sunny weather state</td>
<td>Meteorology</td>
</tr>
<tr>
<td>Q3</td>
<td>Display biological disasters</td>
<td>Disaster</td>
</tr>
<tr>
<td>Q4</td>
<td>Find floods followed by earthquakes</td>
<td>Disaster</td>
</tr>
<tr>
<td>Q5</td>
<td>Locate high precipitation at hydrological disasters occurring places</td>
<td>Multidisciplinary</td>
</tr>
<tr>
<td>Q6</td>
<td>Fire disasters during sunny weather state</td>
<td>Multidisciplinary</td>
</tr>
</tbody>
</table>

MODAC provides data products rather than specific information for a query. For query Q1, it retrieves a data product “QPE”. QPE file contains 7370 records, but only 277 are relevant. MOSDAC data access does not answer query Q2, due to lack semantic knowledge on sunny weather state. The data in the world disaster database are classified, hence the world disaster database search engine can provide better performance for concept based queries such as Q3. However, it does not answer the queries like Q4 which have semantic relations. Either MODAC data access or world disaster database search engine can answer multidisciplinary semantic queries like Q5 and Q6, due to interoperability issues. The current approach provides better query performance for all the queries through achieving interoperability across the weather and disaster disciplines.

Precision, recall, and F-measure are computed for each query executed against the data driven and current approaches. Precision, recall, and F-measure are defined as follows [25]. Figure 5 shows precision analysis, figure 6 shows recall analysis and figure 7 shows F-measure analysis.
From the analysis of current and existing approaches, it is being observed that the current approach showed better performance than the existing search approaches. In the current research, the effectiveness of the query performance is depending information interoperability that is achieved through transforming data as instances of ontology concepts and linking instances of multidisciplinary domains.

**VI. CONCLUSION**

This paper presents a generic framework to achieve information interoperability among multidisciplinary domains data and ontology concept based semantic information retrieval. The current research illustrated the framework through taking data from disaster management and meteorology domains. The framework is evaluated the impact of information interoperability on semantic retrieval using precision, recall, and F-measure as metrics and compared with the existing search approaches. It is being observed that information interoperability among multidisciplinary applications achieve through the current approach enhances semantic information retrieval performance. Further, the research can be extended to achieve information interoperability in big data from multidisciplinary domains.

**VII. ACKNOWLEDGEMENT**

The authors would like to acknowledge the time, effort and constant technical support extended by Dr. PS Acharya, Head (TWF)/ CEO (NSDI)/Scientist ‘G’ (NCSTC)/ and Coordinator (MMGA), Department of Science & Technology, Govt. of India., Dr. N. L. SARDA, Professor, Computer Science & Engineering Department, IIT BOMBAY, Powai Mumbai., Dr. M. K.Ponnuraj, State treasurer, Tamilnadu Government doctors association (TNGDA), Tamilnadu, India. The authors wishes to acknowledge the financial support provided by the Department of Science & Technology (NRDMS, DST), Govt. of India. The authors would also like to thank Mr. A. Radha Harish, Dr M. R. Nayak and Mr. G. Ranganath, for their constant support and review at various stages of this work.

**VIII. REFERENCES**


[5] Sunitha A, Suresh Babu G. Ontology-Driven Knowledge-Based Healthcare System an Emerging


