SEMANTIC WEB IN MEDICAL INFORMATION SYSTEMS

Prashish Rajbhandari[1], Rishi Gosai[2], Rabi C Shah[3], Pramod KC[4]
[1]Department of Information Technology,
Indian Institute of Information Technology, Allahabad, India

ABSTRACT
The complex and heterogeneous storage architecture of hospitals makes it taxing to extract and collect useful information from the data repositories. In this paper we propose the application of an ontology-based method using the concepts of Semantic Web for obtaining interoperability between hospitals in the medical information system domain. We aim to annotate data in Relation Description Format (RDF)/Extensible Markup Language (XML) via common ontology, then interlink the available data from various medical institutions and consequently apply data mining to find meaningful results. The unified data will be stored at a central server accessible by all hospitals

KEYWORDS: *Semantic Web, RDF, XML, Medical Systems, Ontology, Client Server.*

I. INTRODUCTION

Information technology has grown drastically over the last decade due to advancement in networking technologies. Along with the advancement in size, the underlying architecture and the structural complexity has also changed.

Health Information Systems is a coalescent system formed by the intersection of information sciences and health care. It is a source of rich information of medical information, which is of primary importance for the doctors, nurses or other health care experts to treat or diagnose a patient. Health Information includes the storage and maintenance of health records and protection (by law) of patient information.

The systems developed and implemented are proprietary and hence have a custom architecture. The usage of Electronic Health Records is increasing rapidly leading to a huge amount of digital formats. Comprehending the mixed data is becoming increasingly complex for the physicians, which is leading to slower decision making. The integration of the data is a very tedious and difficult process. It would require a considerable amount of programming effort to interlink the data due to different interfaces and architectures. Medical data requires a unified approach because of the promising opportunities in the field of Bio Medical research by world health organizations and agencies.

Semantic web technology provides a common framework such that it is easy to access and process information by machines. The heterogeneous nature of the health care data makes it a very suitable candidate for Semantic Web Application. It aims to analyze the data by comparing the similar ontology and then interlink the data as a whole allowing easier querying by the end user in the repository of information. There are already some efforts in this field.

In this paper, we aim to develop a Medical Information System that will be based on semantic web technology for sharing data between different hospitals. The proposed system can understand data from all the hospitals because of common ontology and semantic interoperability. The data is stored in a common repository which can be used for machine learning or data mining. The clients can then access the results and data from the server.
This paper includes a description of the related works in section II. It is followed by an overview and the details regarding the proposed approach of the SMIS (Semantic Medical Information System) in section III. The results of the implementation of the system is mentioned in the following section. In Section V, the conclusion and future scope of the paper is mentioned.

II. RELATED WORK

In the past few years, scientific community, especially from medical domain, has involved many efforts and resources in the development of eLearning technologies and materials. The Semantic Web is considered as an extension of the World Wide Web where the data present can be integrated and evaluated easily. Although gaps in implementations still exist and the adoption is hampered by the lack of pre-installed base there has been significant research and development in the field of semantic web.
Web Services like MedCIRCLE [4] and MedCERTAIN [1] are Semantic projects with the aim to guide users to health information on the internet and to filter quality health information available on the net. The related work focuses on developing a multi-ontology structure with peer to peer networking. The P2P style avoids physical and semantic bottlenecks in the information exchange. PeerDB [2], an object management system, provides detailed searching services. It is a network of local database on peers. Data is shared without a global schema by using meta-data for each attribute. P2P Information exchange and Retrieval (PIER) is a commonly used query engine [5][6]. It is used for processing queries in a distributed network.
For ubiquitous computing environment [7] provides context aware framework ontology where the rules were used to convert low level context to high level context, for the purpose of providing personalized healthcare services to users. But for companies with less formal training the proposed approach is too complex. Another approach mentioned in [8] proposes merging two healthcare ontologies SNOMED CT and ICNP. Due to the proposed merging the different perspectives of the two ontologies are maintained without losing important information. But there is no efficient technique proposed for merging large ontologies. E – Neuroscience data addresses [9] the integration problem by the using the oracle RDF data model. The data are extracted from brain plus and swan. The proposed system provides the conversion of relational to RDF architecture. This paper [10] presents an organized medical advice for patients suffering from diabetes with the help of food ontology. This ontology is tested to share knowledge between the different stake holders in the pipes project for diabetic control. The approach proposed in [11] states an ontology based knowledge framework that is capable of providing personalized healthcare to users by retrieving all required knowledge such as patient care, drug prescriptions etc. But the system does not use any semantic rule engines in its implementation.
Apart from ontology based medical systems there are some rule based healthcare systems proposed by researchers. The paper [12] presents rule based information extraction from medical documents. The information hence extracted is then grouped and classified into the required complex templates. The method proposed can also be used to select the most prominent and important diagnosis and the patients that require special attention. But the development of the system is a time consuming approach. In [13] the dynamic modification of the workflow is described. If an exception occurs, the system identifies the affected workflow region dynamically corrects the affected workflow. Rules are used to detect semantic exceptions and decide about which activities are dropped or added. The paper [14] mentions a protocol that can handle exceptions and correct the workflow dynamically. A set of predefined guidelines is used. It gives a rule based exception detection and correction method.

III. SEMANTIC MEDICAL INFORMATION SYSTEM (SMIS)

3.1 An Overview

It is an ontology-based medical information system that captures the data from various heterogeneous systems and stores it at a central server. The underlying principle is the same as that of a client-server model. Since the database architecture of every client (Hospital) is specific and incompatible with other client databases, there is a need to unite them under a common structure. The server, in our case SMIS Server, merges these databases and follows a common ontology to unite the different relational databases. The data is transformed according to information contained and is stored in a unified and understandable manner. The SMIS server is accessible by all clients and since it has a common ontology throughout it is easy to query for information or mine for research. For example, consider the disease yellow fever. All the clients would contain relevant information regarding the symptoms about the disease. If a researcher or doctor needs this data he will have to study the individual database structure for each client leading to strenuous and tedious work overhead. The same information can be combined and stored in a central server, under a node titled yellow fever. All the required data can then be queried by only taking into consideration the database layout of the central server because the SMIS server holds information from all the clients (hospitals) connected to it.

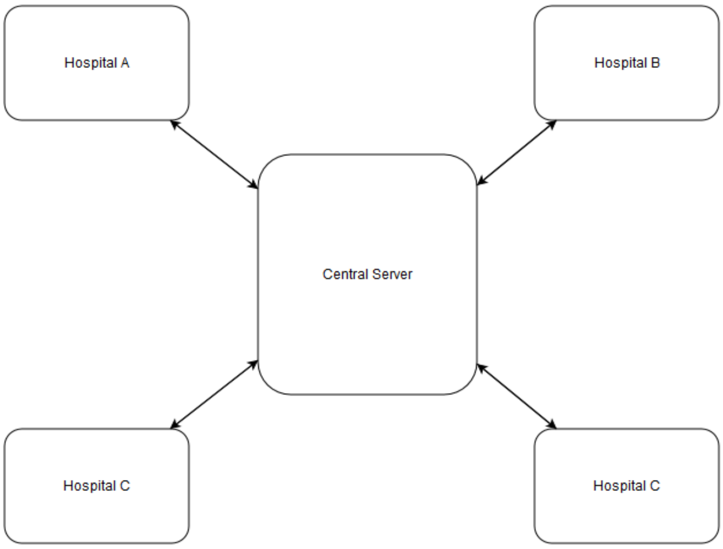

Figure 1. Basic Architecture of Proposed Approach

When a new client is registered to the server, it is authenticated before it can use the services provided by the Semantic Medical Information System. Services include request for information from the central server and updating the data on the server. During registration, a mapping process takes place which uses the ontology of the client to create an RDF file containing the information the client needs to send to the server. This file is then transferred to the server where the RDF graph pertaining to each client is created and merged into a unified RDF graph containing the entire data from all the clients. The client Application can be in the form of a web browser or a desktop application connected to the network containing the SMIS server. Each client is provided with a username and password through which it can access the SMIS server and request information. This data can be used individually by each client for diagnostic or research purposes.

3.2 Details of Proposed Approach

3.2.1. Mapping Phase:

This is the most important phase of the Medical Information System. Since every hospital will have its own database schema, the central server needs to map each hospital's database schema into common ontology classes. This is done using a standardized ontology that is common to all the clients. The primary function of medical ontology is to map data from each hospital to the classes

defined in the ontology. This is the extracted data mapped into known classes that is understandable to the central server and therefore incorporate it in its own RDF Model.
The mapping application that we built for hospitals would initialize itself during setup and ask for the mapping schema of the hospital's database. It is assumed that the database isn't distributed and data is available to the application from a single database server. The application can be triggered for a change in the mapping schema if there are changes to the overall change in database schema. The changes and new data would be then transmitted to the central server for changes in its model.
The application required database privileges of the client during setup. The application then lists down all tables and its corresponding columns so that the administrator can map each table-column with our defined common ontology class. After the administrator maps table-column to classes, D2RQ tool runs in the background to create an RDF Model of the data of relational database. The RDF file is generated in the system that is then transmitted to the central server. After transmission is complete, the file is deleted or kept as backup. Mapping phase is vital because it is where the administrator needs to map hospital's data to ontology classes.
The algorithm for Conversion of RDBMS to RDF Model:

```
BEGIN CONVERSION
    If Connected
        Start Mapping Interface
        For Each Class Map
            If Map present
                Get Database Maps with columns
                Store Database Map
            Else
                Continue
        End For
    If Map present
        Start D2RQ
        Convert into RDF
        Store RDF
    Else
        Start Setup
END CONVERSION
```

3.2.2. Connection Phase:

Mapping phase is specialized to database administrator but this is the phase where normal users (doctors) can use the system to get data from various hospitals. After RDF Model is generated in the client's machine, it needs to connect to the central server. The central server must provide some kind of authentication for each hospital for file transfers. After authentication, the hospital can send its RDF model to the server. The Central server stores all the received RDF Model from all hospitals and merges them to create a common RDF Model. The nodes having the same URI is merged together in the RDF Model so that it is abstracted to have generated from common node classes.
The algorithm to establish connection with central server:

```
START CONNECT CENTRALSERVER
    If Maps complete
        If login successful
            Connect to Server
            Transfer RDF
        Else Login Fail
    Else
        Start Mapping
END CONNECT CENTRALSERVER
```

3.2.3. Communication Phase:

After Connection phase is successful, hospitals can query in the SMIS Server for various information regarding diseases. Since all the data are merged into a common RDF Model, querying about disease will give results from all hospitals. Since, many results can be same, we use duplicate content filtering to filter duplicates from hospitals. The SMIS Server provides specific services to the clients to query into the RDF Model. User forms are created in the application for basic querying in the RDF Model like getting symptoms of a particular disease from all hospitals or selected hospitals.
A single RDF Model is created in the central server but if a hospital queries results from selected list of hospitals, then the server temporarily builds a model merging the models of the hospitals and sends in result. This facility helps hospitals get specific results from registered hospitals.

```
START CLIENT TO SERVER
        If Connected
                While !Log out
                        Query to Central Server
                        Transfer Updated RDF
                End While
        Else
                Terminate
END CLIENT TO SERVER
```

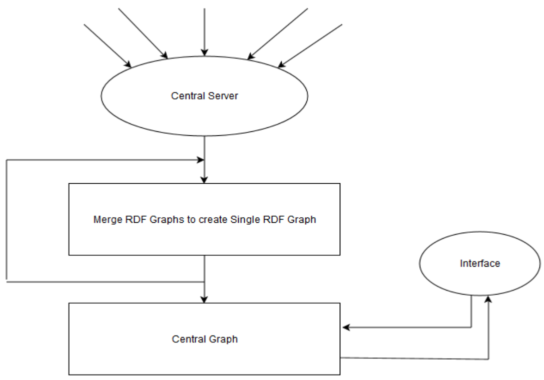

Figure 2. Dataflow for Mapping

IV. RESULTS AND DISCUSSIONS

The implementation of our SMIS was done in Java language using RMI (Remote Method Invoking), D2RQ and JDBC. We used Stanford Protégé tool to create our standard ontology.
The prototype ontology that we used:

Disease Class
  hasSymptoms : Symptom Class
  hasDefinition: Definition Class
  hasCure: Cure Class

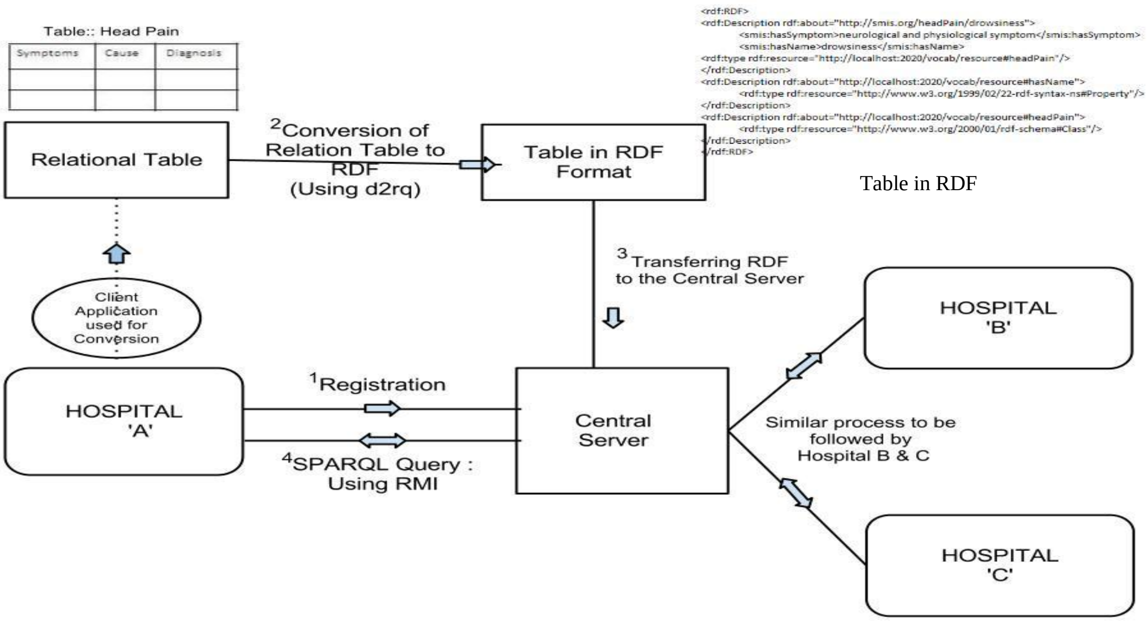

Figure 3: Simulation of the proposed model

Clients were set up in different systems and were connected to the central server in a network. The clients could incessantly communicate with the server using sockets for file transfer. Clients couldn't communicate with each other directly. After clients authenticated with the central server, it could send its RDF Model and also, query remotely to the central server. Clients could query using Java RMI, which is instated in the client application. The services provided to the clients are all remotely invoked in the central server. No data is stored in the client machine to reduce load.

Querying in the server RDF Model was possible using SPARQL – a querying language that can retrieve data/resources stored in RDF. Clients could query for data from any other client or all clients. The central server used Apache Jena to combine RDF Models of clients depending on the clients from which data is to be received.

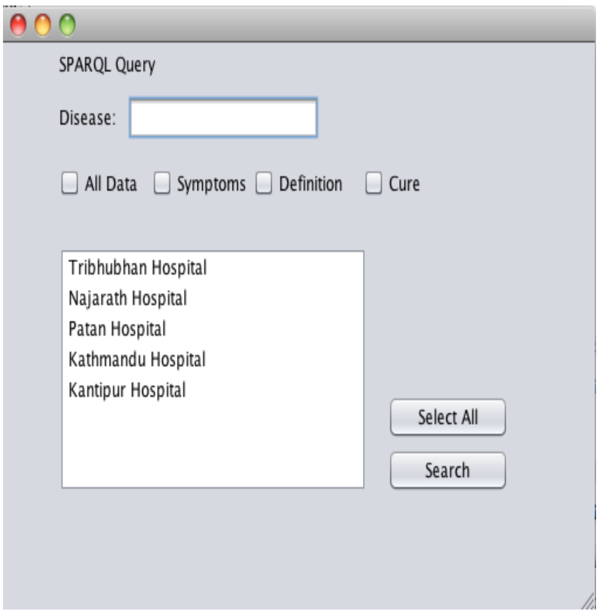

Figure 4: The visual layout of SMIS

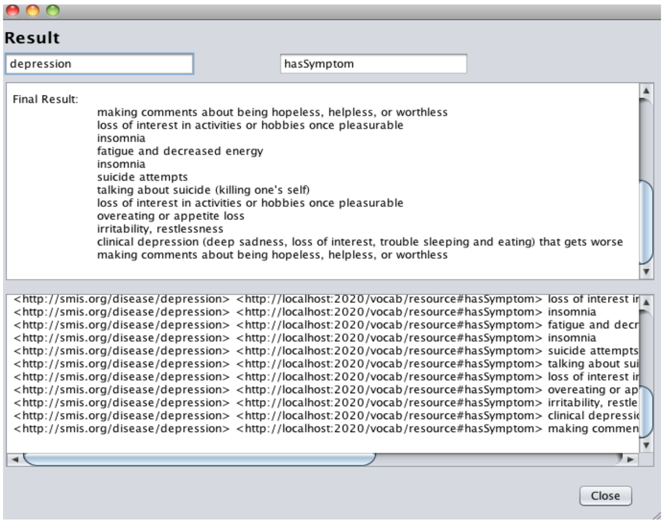

Figure 5: Result of a query

As can be shown in Figure 5 the SMIS system proposed in the paper successfully integrated the required information from all the client hospital and returned the SPARQL query successfully. Most of the existing semantic web approaches do not follow the ACID properties of database. Our SMIS approach has overcome this shortcoming by incorporating the atomicity, consistency, isolation and durability of database transactions. It also provides the reliability of a client server model. Our approach emphasizes on the unification of the heterogeneous clients (Hospitals) in a complete and wholesome manner keeping in mind the database system's integrity.

V. CONCLUSION AND FUTURE WORK

Due to the availability of ready-to-use, mature domain ontologies in medicine, we have worked on ontology-based data integration system architecture to combine the client hospitals information and store it at a central server. Using Semantic technology we were able to create a system which is able to store the data from various heterogeneous sources in a unified manner. The unified RDF graph was successfully created at the server end using Apache Jena and SPARQL query language was used to get the required results at the server.

In the near future, we will investigate the scalability and efficiency of SMIS approach in very large data storing relational databases. Also, we plan to improve the SMIS approach by adding the support for distributed client databases. We plan to expand the scope of the project from integrating client hospitals to mining the unified data at the central server to help in medical research.

ACKNOWLEDGEMENT

This research paper was possible through the wonderful working environment and infrastructure provided to us by our honourable Director (IIITA), Dr. M.D. Tiwari. Also, we would like to thank our parents for providing us encouragement and financial support without which this research paper would not have been successful.

REFERENCES

[1]. S. Gnanambal& M. Thangaraj, "Research Directions in Semantic Web on Healthcare",*International Journal of Computer Science and Information Technologies*, Vol. 1(5) , pp449-453, 2010

[2]. V.S. Agneeswaran, "A Survey of Semantic Based Peer-to-Peer Systems," Technical Report, Distributed Information Systems Lab (LSIR) EcolePolytechniqueF´ed´erale de Lausanne, 2007

[3]. B. McBride, "Jena: A Semantic Web Toolkit," *IEEE Internet Computing*, 6(6), pp55-59, 2002

[4]. Mayer, M. A., Darmoni, S. J., Fiene, M., Eysenbach, G., Köhler, C.&Roth-Berghofer, T.,"MedCIRCLE—modeling a collaboration for Internet rating, certification, labeling and evaluation of health information on the semantic world wide web.", *Proceedings of MIE 2003, Eighteenth International Congress of the European Federation for Medical Informatics*. 95,pp667–7, 2003

[5]. N. Khozoie, "Health Information Management on Semantic Web :(Semanti HIM)," *International Journal of Web and Semantic Technology(IJWest)*, Vol 3, No. 1, 2012

[6]. D. Dou, "Integrating Databases into the Semantic Web through an Ontology-Based Framework," *Data Engineering Workshops*, 2006

[7]. Eunjungko, Hyungjiklee, Jeun woo lee, 2007, "Ontology based context modeling and reasoning for U Healthcare," *Web Information Systems Lab, Chugnam National University*.http://winslab.cnu.ac.kr/resource/labseminar/Seminar2007-1/IEICE_kej_final_paper.pdf

[8]. Fahim Imam, Wendy Maccaull, 2008, " Integrating Healthcare Ontologies: An inconsistency tolerant Approach and Case study" http://www.logic.stfx.ca/docs/prohealthPaper_July_3_Imam_McCaull.pdf.

[9]. Hugo Y.K. Lam, Luis Marenco, Tim Clark, "Semantic web meets e-neuro science: An RDF usecasehttp://www.oracle.com/technology/industries/life_sciences/press/semantic_web_meets_eneuroscience.pdf.

[10]. Jaime cantais, David Dominguez, Valeria Gigante, Loredana Laera, 2007, "An example of food ontology for diabetes control", http://www.csc.liv.ac.uk/~floriana/PIPS/papers/FoodOntology.pdf.

[11]. Jiangbo Dang, Amirhedayati, "An ontological knowledge framework for adaptive medical workflow", *Journal of biomedical informatics* 41 ,829-836 ,2008

[12]. Agnieszka Mykowiecka, Malgorzata Marciniak, "Rule based information extraction from patients clinicaldata", *Journal of biomedical informatics*42,923-936, 2009

[13]. R. Muller, E. Rahm, "Rule based dynamic modification of workflows in a medical domain",*Spinger*,429-448, 1999

[14]. Yan ye, Zhibin jiang, "A semantic based clinical pathway workflow and variance management framework", *IEEE* 978-1-4244-2013, 2009

AUTHORS

Prashish Rajbhandari is currently pursuing his B. Tech (IT) degree in Indian Institute of Information Technology, Allahabad. He is the brain of forty stones team. He has keen interest in Web Development and is fluent in various scripting languages like html/css, php, javascript, jquery and various CMS (Codeignitor, RoR).

Rishi Gosai is an under-graduate student majoring in Information Technology at The Indian Institute of Information Technology Allahabad (IIIT-A). His field of interests include Database Systems, Operating Systems and Artificial Intelligence.

Rabi C Shah is currently pursuing his under-graduate program with a major in Information Technology in Indian Institute of Information Technology Allahabad (IIIT-A). His field of interests include C, C++, Linux and Semantic Web Technology. He is the founder of fortystones.com and has completed many projects in java and php.

Pramod K.C is an ebullient and persistent student pursuing his under-graduate (IT major) at Indian Institute of Information Technology, Allahabad. His interests lie in Networking, Semantic Web Technology, Database Systems, Java and also has good Management skill.