Medical Data Management and Interoperability in E-Health Systems

M. Shamim Hossain*, Mehedi Masud**

*College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

**Computer Science Department, Taif University, Saudi Arabia

mshossaain@ksa.edu.sa, mmasud@tu.edu.sa

***Abstract*—Efficient management of medical multimedia data are important research issues in e-Health systems. Proper management of data facilitates efficient data sharing and querying. In this paper, we propose such an approach for data interoperability among heterogeneous data sources. The approach provides a mechanism for sharing data among the sources by overcoming the heterogeneity issues. The framework has been tested with various user queries and the accuracy of the query results evaluated by means of precision and recall methods.**

***Keywords*—Interoperability, ehealth, medical multimedia data, query processing**

I. Introduction

A collaborative e-Health environment facilitates sharing and accessing patients medical data (picture, video, audio, text, etc.) remotely and communication between parties such as doctors, hospitals, and researchers. In the e-Health collaborative environment there exist different types of multimedia data (e.g. Magnetic Resonance Imaging (MRI), Computed Topography (CT), X-ray images, electrocardiograms (ECG) images as well as medical information in the forms of charts, graphs, and others) to share and store. An appropriate metadata structure is necessary for querying the multimedia items related to patients' medical treatment. A proper metadata structure describes multimedia content effectively, and thus enriches the information related to the multimedia content [16]. Hence, querying multimedia content is faster, while the query results are more accurate. Thus, the challenge is to store and represent the metadata of the diverse medical multimedia data using a framework that ultimately facilitates efficient sharing and querying of such heterogeneous data.

Dealing with the heterogeneity of data is a big issue in a collaborative healthcare system. In a typical e-Health system, data about patients are stored in distributed sources (e.g., physicians, hospitals, laboratories, etc.). These data sources are autonomous and therefore, independently manage data. For collaborative data sharing in healthcare data sharing system, data sources agree to share data with one another with different goals, schemas, and data. A logically integrated view of data sources is required to enable the use of this distributed data by healthcare professionals to improve the care of patients. The integration of distributed heterogeneous data sources for sharing data among the sources in healthcare is challenging, because all the data sources may not agree on a common storage structure. Instead, a data sharing policy is established that presents a logical integrated view of all the autonomous data sources [1]. The data sharing policy considers the issues of resolving data heterogeneity among the sources. The challenge in this case is to provide interoperability as a sharing policy that can resolve the data heterogeneity of the sources. Schema level and data level are two types of heterogeneity of data when considering interoperability of data [2][3]. Generally, data sharing between data sources is provided mainly through the use of views or schema mappings, i.e., queries that map and restructure data between schemas.

Finding a good service provider in the Internet for a particular health care service is another important challenge. For the last years, many web-based information systems have been proposed to provide useful information, recommend better services to the users [4]. These systems use intelligent information filtering and decision support tools aimed at providing product, information, and service recommendations personalized to the user's needs and preferences at each particular request context [5][6]. Although, there are lot of scientific and technological development progress over the recent years in the healthcare delivery, a significant portion of the decision-making information on choosing doctors, clinics, and the treatment of a patient's illness is still based on manual searching and processed using unstructured information or even hand-written notation.

The objective of this paper is to propose a solution of data interoperability that goes beyond just the information systems connectivity but allows sources to share and search medical data efficiently internally or externally from the e-Health service providers world. Mainly, our contribution is to propose an approach for representing multimedia content of medical data in a e-health collaborative environment. For this we propose a metadata management system and a distributed query processing mechanism.

II. Multimedia Content Management

In a collaborative e-Health network, the large amount of data are stored in various systems that may include multimedia content management and archiving systems (e.g.

PACS, RIS, etc.), health information systems with autonomous data sources (e.g., hospitals, doctors, pharmacies, etc.), content repurposing services, and different clients using different types of smart devices for using multimedia contents. The stored data may be structured (e.g. patients information, medical records, drugs, etc.) stored in traditional database systems, text in the form of reports and prescriptions, medical test images (e.g. computed tomography, magnetic resonance, X-rays), videos and voice data. The content is acquired in real time and stored in different distributed locations. In order to deal with the proliferation of multimedia content and metadata it is advised the use of distributed architectures where information can be stored on multiple sites that manage and compute queries [16].

There are fundamental differences between multimedia queries and standard relational database queries [10]. The rich content contained in multimedia resources demands more complex description of the content. We can search multimedia resource from different contexts. For example, an image can be searched not by only "file name" for content but also by image related to patients, diseases, format, etc. A query for a ceratin multimedia resource may be denoted as: patient="David", format="Jpeg"', diagnosis="X-ray".

In distributed systems, answers to queries may require data from different data sources or peers. Therefore, efficient distributed indexing or catalog is needed to find sources and process queries. To process a query, we cannot always move the data pertinent to a query to a single server and do processing there, both technical and political reasons. The lack of global knowledge about data poses challenge for any peer to execute a query over the systems.

In this paper we present an efficient multimedia content management and query processing strategy which is applicable in e-Health distributed systems where different parties maintain their data independently. Our query processing is based on dynamic query decomposition and planning based on the cost of intermediate data processing, cost of data transmission, and up-to-date information of the data.

Consider a case where several parties acquire different types of medical content such as text (medical reports, prescriptions), videos (patients footage of movements, surgeries), images (X-Rays, tomography, ECGs) about patients. The parties are located in different sites (cities, countries) and their acquired multimedia content is stored on local servers. To receive the desired information such as images of X-rays, prescriptions, videos of tumor surgery, the local server must be aware of the distributed servers and capable of searching and providing data for the requested query. In order to find required content efficiently an efficient metadata management and structure for multimedia content is required. Figure 1 shows generic process to produce metadata from multimedia content over the data sources. When a multimedia content is produced in a local peer it follows the process as shown in Figure 1. The metadata extractor is applied to the multimedia content to produce a metadata structure. We can use extractors [7]-[9] to generate metadata from a multimedia resource. The multimedia resource is stored in a local storage and metadata are stored in the local catalog. A subset of the metadata is also stored in an export catalog. The export catalogue allows other nodes in the network to search for multimedia content. It also defines the sharing policy of the local multimedia resources. We describe the process of creating a metadata structure below.

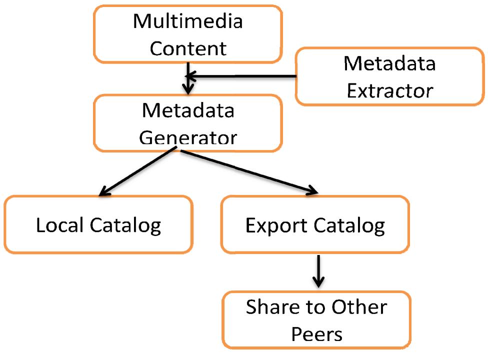

Figure 1. Process of generating metadata from multimedia content for sharing

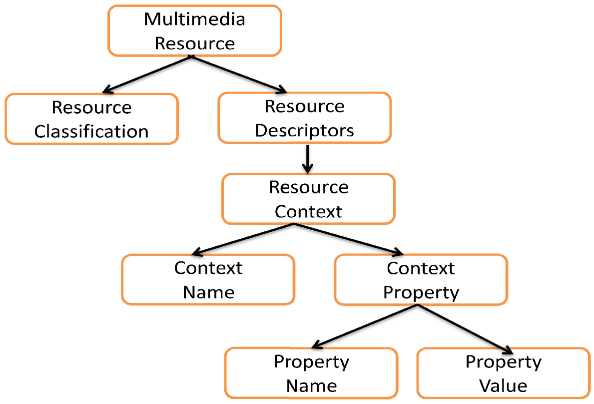

Figure 2. Process of generating metadata structure from multimedia content

For each resource $\boldsymbol{R_i}$ a metadata structure $\boldsymbol{M_i}$ is generated which is formalized below:

$\boldsymbol{M_i} = <\boldsymbol{R_c}, \boldsymbol{R_d}>$, where $\boldsymbol{R_c}$ represents resource classification (e.g. video, audio, text, image etc.) and $\boldsymbol{R_d}$ represents description of $\boldsymbol{R_i}$.

$\boldsymbol{R_d} = \{\boldsymbol{C_i},...,\boldsymbol{C_k}\}$, where is C_i represents context of the resource.

$\boldsymbol{C_i} = <cname, \{p\}>$, where *cname* represents a context name related to the resource and $\{p\}$ represents set of properties of the context. The generation process of metadata structure of a resource is shown in Figure 2.

Example: A patient 'David' visited to a diagnosis center 'Fast' for chest X-Ray referred by his physician 'John'. When the X-Ray image is passed to the extractor the metadata structure as shown in Figure 3 is generated.

The structure allows a user to search a resource associated to resource classification and metadata. When a user submits a query, the system identifies the identifier of the resource from resource classification and metadata by analyzing the query string. Once identifier is found the corresponding resource is $\boldsymbol{R_i}$ is obtained.

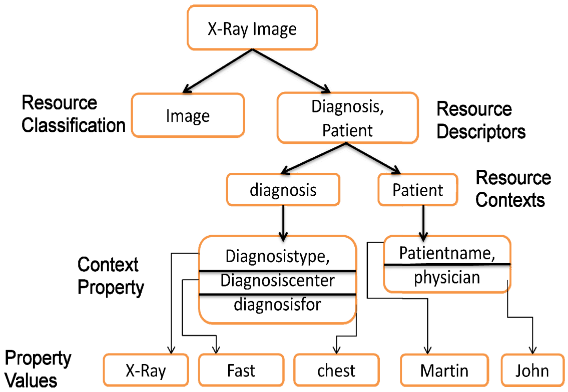

Figure 3. Example of generating metadata structure from multimedia content

III. Query Processing

Now we propose a query processing mechanism to find resources for a user query. The query processing is completed in two-phase.

(i) Resource matching and finding and (ii) Resource collection and query execution.

First, the metadata matching strategy is applied to locate potential resources matching the user query. The metadata of the resources that matches the user query are returned to the query node for two purposes. It allows the user to select the more relevant resources. This is to minimize information overloaded when data may be partially matched or having descriptors but semantically different. Moreover, this can minimize transmitting data that are not useful to the user. Phase two begins after the user has selected the desired resources. In phase two, the query will be directed to the nodes containing the selected resources, and answers are finally returned.

A. Resource Finding:

When a node receives a query it formulates the query into a logical query plan by analyzing the query, local metadata catalog, and export catalog. The query plan is tentative search and resource collection plan matching the query description. After that the node creates an agent corresponding to the user query **q**. The agent performs the following operations.

1. Finds resource classification and descriptors requested by the query.

2. Generates two types of agents: A_1 and A_2. A_1 finds the resource R_i corresponding to the query **q** at the local source.

4. Query **q** and agents A_2 are dispatched to the acquainted nodes. Agent A_2 searches the metadata catalog in the acquainted node and performs resource finding operation.

5. Agent A_1 waits for the initial answers (resource metadata) from acquainted nodes. A_1 sends the initial answers to the users for selection and verification.

B. Resource collection and query execution

After receiving the initial answers, the agent A_1 at the query node selects the nodes from which the resources for the query will be collected. The steps are given below:

1. The query node creates an agent A_3 that performs resource collection.

2. Agent A_3 are then sent to the selected nodes and returns answers directly from the remote node to the query node.

3. After generating the result all the agents killed.

In our approach, we provide a dynamic query execution strategy that may involve data shipping, the query shipping, or the query with data like in MQP [15], which completely depends on the intermediate results of the query in a node. When a query is initiated, the local node generates a query execution plan based on the local information of its acquaintances. The local node then evaluates the query and cache the partial results. Before passing the query to its acquaintances for further execution, a node exchanges some messages with its acquaintances, and takes decision about the prospective execution plan.

IV. Evaluation

In this section, we present performance study conducted to evaluate our approach. We studied the accuracy of query results for different user queries. For the experiemnt, we used a simulator [11] that was implemented in Java to evaluate the proposed query processing mechanism.

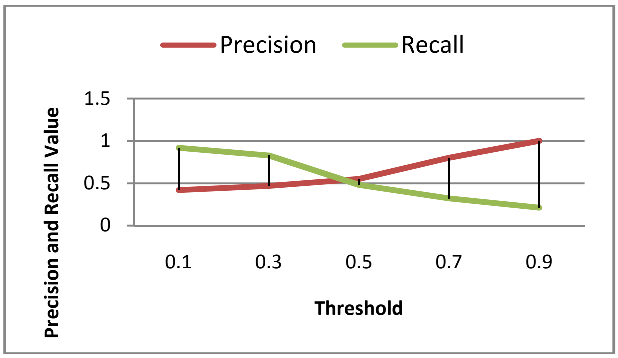

Figure 4. Precision and Recall values of Resource Matching

For the experiment, we generated a large number of metadata for different categories of multimedia content. First, we created a set of classifications C of the resources. In each classification, we produced set of virtual resources. For each resource, we also produced 2−5 descriptors which gives semantic meaning and properties of the resource, since users queries are not expected to have too many metadata. We use the standard precision and recall measures as the performance metrics. Precision measures the purity of search results, or how well a search avoids returning results that are not relevant; recall refers to completeness of retrieval of relevant items. We consider a multimedia content to be relevant to the query if more than k properties from the content match. In our study, we have set k to be 2. For each resource that is examined, we compute its matching score. We varied the threshold value from 0.1 to 0.9. For all the results returned, we compute its precision and recall. The results are shown in Figure 4. We also show the accuracy of query results of our framework to find resources. The results are shown in Fig. 5. We consider queries with different number of conditions in the queries. For this evaluation, we considered five different queries with one to five conjuncts, respectively. The results show that a query with fewer conjuncts returns more resources, but have less accurate results.

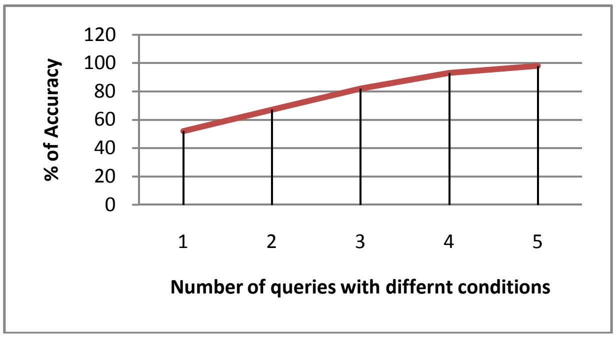

Figure 5. Accuracy of query results

V. Related Work

Fioravanti et al. [12] presented a solution for data collection and storage from a variety of medical devices through. However, we provide solution of patient centric data sharing among the parties in an e-Health system to facilitate collaboration.

Kim et al. [13] proposed a client–server agent-based system that allows sharing and management of medical data. It creates a portal to each system; however, the approach is centralized and all user queries must pass through an agent server, called the multimedia medical agent. On the contrary, our approach is distributed. In our approach, we also use agents to find multimedia resource in a peer-to-peer fashion. The agents also provide a dynamic query execution plan to limit data transfer over the network.

Laborie et al. [14] proposed a framework for managing and querying distributed multimedia metadata that involves transferring only a concise version of the distributed metadata to a central server instead of indexing and storing metadata locally. The central server is used to answer very general queries or locate remote servers that may contain the desired specific information. On the other hand, our approach is totally distributed and the query is processed in a distributed way. Only the peers that want to share data access the export dictionary, which is used for local and remote queries to find sources.

VI. Conclusions

e-Health systems are patient-centered and different health service providers need to share data, knowledge, and services for collaborative healthcare of patients. Managing and searching of medical data distributed over sources are the challenging issues in e-Health systems. Considering these needs we proposed a framework for multimedia content management in the distributed sources. We also proposed a distributed agent-based query processing mechanism to find patients' related medical data. One of the features of our approach is that users receive content information as metadata before receiving the final content. The contents metadata is returned to the user first, and the user decides the desired resources. This reduces irrelevant content transmission over the network.

In our future work we like to propose a service oriented architecture with agent based technique for multimedia service composition, which allows for more scalable and robust multimedia delivery.

Acknowledgment

This work was supported by the Research Center of College of Computer and Information Sciences (CCIS), King Saud University by the research project RC120902. The authors are grateful for this support

References

[1] M. Lenzerini. Data Integration: A Theoretical Perspective. In *PODS*, 2001.

[2] A. Kementsietsidis, M. Arenas, and R.J. Miller. Mapping Data in Peer-to-Peer Systems: Semantics and Algorithmic Issues. In *SIGMOD*, 2003

[3] L. Serafini, F. Giunchiglia, J. Molopoulos, and P. Bernstein, "Local relational model:a logocal formalization of database coordination," Tech. Rep., Informatica e Telecomunicazioni, University of Trento, 2003.

[4] P. Lamber, A. Girardello, F. Ricci, and M. Mitterer. MobiDay: a Personalized Context-Aware Mobile Service for Day Hospital Workflow Support In *Proc. of 12th Conference on Artificial Intelligence in Medicine (AIME)*, 2009

[5] G. Adomavicius, and A. Tuzhilin. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. In *IEEE Transactions on Knowledge and Data Engineering*, vol 17(6): 734-749, 2005.

[6] J. B. Schafer, J. A. Konstan, and J. Riedl. E-commerce recommendation applications. In *Data Mining and Knowledge Discovery*, 5(1/2):115-153, 2011

[7] S. Ayache, J. Gensel, and G.M. Quenot. Clips-LSR Experiments at Trecvid 2006. In *Proc. Trecvid 2006 Workshop*, 2006. http://www-nlpir.nist.gov/projects/tvpubs/tv6.papers/clips.pdf.

[8] V. Roth, J. Peters, and U. Pinsdorf. A Distributed Content-Based Search Engine Based on Mobile Code and Web Service Technology. In *Scalable Computing: Practice and Experience*, vol. 7, no. 4,pp. 101-117, 2006.

[9] N. Hinds and C.V. Ravishankar. Managing Metadata for Distributed Information Servers. In *Proc. 31st Hawaii Int'l Conf. System Sciences*, IEEE CS Press, 1998.

[10] Fuzzy queries in multimedia database systems. In PODS, 1998.

[11] M. Masud and I. Kiringa, "PDST: A peer database simulation tool for data sharing systems," in Proc. 1st Int. Conf. Simulation Tools, 2008, pp. 39:1–39:6.

[12] L. Serafini, F. Giunchiglia, J. Molopoulos, and P. Bernstein, "Local relational model: A logical formalization of database coordination," in *Proc.4th Int. Conf. Modeling Using Context*, 2003, pp. 286–299.

[13] J. Kim, D. D. Feng, T. W. Cai, and S. Eberl, "Integrated multimedia medical data agent in e-Health," in Proc. Pan-Sydney Area Workshop Visual Inf. Process., Sydney, 2001, pp. 11–15.

[14] S. Laborie, A. Manzat, and F. Sedes, "Managing and querying distributed multimedia metadata," *IEEE Multimedia*, vol. 16, no. 4, pp. 12–21, Oct. 2009.

[15] V. Papadimos and D. Maier, "Mutant query plans," *J. Inf. Software Technol.*, vol 44, no. 4, pp. 197–206, 2002.

[16] M. Masud, M. Shamim Hossain, A. Alamri, "Data Interoperability and Multimedia Content Management in e-Health Systems. *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 6, pp. 1015-1023, 2012.