IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-9, Sept. 2018 E-ISSN: 2347-2693

Information Retrieval Mechanism for Dynamic Health Care

C. Dharma Devi[1*], S. Thaddeus[2]

[1]PG & Research Dept. of Computer Applications, Sacred Heart College, Tirupattur, Vellore District, Tamil Nadu, INDIA
[2]Don Bosco College, Yelagiri Hills,Vellore District, Tamil Nadu, INDIA

[*]*Corresponding Author: udharmadevi@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 17/Sept/2018, Published: 30/Sept/2018

Abstract- In spite of the commendable success of Big Data, there are still many challenges with regard to storage, retrieval, analysis and prediction. A framework was proposed for information retrieval by integrating ontology models with Big Data. Evaluation of this framework with different domain models is a invariable work to validate this model. This study applies the framework to health care data and thus identifies the strengths and weaknesses of the framework.

Keywords- MongoDB, Ontology

I. Introduction

Big data analytics offers support in many business sectors, and health care is looking at big data to provide solutions too many disease related issues, age-wise issues and long-term disease and patient management. Health care is also one of the crucial service render by different organizations having more impact in the man power of a country. Advances in health care and technology in recent decades have expanded the attributes of the traditional medical services and have merged technologies such as networks, GIS, artificial intelligence, system engineering, simulation and knowledge management. This scenario has created new opportunities for building applications to provide medical services that are highly dynamic, diverse and efficient. Though the application of such technologies improves the response time in data retrieval and supports quick and effective decision-making, effective information retrieval in very large scale data with great accuracy and speed is a challenge.

The recent researches on integrating Ontology with Big Data [6] in various domains depicts that bridging ontology improves performance in data retrieval. This research work is to implement the framework [1, 5] in a real time medical data to show the enhancement in the data retrieval mechanism.

Section 2 highlights the features of document oriented NoSQL database MongoDB which is suitable for Medical data. Section 3 briefs the benefits of ontology. Section 4, explains the framework for enhanced data retrieval [1, 5] which has been proposed and experimented in the earlier research work. Section 5 illustrates how the process model integration with ontology improves performance through a real time medical data.

II. Document oriented NoSQL Database

MongoDB [2, 3,], a document based database in which each collection holds different documents represented in the form of JSON files. The unique features of MongoDB, being schema less, scalability, variable document length, capability to store both structure and unstructured data made it easily adopted and best suitable to maintain health care records. [3] Especially storing and retrieving semi-structured and unstructured health records are quite easy and flexible in MongoDB.

III. Ontology for Health Care

Ontologies [4, 6] are taxonomies or standard representation of knowledge and data used to build a common vocabulary for the domain, mapping of concepts or integration of data sources. Researches are creating number of Ontologies for various domains. The main benefit of creating and integrating ontology in health care is to provide a standard representation for structured and unstructured patient EHR (Electronic Health Records) and diagnosis results in the form of RDF (Resource Description Framework) with the ability of reprocessing the existing knowledge through ontologies (OWL). [7, 8] Using RDF like OBO (Open Biomedical Ontologies), semEHR (semantic Health Record) ontology for health care improves the retrieval, analysis and prediction on patient and diseases which has been clinically proven.

IV. A Road Map of Enhanced Information Retrieval Framework

It is identified from the previous research work of applying and testing the Geographical data in the framework [1, 5] that integrating ontology with Big Data enhances its

retrieval. Referring the results of the previous works the ontology bridge with MongoDB is now applied for efficient Medical data retrieval shown in the Fig. 4.1.

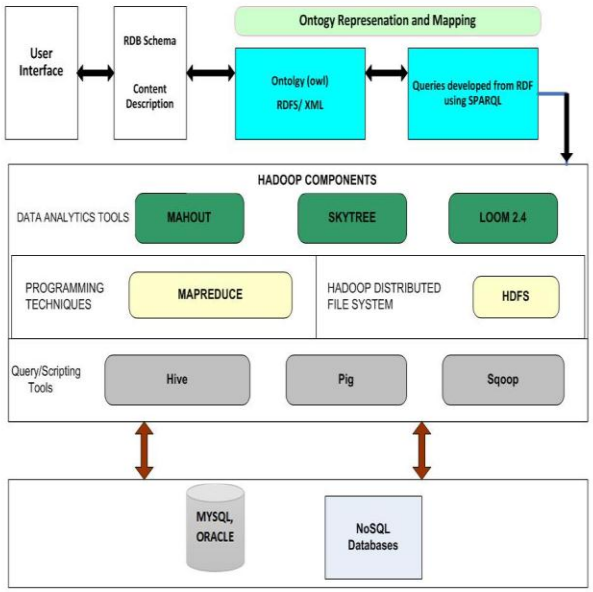

Fig. 4.1 A Model for Ontology Integration with Big Data

The retrieval model shown in Fig. 4.1 is a high-level architecture depicts an enhanced mechanism for information retrieval and big data analytics. Different elements in the model are explained below:

Ontology Representation
Ontology Representation is the process of creation and representation of ontology for a domain. The ontologies (owl files) are created using the existing tools like protégé.

Ontology Mapping
It is the process of representation of developed ontologies as RDF / RDFS Triple Stores / XML. Frameworks like HAIL, Jena can be used for automatic conversion of ontologies into RDFs and stores it into HDFS.

Relational data and NoSQL data
If the existing massive data is in the form of relations they are imported and converted as NoSQL data. In the process of ontology mapping the data can be renovated into RDF triple stores.

Information Retrieval
This process involves Hadoop components (HDFS, Map Reduce) and NoSQL database like MongoDB represents in the Fig. 4.1. From the RDF triple stores generated, we can develop any type of queries using SPARQL like query language. The queries are stored in the distributed file system HDFS. Then the parallel processing engine Map Reduce reads the queries, create the tasks, execute the tasks, fetch values from databases and produce the results very fast.

V. Evaluation of Process model

A sample of structured medical data stored using MongoDB has been taken. A Medical Ontology has been derived using protégé tool. A few set of MongoDB queries and SPARQL queries has been created to verify the factors like time consumption and memory space. For the same sample of medical records, information are retrieved using MongoDB and the results with time of execution, memory space used were recorded. Then, to retrieve the same information SPARQL queries are executed over the medical ontology developed and results were logged.

The graphical representation of medical records retrieval is shown in Table 5.1 and Fig. 5.2.

TABLE 5.1 Evaluation table for MongoDB Query in Hadoop and SPARQL in GraphDB

#	Query	Execution Time in ms	Memory Space in MB
MongoDB Query executed in Hadoop vs SPARQL query executed in GraphDB	db.pat_ind_seg.find({patient_name : /^R. +i })	94	0.633
	Select ?patient_name ?age_year where { ?s rdf:type patient: filter (regex(? patient_name, "^R","i"))}	36	0.312
	db.pat_idn_seg.find({gender : 2, {AGE_YEAR : {$gt : 50}})	2	0.633
	Select ?gender ?patient_name where { ?s rdf:type gender=2 and age_year > 50 } order by ?Gender	0.08	0.312
	db.orders.group({ key:{bed_no: b1}, cond:{adm_dt: { $gt: new Date('01/01/2012')} }, reduce: function (curr, result) {}, initial: { } })	46	0.514
	Select (COUNT (?Bed_num) as ?Total)	21	0.218

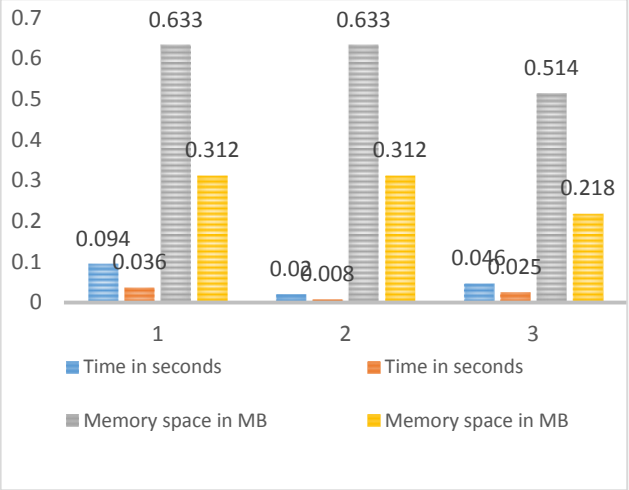

Fig. 5.2. Graphical Representation of MongoDB in Hadoop vs SPARQL in GraphDB

From the table 5.1 and the graph 5.2, it is identified that both time of execution and memory space occupied for big data retrieval integrating with ontology is better than the data retrieval using existing NoSQL tools.

Conclusion

Information retrieval in big data is good when using parallel processing tools like HADOOP and HAIL. But when integrating with ontology, it shows that the performance is better. Moreover, a sample of patient medical records of a hospital is used in this work for testing and visualized the improvement in data retrieval. The same ontologies can be used to integrate few more hospitals to have a comparative analysis and prediction in a minimal cost.

References

[1] Fr. Thaddeus, C. DharmaDevi, S. Deepa, "A Framework for ontology-driven big data applications", IJAER, Vol.10 (No.79), ISSN 0973-4562, 2015.
[2] Karl Seguin, "The Little MongoDB for 2.6".
[3] Chieh Ming Wu, "Comparisons between MongoDB and MS-SQL Databases on the TWC Website".
[4] Natalya F.Noy and Deborah L. McGuinness, "Ontology Development".
[5] Fr. Thaddeus, C. DharmaDevi, "Ontology Bridge for Information Retrieval", in the Proceedings of International Conference on Computing Paradigms, 2016.
[6] Furkh Zeshan and Radziah Mohamada, "Medical Ontology in the Dynamic Healthcare Environment", in Elsevier Ltd, 1877-0508, 2012.
[7] Sreekanth R, Golajapu Venu Madhava Rao, Srinivas Nanduri, "Big Data Electronic Health Records Data Management and Analysis on Cloud with MongoDB: A NoSQL Database", in IJAGT, 2309-4893, 2015.
[8] Manju K.K, Srinitya G, "Analysis and Prognosis of Cancer with Big Data Analytics", in IJRASET, 2321-9653, 2016.
[9] Muthulakshmi P and Udhayapriya S, "A Survey on Big Data Issues and Challenges" in IJCSC, Vol.6 (Issue 6), pp.1238-1244, 2018.
[10] Rabie A. Ramadan, "Big Data Tools-An Overview" in IJCSE, Vol.2 (125), 2456-4451, 2017.

