

SEMANTIC WEB BASED INFERENCE MODEL FOR LARGE SCALE ONTOLOGIES FROM BIG DATA

J.RAVI RAJESH

PG Scholar

Rajalakshmi engineering college

Thandalam, Chennai.

ravirajesh.j.2013.mecse@rajalakshmi.edu.in

Mrs. ROXANNA SAMUEL

Assistant professor,

Rajalakshmi engineering college

Thandalam, Chennai.

roxanna.samuel@rajalakshmi.edu.in

Abstract- Hadoop is an open source file system that can have a framework is processed over the big data. The fast growth of ontologies nowadays that can grow significantly performs normally and also some major issues in the efficiency and scalability reasoning methods. The traditional and centralized reasoning methods do not handle large ontologies. The system proposed a large scale ontologies for healthcare is applied to use map reduce and hadoop framework. The system also proposed a transfer inference forest and effective assertional triples for reduce the storage for reasoning methods and also simplified and accelerate. The Ontology Web Language which provides the semantic web access to all the relationships maintained by the syntaxes, specifications and expressions. These Ontology web based algorithm is used for maintain the strong relationships for the prefix names of subject and object of the large resource description framework that also process to knowledge base. The prototype system is implemented on a hadoop framework and the large scale ontologies are to provide the deduplication for the large datasets. A real world application of the healthcare is to validate the effectiveness of these large datasets. The experimental results show the high performance reasoning and runtime searching in the big data.

Keywords: Semantic Web, Transfer Inference Forest, Ontology bases, Effective Assertional Triple, RDF, MapReduce.

INTRODUCTION

Semantic Web is the large volume of the information that contains the information in the web used to describe the shared resources. The conversion of all the information is converted to the current web that is split into semi-structured and unstructured

documents into the web of these data. The technology of the semantic web is to create the data in the web, design a structure for resource description framework and specify vocabularies and set the rules for handling the data. The ontology web language algorithm is implemented in this concept to represent and handle large scale ontologies with the representation of transformation. Healthcare is mainly having the social potential that can be explored in semantic web. The inference model is used to represent the large datasets that can execute a fast process in the online queries. Existing reasoning methods can take too much of time to handle large datasets. Resource description framework is the representation of ontologies that can be describing the knowledge in the semantic web. In the resource description framework two functions are used to reduce the storage and efficient reasoning process that are transfer inference forest (TIF) and effective assertional triple (EAT). The relationship between the new triples is updated by these two methods to describe subject, predicate and object by implementing OWL algorithm.

MapReduce is the process for handling large datasets in the hadoop environment. MapReduce is useful in a wide range of applications, including distributed pattern-based searching, distributed sorting, web link-graph reversal, and term-vector per host, web access log stats, inverted index construction, document clustering, machine learning, and statistical machine translation. The MapReduce model has been adapted to several computing environments like multi-core and many-core systems, desktop grids, volunteer computing environments, dynamic cloud environments, and mobile environment. The various applications of healthcare and life sciences are applied in the technologies which retrieve a large datasets contain a semantic

web. A real world application of healthcare is to validate the efficient results to retrieve the large datasets using big data. By using these transference inference forest and effective Assertional triple methods time consuming is efficient and produce the efficient output functions for large scale ontologies. A real time data is to be recognized and implement the ontology web language algorithm is used to describe the efficiency of semantic web information. A novel representation of all the ontologies which can gives the efficient results from the queries. Every datasets contains relevant information that represents a ontology bases in a reasoning time with efficient retrieval from the queries. A Hadoop cluster is too configured in the computing nodes which are in the distributed computing. There are some eight cluster nodes that can execute a large datasets perhaps in the computing nodes performs an efficient query results. A billion triple challenges are to evaluate the data with a efficient results shows the performance in the experiments through semantic web. A knowledge base have to search the process of queries with patient, illnesses and drugs which gives the real time data to gives efficient results. An inference model is to represent the medical ontologies that give an efficient reasoning time and speed which is compatible with a challenging big data.

RELATED WORK

An Inference model is a method that directly integrated into the distributed method that affects the datasets in the knowledge bases. Semantic inferences that can be deriving from all the datasets which is to validate a small amount of ontologies is handled with a complex of an inference model. In the existing method of the knowledge bases have an incorrect values or tasks that execute the queries in a short time interval for updating queries. An incremental model is to identify the deriving inferences which represent the large scale ontologies in the web. The data volume is increased in the entire datasets the ontology bases is to fails to describe the resource description framework for matching the datasets with inaccurate values or produce duplicate files that collect from the web. There are many semantic inferences engines are available in the semantic web. Moreover many duplicate files are retrieved and updated on the inference model which deals with

large ontologies because of taking too much of time to show the results. The online queries is also affected by this duplicated files which retrieve from the current web and also make it to difficult to response from the users. In the statistical inference model the resource description framework is deal with a noisy data. These inference which fails to update the current information from the web and produce only existing one because of noisy inferences available in the semantic web. Semantic web is a representation in World Wide Web consortium also extends in the terms of the information and services. RDF contains the representation of ontology bases that collects from knowledge bases. A temporal extension of the time dependent information is retrieved from the ontology web language. A single machine or a local server dealing with large datasets that have to show the inferences in the reasoning time depends upon the real values. RDF can directly integrated with the ontology formation that represents the link changed into an inference model separates the data into a transformation files. An incremental and inference model is validating the large datasets with a single machine and it also integrated the ontology files from semantic web that updates the formation of resource description framework through a search engine. A large amount of RDF datasets have a reasoning time for its storage space is high to assured the otology bases that can adopt in the low performance of storage and also fails to response at the demand query for the online users. Semantic web data is a basic representation of all the information in the current web retrieves from the search engine process. It is also known as triples. An inference engine which builds the references for the triples is to identify the duplicate data available in the current web for endorsement in a schema representation. Semantic web also identify the duplicate inferences available in the syntaxes and specifications in the resource description framework. RDF must contain the rule set in the large datasets that can be represented by the size of the data in the load balancing problems available in the semantic web. An incremental method which finds the duplicate inferences for establishing the RDF datasets to transform the rule set into an original datasets which causes the low time complexities. WebPIE is a Web-scale Parallel Inference Engine also encodes the process of

semantic web reasoning into integrated sets of a large datasets that can be handled by two main operations ie. Map and Reduce process in big data. WebPIE is the process of identify the reasoning time for the datasets is mapped into rule set for optimizing the resources in semantic web. It also executed in a single machine or a group of cluster that runs parallel in distributed computing nodes. Duplicate inferences cause the load balancing problem and these datasets create a l group for every statement and also reduce the reasoning time in the iterations for rule set. WebPIE also remove the duplicate inferences and optimize the load balancing problems in the occurrence of an incremental and distributed inference method (IDIM). It also makes the inference model and datasets is transformed quickly to the query process with efficiency and reasoning time is compatible with highly reduce the storage requirements in the reasoning process.

PROPOSED SYSTEM

In the proposed system semantic web contains all the information is retrieved from the current web is used to represent the ontologies which is available in the search process. An inference model is provide the deduplication for the ontology bases is used for reduce the reasoning time for updating the transfer inference forest and effective Assertional triple method to represent knowledge bases. Resource description framework is used to describe the representation of ontologies that can transform the format for all the schemas that describe the specification and attributes. For handling large scale ontologies ontology web language algorithm is used to specify its attributes and functions for the RDF formation through ontology bases. Semantic web encodes all the relevant information available in the resources for describing the attributes and specifications in the ontology bases for updating TIF/EAT method. The representation of schemas is assigned the functions that can transform the RDF to a attributes and specifications that available in the semantic web. TIF/EAT provides the triples for expressed in statements and have a computation with a relationship between the newly updated triples with efficient formation. An inference model is a basic representation for specify attributes and integrated by the ontology bases for the removal of duplicate

inferences. By using transfer inference forest (TIF) and effective Assertional triple (EAT) reduce the storage requirements and simplify the reasoning time in the knowledge bases for handling large scale ontologies. A real world application of healthcare is used for large datasets like patient details, disease names, illnesses, drugs and side effects for validating all the details in a database.

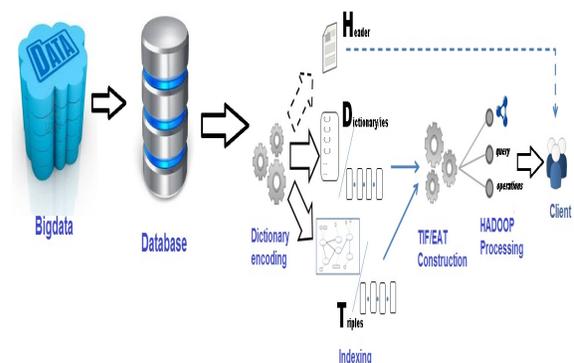


Fig 1. Proposed System Architecture

The resource description framework is used for a representation of the ontology bases that may cause a deduplication process for updating the triples in a formation. There are three main methods which specify the ontologies i.e. Subject, predicate and object. The computation of the resources is executed in parallel with eight cluster nodes or local server and single machine. In the ontology bases dealing with an assertional triple that can be expressed in statements for validating the healthcare or medical databases integrated with a single machine or eight cluster nodes.

Figure 1 shows semantic web is the process of describing current web content is access to the resource description framework. Transfer inference forest and effective Assertional triple is generated for reduce the storage space requirements and simplify the reasoning time. The computation of the cluster node and single server is communicating efficient process by implementing hadoop processing.

Figure 2 shows that collection of large datasets that is directly sent to the RDF datasets for transformation in the ontology bases. The computation of TIF/EAT needs to be updated by triples and performs an efficient process. These RDF datasets is integrated into deduplication inferences.

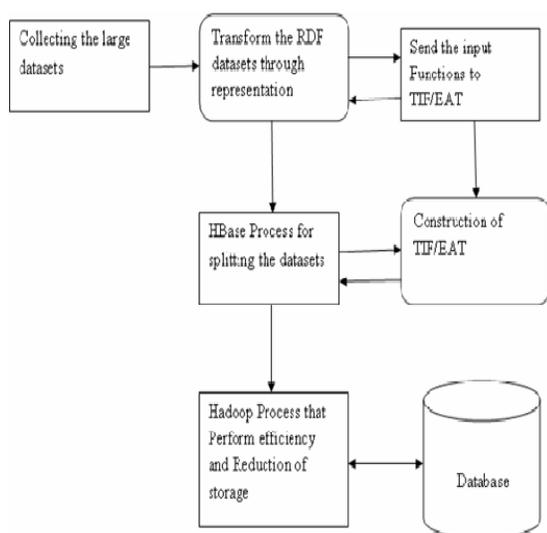


Fig 2. Flow diagram for the large datasets

The construction of transfer inference forest and effective Assertional triple which updates the new triples for simplify the reasoning time and process. The computation of a single machine or local server and eight cluster nodes deals with the big data for performing the real time datasets using hadoop functions.

ONTOLOGY WEB LANGUAGE ALGORITHM:

- a. The ontology web language algorithm provides the description of ontologies with attributes and specifications.
- b. OWL extends the resource description framework and semantics for the formation of XML and rule set for specifying standards.
- c. The standards of OWL which represent RDF document are to specify taxonomies and denote concepts of subject predicate and object to represent syntaxes and properties.
- d. The specifications of the ontology bases that referred the concepts which insisted o three main functions are subject, object and predicate in the large datasets. The standards

are represent the relations through properties.

MODULES

- Large data collection
- Transform the RDF formation
- Construction of TIF/EAT
- Hbase Processing
- Hadoop Processing

Large Data Collection

In this module, collection of a large datasets is very large that makes a challenging in the traditional applications. These large datasets is collected from various applications in the real time datasets. There are so many large datasets which is collected from the knowledge representation. In the healthcare domain also collect a large datasets with millions of files are running in every countries. These datasets are maintained by security and provides a high transmission range for communicating the system.

Transform the RDF formation

In this module semantic web is provide the resource description framework from the content of all the ontology bases that representing the schemas. RDF is a basic representation of the ontologies that can specify the semantics and syntaxes for the knowledge bases through schemas. The updated triples are expressed with the statements and attributes that define the taxonomies with the specifications of ontologies.

TIF/EAT Construction

Transfer Inference forest and Effective Assertional Triple is the process of constructing an RDF formation of specifying to functions are Domain and range. This subset of domain and range specify the attributes which convert the information available in the medical datasets like patient details disease name, illnesses and drugs which specify its value to the range and domain. Forest is the single or multiple trees which generate the relations that is linked with property of the RDF subsets.

Hbase Processing

Hbase is a distributed and a scalable process of integration of the data store in the hadoop. An

Hbase function provides a real time data that can be access and generates the read and write access to very large tables. It also performs the commodity hardware with cluster nodes and generates six tables. Hbase also store the intermediate results from the real time data and return the results from the queries.

Hadoop Processing

Hadoop is open source software that is integrated of the distributed processing and compatible with large datasets. It also provides a framework for commodity hardware and a cluster nodes access from the large datasets. Hadoop also designed with the scale up of single sever to thousand of machines with a very high degree of fault tolerance. The reliability of software and the process for detection of failure in hardware can be easily found in the hadoop environment. By implementing the hadoop is to run the process in linux and performs a main function is MapReduce concept is performed for efficiency in a real time data.

CONCLUSION

In the big data reasoning process in a web scale is integrated into a large volume of data has increasingly very complex in the semantic web. The construction of transfer inference forest and effective Assertional triple is effectively design the concept of reasoning time that is updated by the new triples for each ontology bases. Resource description framework is a schematic representation of the large scale ontologies specifies its attributes related to describe syntaxes and semantics. By using TIF/EAT the storage requirements is reduced and simplify the reasoning time in the medical datasets that response the online queries very quickly to end users. The computation of the queries is to validate all the benchmark in the billion triples challenge in the experiments. Handling large scale ontologies with a cluster of nodes by using hadoop environment.

REFERENCES

[1] M. S. Marshall et al., "Emerging practices for mapping and linking life sciences data using RDF—A case series," *J. Web Semantics*, vol. 14, pp. 2–13, Jul. 2012.
[2] M. J. Ibáñez, J. Fabra, P. Álvarez, and J. Ezpeleta, "Model checking analysis of semantically annotated business processes," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 42, no. 4, pp. 854–867, Jul. 2012.
[3] V. R. L. Shen, "Correctness in hierarchical knowledge-based requirements," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 30, no. 4, pp. 625–631, Aug. 2000.

[4] J. Guo, L. Xu, Z. Gong, C.-P. Che, and S. S. Chaudhry, "Semantic inference on heterogeneous e-marketplace activities," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 42, no. 2, pp. 316–330, Mar. 2012.
[5] J. Cheng, C. Liu, M. C. Zhou, Q. Zeng, and A. Ylä-Jääski, "Automatic composition of Semantic Web services based on fuzzy predicate Petri nets," *IEEE Trans. Autom. Sci. Eng.*, Nov. 2013, to be published.
[6] D. Kourtesis, J. M. Alvarez-Rodriguez, and I. Paraskakis, "Semantic based QoS management in cloud systems: Current status and future challenges," *Future Gener. Comput. Syst.*, vol. 32, pp. 307–323, Mar. 2014.
[7] "Linking Open Data on the Semantic Web" [Online]. Available: <http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/Datasets/Statistics>
[8] M. Nagy and M. Vargas-Vera, "Multiagent ontology mapping framework for the Semantic Web," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 4, pp. 693–704, Jul. 2011.
[9] J. Weaver and J. Hendler, "Parallel materialization of the finite RDFS closure for hundreds of millions of triples," in *Proc. ISWC*, Chantilly, VA, USA, 2009, pp. 682–697.
[10] J. Urbani, S. Kotoulas, J. Maassen, F. V. Harmelen, and H. Bal, "WebPIE: A web-scale parallel inference engine using mapreduce," *J. Web Semantics*, vol. 10, pp. 59–75, Jan. 2012.
[11] J. Urbani, S. Kotoulas, E. Oren, and F. Harmelen, "Scalable distributed reasoning using mapreduce," in *Proc. 8th Int. Semantic Web Conf.*, Chantilly, VA, USA, Oct. 2009, pp. 634–649.
[12] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, 2008.
[13] C. Anagnostopoulos and S. Hadjiefthymiades, "Advanced inference in situation-aware computing," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 39, no. 5, pp. 1108–1115, Sep. 2009.
[14] H. Paulheim and C. Bizer, "Type inference on noisy RDF data," in *Proc. ISWC*, Sydney, NSW, Australia, 2013, pp. 510–525.
[15] G. Antoniou and A. Bikakis, "DR-Prolog: A system for defeasible reasoning with rules and ontologies on the Semantic Web," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 2, pp. 233–245, Feb. 2007.