ISSN 2350 – 0328

International Journal of AdvancedResearch in Science, Engineering and Technology

Vol. 2, Issue 10, October 2015

A Survey on Semantic Inference Method Using Ontologies

A.Vijayalakshmi, Dr.S.Babu

P.G Student, Department of Computer Engineering, IFET College of Engineering,Villupuram,Tamil Nadu, India.

Associate Professor, Department of Computer Engineering, IFET College of Engineering, Villupuram, Tamil Nadu, India.

ABSTRACT: Semantic inference method attracts much attention of users from all fields. Many inference engines have been developed to support the reasoning over semantic web. With a large volume of Semantic Web data and their fast growth, diverse applications have emerged in a plurality of domains poses new challenges for ontology mapping. Ontology mapping can provide more correct results if the mapping process can deal with uncertainty effectively that is caused by the incomplete and inconsistent information used and produced by the mapping process. As it is evolving into a global knowledge-based framework, supporting knowledge searching over such a big and increasing dataset has become an important issue. A survey was made for different reasoning approaches that focus on semantic inferences. This paper describes about how the reasoning approaches process on users' queries.

KEYWORDS: Ontology reasoning, RDF, MapReduce, IDIM

I. INTRODUCTION

An Extension of Web is the Semantic Web [2]. The concept of the Semantic Network Model was formed in the 1960s to represent semantically structured knowledge. When applied in the context of the modern internet, it extends the network of hyperlinked human-readable web pages by inserting machine-readable metadata about pages and how they are related to each other. This enables agents to access the web more intelligently and perform more tasks on behalf of users.

Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise and business boundaries. Semantic Web data increases its volume by their fastest growth of ontology bases that brought significant challenges in performingefficient and scalable reasoning. Ontology is the "theory of existence" and an explicit specification of conceptualization. Ontology is a body of knowledge describing some domain, typically common sense knowledge domain. It is a model of reality of the world and the concepts in the ontology must reflect this reality.

Definition: A web of data that can be processed directly and indirectly by machines.

Semantic web [8] is regarded as an integrator across different content, information applications in publishing, blogging and many other areas. It was estimated to contain 4.4 billion triples in 2009 and has now reached over 20 billion triples. Its growth rate is still increasing. It takes the solution of publishing the data in ontological languages [10] like Resource Description Framework (RDF), Web Ontology Language (OWL) and Extensible Markup Language (XML). HTML describes documents and the links between them. The machine readable descriptions enable content managers to add meaning to the content i.e., to describe the structure of the knowledge about that content. By this way, machine can process knowledge itself, instead of text, using processes similar to human deductive reasoning and inference to obtain meaningful results.

II. SEMANTIC ONTOLOGY LANGUAGES

The semantic part is enabled by a stack of evolving languages:

ISSN 2350 – 0328

International Journal of AdvancedResearch in Science, Engineering and Technology

Vol. 2, Issue 10, October 2015

A. RDF

Resource Description Framework (RDF) is a basic representation of ontologies used to describe the knowledge in the semantic web. It integrates a variety of applications by using extensible markup language (XML) for syntax and universal resource identifier (URI) for naming. RDF is an assertional language intended to be used to express propositions via precise formal vocabularies. An RDF [1] data model is similar to classic conceptual modelling approaches, as it is based on the idea of making statements about resources. The fundamental unit of RDF is a triple, which describes a relationship between a subject and object. Its formal definition is <subject, predicate, object>, in which subject denotes a resource, and predicate denotes properties or aspects of the resource and expresses a relationship between the resource and the object. Some triples in conjunction with each other can give rise to new knowledge.

B. RDF Schema

RDFS extends RDF and it is a set of classes with certain properties in RDF. It offers a simple vocabulary, intended to structure RDF resources and axioms for object-oriented modelling. These resources can be saved in a triple store to reach them with the simple protocol and RDF query language (SPARQL). RDF closure is a way to realize an RDF query.

C. OWL

Web Ontology Language (OWL) [9], a family of knowledge representation language that adds more vocabulary for describing properties and classes: among others, relations between classes, cardinality, equality, richer typing ofproperties, characteristics of properties and enumerated classes.

III. METHODOLOGIES

A. Fuzzy set theory

Context awareness (CA) is a very important computing paradigm. Context is any information that can be used to characterize the situation of a person, place, or object that is considered relevant to the integration between a user and an application, including the user and the application themselves. CA is the ability of a system to sense, interpret, and react to changes in the environment a user is situated in. The capability of a context (or situation)-aware system [6] to *classify* context and *infer* specific situations can be facilitated by properknowledge-representation (KR) models. A Fuzzy-set-based model can accommodate the vagueness inherent in context capturing. A *fuzzy set* is used for representing imprecise context in a human understandable form. This methodology is generic and can be applied to different inference schemes in order to improve the inference capability of the classifier and deal with mutual-exclusion inference. This model generates specific complementary fuzzy rules used for increasing the accuracy of the classification process for the well-specified information in Semantic web. Applications can handle context as flexibly as their users would expect by using this method but it is not suitable for all situations of user.

B. RuleXPM

The RuleXPM (XML Product Map) [2] approach is an integrated model that combines a set of representations of various types of concepts, some e-marketplace participating systems, and an inference process. The method consists of several major constituents that include a collaborative ConexNet (Concept exchange Network), an e-marketplace network (EMpNet), and an inference engine.

ISSN 2350 – 0328

International Journal of AdvancedResearch in Science, Engineering and Technology

Vol. 2, Issue 10, October 2015

The RuleXPM architecture supports the concept separation strategy and makes the designed RuleXPM inference engine generic and suitable for use in all types of EMpNetparticipating systems. In this architecture, the inference engine is modular, i.e., each inference module is independent and reusable and the data in use can be dynamically generated, and is contextual.

The outgoing activity is semantically generated by separating the inputs as a multiphase forward-chaining inference, where different inputs are separated into external inputs requiring universal semantic understanding and internal inputs independent of external environment. The match–act cycle is built in several phases in the entire processing of the outgoing activity, namely, the RuleXPM inference procedure. This procedure is processed based on five types of resources namely XPMR document, XPMT library, stored execution rules (SERs), RuleXPM-based BPPs and knowledge base. Although this method is interoperable and inferred from one entity to another, it is not possible to implement it on an automated offering system and an automated negotiation system

C. Similarity transition

A linked dataset [12] is a kind of labelled directed graph cross domain, which is used for knowledge presentation and cognitive model foundation. Each link represents a kind of relationship between two resources and they can be represented as a statement in RDF. In these statements, since objects are the property value of subjects and describe their features, the similarity between two subjects can be calculated from the similarity between their corresponding sets of objects.

If the linked dataset is considered as a whole semantic graph, then the calculated similarity value between two subjects can be further transited to their own related subjects, which may make the similarity between the related subjects more accurate. This calculation is referred as similarity transition [7] that utilizes node and link types together with the topology of the semantic graph to derive a similarity graph from linked datasets. This method enables smooth interaction and visualization of the similarity graph which is derived based on the calculated similarity of two resources but effectiveness of this method is less as similarity weight of each link type is given by experience.

D. Distributed reasoning methods

The above described methods are applicable for small databases. To deal with large base, some researchers turn to distributed reasoning methods.

1) *Parallel Materialization:* Parallel Materializationof the Finite RDFS [14] is the first method to provide RDFS inference on such large data sets in such low times and scalable manner. This maintains soundness and completeness without requiring any cumbersome preparation of the data. This method increases the processing speed by means of parallel inference. It lacks with scalability and expressivity.

2) *Scalable distributed reasoning:*Scalable distributed reasoning[4]presents some non-trivial optimisations for encoding the RDFS ruleset in MapReduce and exploits the MapReduce framework for efficient large-scale Semantic Web reasoning and implements on the top of Hadoop. This reasoning technique performs quick reasoning using HDFS and high data correlation. The drawback of using this method is it does not focus on quality of reasoning.

3) MapResolve: MapResolve [11]solves the problem by adapting the standard method for distributed resolution that avoids repetition of resolution inferences. For the limited expressivity of RDFS, the repetition can be avoided because every MapReduce [5] job is executed only once. The clause sets are parsed and written to disc for each iteration, generating needless overhead.

ISSN 2350 – 0328

International Journal of AdvancedResearch in Science, Engineering and Technology

Vol. 2, Issue 10, October 2015

E. Incremental reasoning methods

However, the distributed reasoning methods considered no influence of increasing data volume and did not answer how to process users' queries. As the data volume increases and the ontology base are updated, these methods require the re-computation of the entire RDF closure every time when new data arrive. To avoid such time-consuming process, incremental reasoning methods are proposed.

1) *WebPIE*:WebPIEis a Web-scale Parallel Inference Engine [3] using MapReduce. This method calculates the RDF closure based on MapReduce for large-scale RDF dataset by adopting algorithms to process the statements based on input data as incremental reasoning. This technique identifies the accurate status which does either exist or new ones but it does not provide the relationship between the newly arrived and existing data.

2) *Incremental Ontology Reasoning:* Incremental Ontology reasoningapproach [13] based on modules that can reuse the information obtained from the previous versions of an ontology which is best suitable for OWL. Reasoning speed is a huge problem while using this method.

F. IDIM

An Incremental and Distributed inference method [15] is based on MapReduce and Hadoop. This method speeds up the updating process with newly-arrived data and fulfils the requirements of end-users for online queries that leverage the old and new data to minimize the updating time and reduce the reasoning time when facing big RDF datasets. Though this inference method speeds up the updating and reasoning, it lacks in number of clustering nodes.

IV. PROPOSED WORK

The existing incremental and distributed reasoning methods focus on computing RDF closure for reasoning which takes much time and space. Moreover, each time when new RDF arrives, full re-reasoning over the entire dataset is needed to compute the new RDF closure. WebPIE distinguishes newly-arrived RDF triples and old ones but fails to consider the relationship between them, thus resulting in a huge number of duplicated triples during the reasoning thereby hampering its performance. The proposed method overcomes the problem of mapping two ontologies effectively and efficiently which is a necessary precondition to integrate information on the semantic web by means one-class clustering tree technique.

V. CONCLUSION

A survey of different reasoning approaches that focus on semantic inferences was discussed that describes about how those approaches process on users' queries which was specific to certain domains. In order to improve the processing for more datasets, these methods can be made suitable to support for many ontology languages.

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] 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.

[3] 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.

ISSN 2350 – 0328

International Journal of AdvancedResearch in Science, Engineering and Technology

Vol. 2, Issue 10, October 2015

[4] J. Urbani, S. Kotoulas, E. Oren, and F. Harmelen, "Scalable distributed reasoning using mapreduce," in *Proc. 8th Int. Semantic Web Conf.*, Chantilly, VA, USA, pp. 634–649, Oct. 2009.
[5] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, 2008.
[6] 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.
[7] H. Paulheim and C. Bizer, "Type inference on noisy RDF data," in *Proc. ISWC*, Sydney, NSW, Australia, pp. 510–525, 2013.
[8] 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.
[9] V. Milea, F. Frasincar, and U. Kaymak, "tOWL: A temporal web ontology language," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*,vol. 42, no. 1, pp. 268–281, Feb. 2012.
[10] D. Lopez, J. M. Sempere, and P. García, "Inference of reversible tree languages," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*,vol. 34, no. 4, pp. 1658–1665, Aug. 2004.
[11] A. Schlicht and H. Stuckenschmidt, "MapResolve," in *Proc. 5th Int. Conf. RR*, Galway, Ireland, pp. 294–299, Aug. 2011.
[12] *Linking Open Data on the Semantic Web* [Online].Available:http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/DataSets/Statistics
[13] B. C. Grau, C. Halaschek-Wiener, and Y. Kazakov, "History matters: Incremental ontology reasoning using modules," in *Proc. ISWC/ASWC*, Busan, Korea, pp. 183–196, 2007.
[14] J. Weaver and J. Hendler, "Parallel materialization of the finite RDFS closure for hundreds of millions of triples," in Proc. ISWC, Chantilly, VA, USA, pp. 682–697, 2009.
[15] Bo Liu, *Member, IEEE*, Keman Huang, Jianqiang Li, and MengChu Zhou, "An Incremental and Distributed Inference Method for Large-Scale Ontologies Based on MapReduce Paradigm," *IEEE Trans. on cybernetics*, vol. 45, no. 1, january 2015.