International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
National Conference "NCPCI-2016", 19 March 2016
Available online at www.ijrat.org

Standardizing Data Exchange using a Benchmark in the Field of Semantic Web Ontologies

Prof. Snehal Rathi [1], Jyoti Shinde[2], Akshata Sonwane[3], Mohini Suryawanshi[4], Mital Shah[5]
Email: snehal.rathi@viit.ac.in [1], akshata025@gmail.com[2], mohini.suryawanshi33@gmail.com[3], jyotishinde333@gmail.com[4], mital.shah113@gmail.com[5]
[12345]Computer Department, Savitribai Phule Pune University

ABSTRACT – In today's world people used to search various types of information which should be knowledge based, appropriate, easily accessible. With the changing world requirements are changing, and different types of browsers came into the picture. The handling of back-end operations for some applications seems to be a very tedious job. This problem is solved by using some specific standards and properties. The data is raw information, where we need to process and integrate it properly. While doing all these things, we must consider their use of resources, the time required for operation. There are various techniques of data exchange and approaches such as rule based and machine learning. The approach, we are using is rule-based approach and it provides three real world problems and seven synthetic patterns to solve problems. The various benchmarks have been provided by implementation techniques. We are using MostoBM benchmark which is an integral part of rule-based approach. This provides us a better performance evaluation and less integration cost with manageable resources. This term is applicable in Scientific and Engineering Fields, Military Applications, Predicting Weather Conditions. The operations have been provided by easy way and better working for handling semantic web data.

INDEX TERMS: Benchmark, OWL(Ontology Web Language), Ontology, RDFS(Resource Description Framework Schema), MostoBM, Semantic Web.

1. INTRODUCTION

The Semantic web is an important term in the era of the internet. It provides a common base to work upon it and perform certain operations where data can be shared, reused and processed in distributed environments. Thus, this provides a way for data exchange which is applicable for various browsers. In this project, we get the abstract data from metadata. We are going to use MostoBM benchmark which has been used for rule-based approach. The term ontology is nothing but any real world fact, thing. SPARQL Query engines are used to get faster access. Some patterns are applied to work on real world problems. This concept is beneficial to save time, reduce the use of resources and their costs. It provides smooth handling of data extraction and data integration. It focuses on a variety of domains such as government, life sciences, geographic media or applications.

2. LITERATURE SURVEY

[3] M. Schmidt, T. Hornung, G. Lausen, and C. Pinkel, "SP2Bench: A SPARQL Performance Benchmark," Proc. Int'l Conf. Data Eng. (ICDE), pp. 222-233, 2009:

The Resource Description Framework (RDF) has become the important standard format to encode machine-readable information of the Semantic Web data. RDF databases can be represented by labeled directed graphs, where each edge connects as a subject node to an object node under label predicate. The object denotes the value of the subject's property predicates. Supplementary to RDF, the W3C has recommended the declarative SPARQL query language, which can be used to extract information from RDF graphs. Also the operators are used to performing relational joins, unions, left outer joins, selections, and projections. These can be combined to build more expressive queries. These approaches comprise a range of optimization techniques, including normal-forms, graph pattern reordering based on selectivity estimations (similar to relational join reordering), syntactical rewriting, specialized indices and storage schemes for RDF, and Semantic Query Optimization, graph matching.

[14]Benchmarking the Performance of Linked Data Translation Systems Carlos R. Rivero Andreas Schultz Freie Universität Berlin, Christian Bizer, Freie Universität Berlin, German, David Ruiz, University of Sevilla, Spain

The benchmark is developed to test the performance of two data translation systems, like Mosto, LDIF and compare the performance of the systems with the SPARQL 1.1 CONSTRUCT query performance of the Jena TDB RDF store. To solve the problems of heterogeneity, mappings are used to perform data translation, i.e., exchanging data from the source data set to the target dataset [18,19]. Data translation, a.k.a. data exchange, is a major research topic in the database community, and it has been studied for relational, nested relational, and XML data models [15,16,17].

3. PROPOSED SYSTEM

In this system we use a benchmark for testing data exchange systems using ontology mapping and SPARQL Query Engine. It provides a list of three real-world and seven synthetic data exchange patterns[2]; seven parameters to construct scenarios that are creation of the patterns[2]; and a publicly available tool that facilitates the creation of them and the collecting of data about the performance[2]of systems. Knowledge the evaluation methodology is a better choice to see the performance. The three real-world problems are relevant data exchange problems for processing the Linked Open Data.

4.ADVANTAGES OF PROPOSED SYSTEM

a) System presents a benchmarking strategy for testing data exchange systems to find relevant ontologies and mapping by query engines.

b) System provides an evaluation methodology that will allow us to use efficient data exchange system which will deliver better performance.

c) This is helpful to know the correspondence between the entities of ontology.

d) It is applied to a number of patterns, systems and after specific operation it gives ranking.

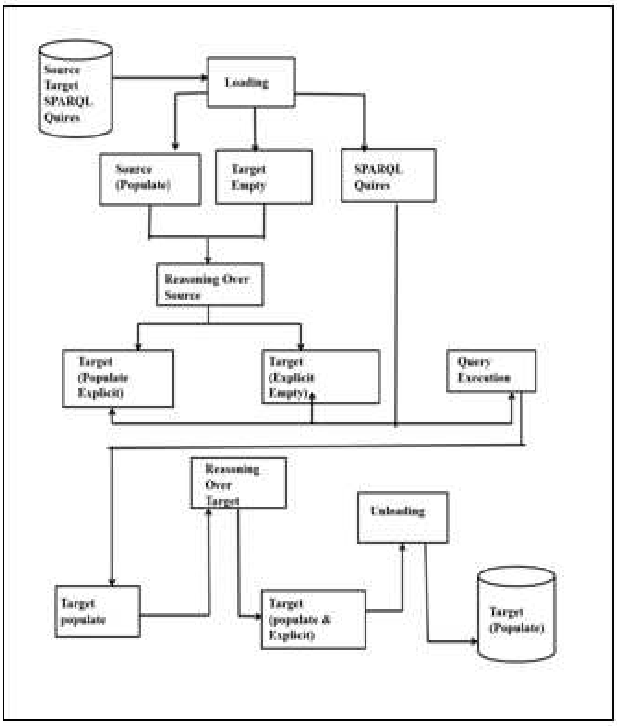

Figure: Proposed System Architecture [4]

5. SYSTEM MODULES

1. Data Exchange:

Data exchange is the process where data from source schema are transformed into target schema, so the target data is the actual representation of source data. Data exchange technique is used to perform the information exchange between the abstract data and a large amount of data by using metadata. If there is any problem in exchange method, it may affect on integration cost of the product.

2 . Data Selection:

In this feature, it will select the data within dataset which is loaded previously. Data selection is the process of identifying the appropriate form of the source. Data selection does the activity of collecting data. The process of selecting data may leave an impact on data integration.

3. Benchmark Real World Patterns:

he properties like synthetic pattern matching are used for computing various patterns and finding relationships among ontologies.

4. Scenario Execution:

In this feature, it will get the minimum number of scenarios. Execute those scenarios. Consider the proper sensitivity analysis and compute the new of them, manage if there is any variation in performance. Compute final decision.

6. THREE REAL WORLD PROBLEMS

Our benchmark provides three real-world data problems which are used to create the variety of scenarios using certain parameters. The problems are Evolution of ontology, Vocabulary adaptation, Publication of Linked Open Data.

Evolution of Ontology:

Usually, things change with the response to a certain need, including that the field of interest has changed, and the perspective under Large-scale Human Locations using Probabilistic Topic Models [11], which the domain is viewed needs to change, or due to design flaws in the original ontology. In this context, the source [10], ontology is the ontology before changes are applied and the target ontology is the ontology after changes are [10][11], applied. This pattern focuses on DBpedia, which is a community effort to annotate and make the data stored[11], at Wikipedia accessible using an ontology.

Vocabulary Adaptation:
It is not uncommon that two ontologies have the same data structured according to various vocabularies. In this context, it is necessary to get the vocabulary of a source ontology to the vocabulary of a target ontology.

Publication of Linked Open Data:
Many existing ontologies do not inherit to the principles of the Linked Open Data contents, and there is a need to deal with unstructured data.

7. SEVEN SYNTHETIC PATTERNS

Lift Properties:

These are used to extract common properties to a super class in a hierarchy. So, therefore, the data properties of a set of subclasses are moved to a common super class.

Sink Properties:

These Properties are used to narrow the domain of some properties. The data properties of a super-class are moved to some subclasses.

Extract Subclasses:

These Properties are used to specialize a class. Therefore, a source class is divided into several, subclasses and the domains of the target data properties are selected amongst the subclasses.

Extract Super-classes:

These properties are used to generalize a class. So, a class is split [10], into several super-classes, and data properties have distributed amongst them.

Extract Related Classes:

These properties are used for extraction of some classes built on a single class. Hence, the data properties have this single class as domain changes their domains as per new classes.

Simplify Specialization:
These properties are used to simplify the specialization of available classes.
Simplify Related Classes:
These properties are used to simplify the classes of entities which are related to specific ontology.
Improved factors:
1. The cost required for Integration is reduced by the use of proper benchmarks and properties.
2. The System gives better performance in the context of rule-based approach.
3. The processing is done according to specific standards.
4. It is applied some patterns and systems, allows ranking which system performs better and it is a Cost- efficient software.

8. IMPLEMENTATION

In this paper, we are using benchmark MostoBM. The implementation has done by considering three real world problems and seven synthetic patterns.

1. First of all, load the dataset and extract the information from it.
2. Fire query using SPARQL as internal operation and get the information.
3. Create knowledge from the information the information obtained from the query DBpedia.

4. Finding relations, mapping the patterns and use it for data exchange operation.
5. Show the graphical view after mapping various ontologies and their relations.
6. Finally, the data will be extracted using metadata which will be in an abstract data form.

Extract Ontology:

Retrieve Ontology:

Compute Patterns:

9. CONCLUSION

In this paper, we present a benchmarking strategy to test data exchange systems containing ontologies. There is common base provided for various integration problems based on current approaches in the context of ontology evolution. This list of patterns is not meant to be exhaustive: we need a community effort to extend them. These had created in the form of synthetic scenarios by using seven parameters that allow controlling the construction of both the structure and/or data of ontologies. The scaling of the patterns helps to analyse the performance of the systems when data exchange problems increase their scale in structure and/or data. Finally, our benchmark provides an evaluation methodology that allows comparing of the systems and make informed and statistically sound decisions about their performance. It has applied some patterns and systems which allows ranking of the system based on the performance.

10. ACKNOWLEDGEMENT

We would like to thank our internal guide Prof. Snehal Rathi for helping and guiding us . We are thankful to her for sharing new ideas and motivating us .

We are grateful to prof. shelke for important guidance. Both the guides helped us a lot. They consistently allowed this paper to be our own work, but showed us the right direction whenever they thought we needed it.

REFERENCES

[1] STBenchmark: Towards a Benchmark for Mapping Systems Bogdan Alexe
UC Santa Cruz abogdan@cs.ucsc.edu
Wang-Chiew Tan UC Santa Cruz
wctan@cs.ucsc.edu Yannis Velegrakis∗
University of Trento velgias@disi.unitn.eu.
[2] C.R. Rivero, D. Ruiz, and R. Corchuelo, "On Benchmarking Data Translation Systems for Semantic-Web Ontologies," technical report, Univ. ofSevilla,
http://tdgseville.info/Download.ashx?id=205, 2012.
[3] M. Schmidt, T. Hornung, G. Lausen, and C. Pinkel, "SP2Bench: A SPARQL Performance Benchmark," Proc. Int'l Conf. Data Eng. (ICDE), pp. 222-233, 2009.
[4]Prof. Snehal Rathi, Akshata Sonawane, Jyoti Shinde, Mital Shah, Mohini Suryawanshi
DATA-EXCHANGE -AMONG-SEMANTIC-WEB-ONTOLOGY-USING-BENCHMARK..pdf, ISSN (Online): 2277-6370 & ISSN (Print):2394-0921
[5] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, "DBpedia: A Crystallization Point for the Web of Data," J. Web Semantics, vol. 7, no. 3, pp. 154-165, 2009.
[6] B. Glimm, A. Hogan, M. Kro¨tzsch, and A. Polleres, "OWL: Yet to Arrive on the Web of Data?" Proc. Conf. Linked Data on the Web (LDOW), 2012.
[7] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, "DBpedia: A Crystallization Point for the Web of Data," J. Web Semantics, vol. 7, no. 3, pp. 154-165, 2009.
[8] M. Klusch, B. Fries, and K.P. Sycara, "OWLS-MX: A Hybrid Semantic Web Service Matchmaker for OWL-S Services," J. Web Semantics, vol. 7, no. 2, pp. 121-133, 2009.
[9] Efficient Extraction of Ontologies from the Semantic Web for Benchmarking Data Exchange
Smita U.Lokhande, BE Student, Computer, JSPM, NTC, Amruta S. Gogawale, Manoranjani A. Kulkarni, Sneha S. Satpute, G.P.Sagar, Pune, India
[10] Benchmarking Data Exchange among Semantic-Web Ontologies Carlos R. Rivero, Inma Hernandez ,David Ruiz, Rafael Corchuelo, University of Sevilla, http://doi.ieeecomputersociety.org/10.1109/TKDE.2012.175 ,Sevilla.
[11] DBpedia SPARQL Benchmark – Performance Assessment with Real Queries on Real Data Mohamed Morsey, Jens Lehmann, Sören Auer, and Axel-Cyrille Ngonga Ngomo
Department of Computer Science, University of Leipzig Johannisgasse 26, 04103 Leipzig, Germany.
[12] Schema Matching and Mapping. Springer, 2011.Z. Bellahsene, A. Bonifati, and E. Rahm, editors**.**
[13] SRBench: A Streaming RDF/SPARQL Benchmark Ying Zhang1, Pham Minh Duc1, Oscar Corcho2, and Jean-Paul Calbimonte2
[14] Benchmarking the Performance of Linked Data Translation Systems Carlos R. Rivero, Andreas Schultz Freie Universität Berlin, Christian Bizer,Freie Universität Berlin, Germany, David Ruiz, University of Sevilla, Spain.
[15] M. Arenas and L. Libkin. XML data exchange: Consistency and query answering. J. ACM, 55(2), 2008.
[16] R. Fagin, P. G. Kolaitis, R. J. Miller, and L. Popa.
Data exchange: semantics and query answering. Theor. Comput. Sci., 336(1):89–124, 2005.
[17] A. Fuxman, M. A. Hernandez, C. T. H. Ho, R. J. Miller, P. Papotti, and L. Popa. Nested mappings: Schema mapping reloaded. In VLDB, pages 67–78, 2006.
[18] C. R. Rivero, I Hernandez, D. Ruiz, and R. Corchuelo. Generating SPARQL executable mappings to integrate ontologies. In ER, pages 118–131, 2011.
[19] C. R. Rivero, I. Hernandez, D. Ruiz, and R. Corchuelo, On benchmarking data translation systems for semantic-web ontologies. In CIKM, pages 1613–1618, 2011.
[20] B. Glimm, A. Hogan, M. Kro¨tzsch, and A. Polleres, "OWL: Yet to Arrive on the Web of Data?" Proc. Conf. Linked Data on the Web (LDOW), 2012.