International Journal of Advance Engineering and Research Development (IJAERD)
Volume 1,Issue 5,May 2014, e-ISSN: 2348 - 4470 , print-ISSN:2348-6406

Performance Evaluation & Comparison of Inference Engines for Different Biomedical Ontologies

Ravi L Verma[1], *Prof. Rashmi Gupta*[2], *Prof. Arjun Rajput*[3]

[1]*Information Technology, Technocrats Institute of Technology*
[2] Computer Engineering, *Technocrats Institute of Technology, Excellence*
[3]*Computer Engineering, Technocrats Institute of Technology, Excellence Bhopal, MP*

Raviverma.vier@gmail.com[1],*rgupta6773@gmail.com*[2],*Rajarjun07@gmail.com*[3]

Abstract-- Semantic web can be defined as expansion of the existing web, which gives an easier way to search, share, combine and re-use information. Semantic web is intended to form a huge distributed knowledge based system. Focus of semantic web is to share data instead of documents. Based on Machine readable information & developed on XML technologies, that is capable of defining modified tagging schemas and RDF (Resource Description framework) flexible approach for representing data. For understanding and using the knowledge &data, which is encoded in semantic web documents, we require an inference engine, also called reasoners, can be defined as software applications that describe new facts from the existing information. Here different Bio-Medical ontologies are used to analyze the performance of various Inference engines. The paper focuses to analyze different parameters for number of inference engines and generate statistic based on suitability of inference engine with respect to particular domain.

I. INTRODUCTION

Semantic web is a web of actionable information, information derived from data through a semantic theory for interpreting symbols. Semantic web provides an account of “Meaning” in which the logical connection of terms establishes interoperability between systems. An idea of having data on the web defined and linked in such a way that it can be used by machines not just for displaying purpose but that can also be used for automatic integration and reuse of data across various applications. Semantic computing services visions to provide information & services in intelligent manner with more coherent interfaces that helps user to be served, anywhere, anytime with any device, without interference of the existing user. Correctness and quality of ontologies play an important role in semantic representation and sharing of knowledge. To confirm the quality of ontologies, a need arises for dealing with the uncertainty and inconsistency in the ontologies of the real-world applications. In consistent ontology means there exists an error or conflict in ontology and so some concepts in the ontology cannot be interpreted correctly. Inconsistency will result in false semanticunderstanding and knowledge representation. Reasoners (Inference engines) are used, as they reduce the redundancy of information in knowledge base and find the conflicts in knowledge content. Hidden knowledge can be extracted

by using reasoning with rules that are useful and relevant information within open web environment.

Keywords: Inference engine, Classification time, OWL DL, Performance.

II. INFERENCE ENGINES

A. Racerpro: Racerpro formally known as Racer is a reasoner for OWL DL [7]. It supports the optimization techniques of SWRL-IQ (Semantic Web Rule Language Inference Query Tool) is a Plugin for Protégé 4.2 that allows the user to edit, save & submit queries to an underlying inference engine based on XSB prolog.

B. Fact++: Fact++ is an extended version of FACT, which works on tableaux algorithm, for SROIQ description logic and is implemented in C++, but has very limited user interface and services as compare to other reasoners. Follows strategies like absorption, synonym replacement, model merging, told cycle elimination and taxonomic classification.

C. Hermit: Description logic reasoning system. Hermit is a reasoner for ontologies that are written using OWL (Web Ontology Language). For a given OWL, the reasoner Hermit can determine whether or not the ontology is consistent, identify the relationship between classes. Hermit is an OWL reasoner which works on Hypertableau calcus algorithm, which provides more efficient reasoning as compared to previously used algorithm. Ontologies that required minutes and hours to be classified now can be classified in seconds by Hermit reasoner.

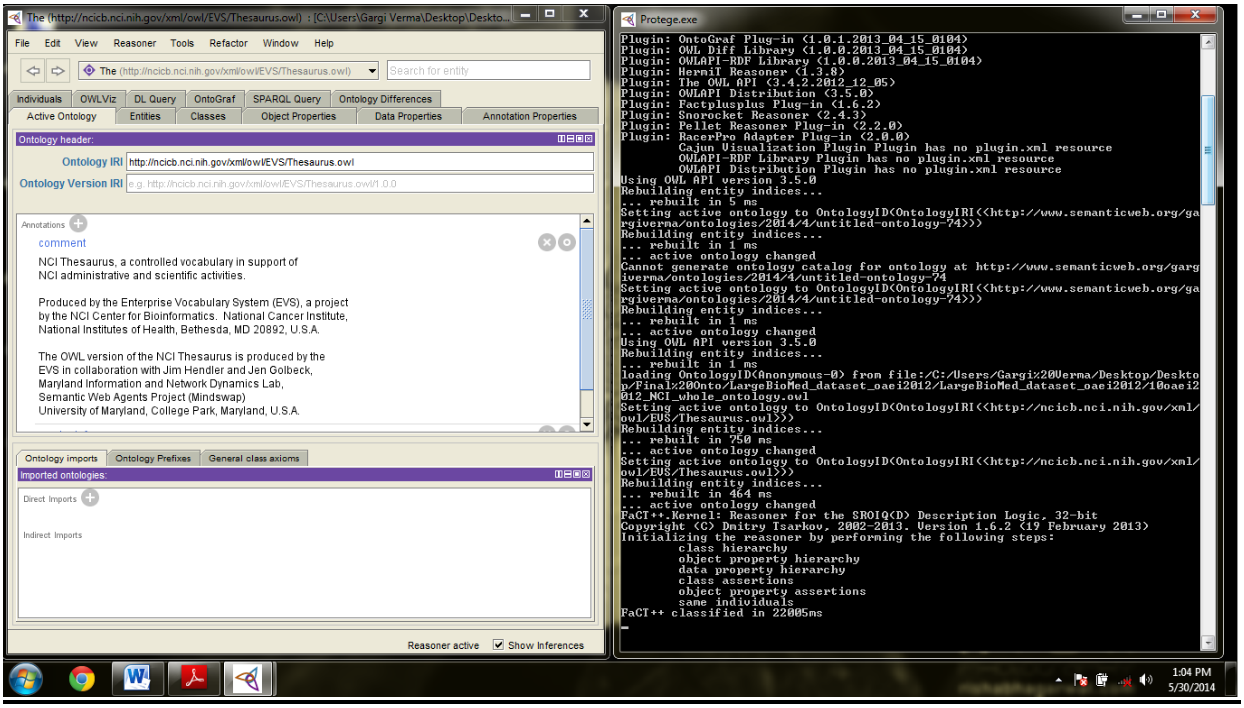
D. Pellet: Pellet is an open source Java Owl Dl reasoner [6]. It supports expressivity of SROIQ (D).Supports SWRL rules. Applies Reasoning on SHIN (D) with strategies like TBOX portioning, semantic branching, absorption and dependency directed back jumping. Individuals reasoning, Data type reasoning and optimization in ABOX query answering makes it more suitable for semantic web applications. Includes number of optimization techniques like optimization for nominal, incremental reasoning & conjunctive query answering.

III. PERFORMANCE EVALUATION

A. About Biomedical Ontologies & Performance evaluation:

FMA (Foundational Model of Anatomy), NCI (National Cancer Institute Thesaurus) are semantically very large ontologies and contains tens of thousands of classes. Existing matching tools can efficiently deal with moderately sized ontologies, but large-scale ontologies such as FMA, SNOMED and NCI are still very difficult to classify and obtain their results. In this paper we have taken around 10 biomedical ontologies, some of them are mixed & some are taken as whole. We have classified these ontologies with 4 different inference engines & their classifications time have been recorded, in order to reach to a conclusion that which inference engine is more capable for classifying the large biomedical ontologies & so based on that inference engine can be ranked and can be used for future enhancements of other large size ontologies with respect to a specific domain.

This test were executed on a Intel(R) core(TM) i3 CPU M380 @2.53 GHz,with 4 GB RAM, running on Windows 7 Ultimate, Java SE 1.6, Protege_4.1.we have focused on the classification time taken by various inference engines for classifying same sized ontologies. Comparing all the inference engines, we concluded that among all the inference engines FACT++ has the least classification time for almost all ontologies.We have therefore analyzed inference engines with respect to different sized biomedical ontologies. In fig 1, we have shown ontology running in Protégé 4.1 Tool for calculating the classification time.



Ontology IRI: http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl

comment

NCI Thesaurus, a controlled vocabulary in support of NCI administrative and scientific activities.

Produced by the Enterprise Vocabulary System (EVS), a project by the NCI Center for Bioinformatics. National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, U.S.A.

The OWL version of the NCI Thesaurus is produced by the EVS in collaboration with Jim Hendler and Jen Golbeck, Maryland Information and Network Dynamics Lab, Semantic Web Agents Project (Mindswap) University of Maryland, College Park, Maryland, U.S.A.

```
Plugin: OntoGraf Plug-in (1.0.1.2013_04_15_0104)
Plugin: OWL Diff Library (1.0.0.2013_04_15_0104)
Plugin: OWLAPI-RDF Library (1.0.0.2013_04_15_0104)
Plugin: HermiT Reasoner (1.3.8)
Plugin: The OWL API (3.4.2.2012_12_05)
Plugin: OWLAPI Distribution (3.5.0)
Plugin: Factplusplus Plug-in (1.6.2)
Plugin: Snorocket Reasoner (2.4.3)
Plugin: Pellet Reasoner Plug-in (2.2.0)
Plugin: RacerPro Adapter Plug-in (2.0.0)
        Cajun Visualization Plugin Plugin has no plugin.xml resource
        OWLAPI-RDF Library Plugin has no plugin.xml resource
        OWLAPI Distribution Plugin has no plugin.xml resource
Using OWL API version 3.5.0
Rebuilding entity indices...
... rebuilt in 5 ms
Setting active ontology to OntologyID(OntologyIRI(<http://www.semanticweb.org/ga
rgiverma/ontologies/2014/4/untitled-ontology-74>))
Rebuilding entity indices...
... rebuilt in 1 ms
... active ontology changed
Cannot generate ontology catalog for ontology at http://www.semanticweb.org/garg
iverma/ontologies/2014/4/untitled-ontology-74
Setting active ontology to OntologyID(OntologyIRI(<http://www.semanticweb.org/ga
rgiverma/ontologies/2014/4/untitled-ontology-74>))
Rebuilding entity indices...
... rebuilt in 1 ms
... active ontology changed
Using OWL API version 3.5.0
Rebuilding entity indices...
... rebuilt in 1 ms
loading OntologyID(Anonymous-0) from file:/C:/Users/Gargi%20Verma/Desktop/Deskto
p/Final%20Onto/LargeBioMed_dataset_oaei2012/LargeBioMed_dataset_oaei2012/10oaei2
012_NCI_whole_ontology.owl
Setting active ontology to OntologyID(OntologyIRI(<http://ncicb.nci.nih.gov/xml/
owl/EVS/Thesaurus.owl>))
Rebuilding entity indices...
... rebuilt in 750 ms
... active ontology changed
Setting active ontology to OntologyID(OntologyIRI(<http://ncicb.nci.nih.gov/xml/
owl/EVS/Thesaurus.owl>))
Rebuilding entity indices...
... rebuilt in 464 ms
... active ontology changed
FaCT++.Kernel: Reasoner for the SROIQ(D) Description Logic, 32-bit
Copyright (C) Dmitry Tsarkov, 2002-2013. Version 1.6.2 (19 February 2013)
Initializing the reasoner by performing the following steps:
        class hierarchy
        object property hierarchy
        data property hierarchy
        class assertions
        object property assertions
        same individuals
FaCT++ classified in 22005ms
_
```

Fig:1 Protégé GUI.

Output of Classified Time for Different Biomedical ontologies with respect to various Inference Engines.

Inference Engine	Ontology	Domain	Data Size (MB)	Description Logic.	Classified Time.

FACT++	1) FMA extended overlapping NCI	**Biomedical**	15.353 MB	**SROIQ(D)**	2371 ms
	2)FMA extended overlapping SNOMED		27.439 MB		out of Memory
	3) FMA small overlapping NCI		2.023 MB		266 ms
	4) FMA small overlapping SNOMED		5.845 MB		out of memory
	5) FMA Whole Ontology		44.834 MB		7660 ms
	6) NCI extended overlapping FMA		21.480 MB		10385 ms
	7) NCI extended overlapping SNOMED		35.839 MB		17659ms
	8) NCI small overlapping FMA		4.744 MB		1326 ms
	9) NCI small overlapping SNOMED		18.051 MB		10280 ms
	10) NCI Whole ontology.		48.841 MB		22979 ms
PELLET	1) FMA extended overlapping NCI	**Biomedical**	15.353 MB	**SROIQ(D)**	10171 ms
	2)FMA extended overlapping SNOMED		27.439 MB		19532 ms
	3) FMA small overlapping NCI		2.023 MB		1497 ms
	4) FMA small overlapping SNOMED		5.845 MB		3557 ms
	5) FMA Whole Ontology		44.834 MB		47174 ms
	6) NCI extended overlapping FMA		21.480 MB		148000 ms
	7) NCI extended overlapping SNOMED		35.839 MB		194563 ms
	8) NCI small overlapping FMA		4.744 MB		15367 ms
	9) NCI small overlapping SNOMED		18.051 MB		168621 ms
	10) NCI Whole ontology.		48.841 MB		1070942 ms

HERMIT	1) FMA extended overlapping NCI	**Biomedical**	15.353 MB	ALCQHIR+(D-)	9782 ms
	2)FMA extended overlapping SNOMED		27.439 MB		15007 ms
	3) FMA small overlapping NCI		2.023 MB		1404 ms
	4) FMA small overlapping SNOMED		5.845 MB		3089 ms
	5) FMA Whole Ontology		44.834 MB		29874 ms
	6) NCI extended overlapping FMA		21.480 MB		395645 ms
	7) NCI extended overlapping SNOMED		35.839 MB		out of memory
	8) NCI small overlapping FMA		4.744 MB		33961 ms
	9) NCI small overlapping SNOMED		18.051 MB		427238 ms
	10) NCI Whole ontology.		48.841 MB		out of memory
RACER-PRO	1) FMA extended overlapping NCI	**Biomedical**	15.353 MB	**SHIQ**	123387 ms
	2)FMA extended overlapping SNOMED		27.439 MB		84347 ms
	3) FMA small overlapping NCI		2.023 MB		9871 ms
	4) FMA small overlapping SNOMED		5.845 MB		18807 ms
	5) FMA Whole Ontology		44.834 MB		160486 ms
	6) NCI extended overlapping FMA		21.480 MB		121960 ms
	7) NCI extended overlapping SNOMED		35.839 MB		261315 ms
	8) NCI small overlapping FMA		4.744 MB		42754 ms
	9) NCI small overlapping SNOMED		18.051 MB		142994 ms
	10) NCI Whole ontology.		48.841 MB		233462 ms

IV: CONCLUSION

Semantic web will provide us a world where web pages will be having meaningful content & software agents or intelligent software's will take information from pages to pages for reaching a meaningful result. Inference engines will be required for classifying number of ontologies with respect to millions of domains & therefore it will be necessary to determine the suitability of inference engine with respect to a particular domain & to decide which inference engine will be faster for classifying which ontology. so here we have performed this test on some large sized biomedical ontologies & found that among all inference engines used, FACT++ was more suitable as compared to other inference engines for large sized ontology classification.

V. REFERENCES

1) Ajaysinh Rathod et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3), 2012, 4199-4201
2) Tim Berners-Lee, James Hendler, and Ora Lassila. *The semantic web*. *Scienti_c American*, 284(5):34{43, 2001
3) "*HermiT: A Highly-E_cient OWL Reasoner*" By Rob Shearer, Boris Motik, and Ian Horrocks, Oxford University Computing Laboratory, Oxford, OX1 3QD, UK
4) "*Sementic Web"* by Dieter Fersel and Kiorpaes, 2010.
5)https://bioportal.bioontology.org/ontologies.
6) *Pellet: An owl dl reasoner*. B. Parsia, E. Sirin, In: Proc. International Semantic Web Conference. 2005.
7) Research.ijcaonline.org/volume57/number17/pxc3883748.pdf
8) LogMap results for OAEI 2011 Ernesto Jim´enez-Ruiz, Ant´on Morant, and Bernardo Cuenca Grau Department of Computer Science, University of Oxford {Ernesto, anton.morant, berg}@cs.ox.ac