

Research on Implementation of Knowledge Convergence Based on Apache Jena3

Xiabai Yang^{1,*}, Ming Yang¹, Deqiang Yang² and Yu Huang³

¹Yunnan College of Finance and Economics, Kunming 650222,China

²College of Computer Science and Information Technology, Yunnan Normal University, Kunming 650092, China

³Kunming Railway Vocational Technical College, Kunming 650208,China

*Corresponding author

Abstract—With the development of Internet and Semantic Web technology, a huge amount of information appears on the network. Meanwhile, more and more knowledge was described based on XML (eXtensible Markup Language) and RDF(Resource Description Framework)format. It is a worthy study on problem of Knowledge Convergence on “how to make the computer automatically to extract useful knowledge from the existing knowledge through the Set Operation and reasoning”. This paper mainly studies on the architecture of Apache Jena3, operations of Union, Intersection and Difference. And these operations are Knowledge Convergence implemented by Java programming based on Apache Jena3 system. The results have been validated by online Parser “W3C RDF Validator”, which proves that the programming design implementing Knowledge Convergence based on Apache Jena3 is correct.

Keywords—*apache jena3; semantic Web; RDF; RDFS;OWL; knowledge convergence; union operation; intersection operation; difference operation; triple*

I. INTRODUCTION

In February 1998, XML1.0 became the standard recommended by the W3C(WWW Consortium). RDF was adopted as a W3C recommendation in 1999. The RDF1.0 specification was published in 2004, the RDF1.1 specification in 2014. RDFS(Resource Description Framework Schema Specification) is an extension of the symbol and syntax of the XML language, which can be said of RDF data and grammar Specification.

With the development of the semantic WEB, the Internet knowledge based on XML format and RDF format is increasing, and it is a worthy study on problem of Knowledge Convergence on “how to make the computer automatically to extract useful knowledge from the existing knowledge through the Set Operation and reasoning”. Since 2003, my team has studied the semantic WEB architecture, Jena2 architecture, RDF API of Jena2 system, Triple (subject, predicate, object) and Triple Store, based on Jena2 system using the Java programming to realize the Union, Intersection and Difference Operation of Knowledge Convergence, and the relevant papers have been published. This paper mainly studies on the architecture of Apache Jena3 system, operation of Union, Intersection and Difference.The operations are Knowledge Convergence achieved by Java programming based on Apache Jena3 system, and the program results have been validated by online Parser “W3C RDF Validator”, which proves that the

programming design achieving Knowledge Convergence based on Apache Jena3 is correct.

II. SEMANTIC WEB ARCHITECTURE

In 2001, Tim Berners-Lee, the founder of WWW proposed the semantic Web architecture is made up of seven layer[1], as shown in figure I, from down to up in turn to Unicode+ URI, XML+NS+XMLS, RDF+RDFS, Ontology Vocabulary, Logic, Proof and Trust, which lays the foundation for the development of the semantic Web.

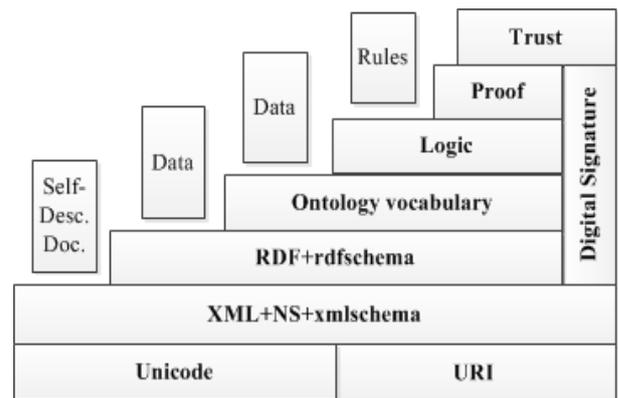


FIGURE I. SEMANTIC WEB ARCHITECTURE

III. APACHE JENA3 ARCHITECTURE OVERVIEW

Jena1 was the first version released in 2000, Jena2 in 2003, Apache Jena3(Jena3) in 2015.

Jena3 is a free and open source Java framework for building Semantic Web and Linked Data applications. It is a toolkit for processing semantic WEB written in the Java language[2].Jena3 is a major version release for Jena - it is not binary compatible with Jena2. The migration consists of package renaming and database reloading. Packages with a base name of com.hp.hpl.jena become org.apache.jena.

At Jena3’s core, it stores information as RDF triples in directed graphs, and allows programming to remove, add, store, manipulate and publish that information. Jena3 consists of several subsystem such as Store API, Inference API, Parsers and writers, RDF API, Ontology API, SPARQL API and Fuseki, etc. As shown in figure II.

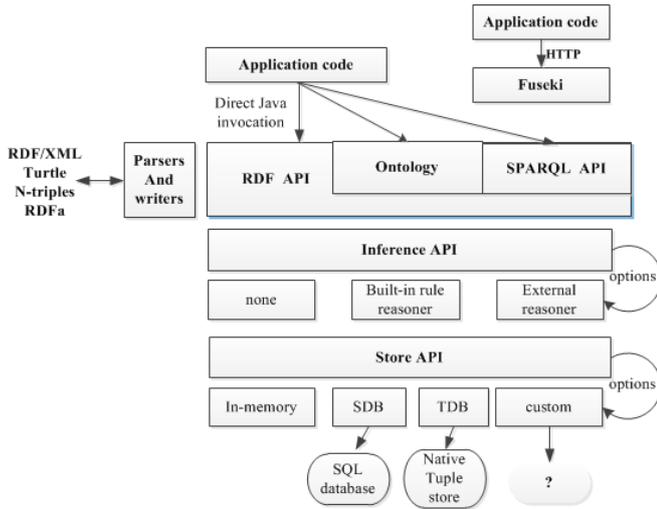


FIGURE II. THE APACHE JENA3 ARCHITECTURE

A. RDF API

The RDF API provides Java application interfaces for users that create and read RDF graphs, and support for RDF/XML or Turtle formats. SPARQL (Simple Protocol and RDF Query Language) is an engine for querying RDF data, which is a Query Language and data acquisition Protocol for RDF development and connects web3.0 and Semantic web, supporting remote federated queries and free text search.

B. Store API

Store API mainly provides TDB and SDB applications, TDB is the data saved in a Triple format, which supports the full range of Jena APIs. SDB is data saved in SQL database format.

C. Ontology API

Ontology API usage models, RDFS and the Web Ontology Language (OWL) add extra semantics for RDF data. RDF has two ontology languages: RDFS and OWL. OWL is much more expressive, and Jena3 supports both languages through the Ontology API.

D. Inference API

Inference API provides an application interface for reasoning, which can work based on the rules of inference or using built-in OWL and RDFS reasoners.

The main features of semantic web applications: the semantic rules of RDF, RDFS and OWL can be used to infer unknown information that is contained in RDF direct graph. For example, if C is the biological father of B, and B is the biological father of A, then by implication C is the grandfather of A, and A is the grandson of C.

E. Fuseki

Fuseki provides interaction with your server-side RDF data over HTTP. That is to say: Fuseki is a server, and it can access RDF models data over the web using SPARQL and HTTP.

IV. DESIGN AND IMPLEMENTATION OF KNOWLEDGE CONVERGENCE BASED ON APACHE JENA3

The website of Apache Jena (<http://jena.apache.org>) is an open source web site which studies and researches the Jena system. Apache Jena provides common Apache-jena and Apache-jena-fuseki for semantic WEB users.

Apache-jena contains API, SPARQL engine, TDB (native RDF database) and command-line tools. Apache-jena-fuseki is the Jena SPARQL server. In 2017, the new version released by Apache Jena is Apache-jena-3.6.0.Zip and Apache-jena-fuseki-3.6.0.Zip. Jena3 requires Java8 (from Jena version 3.0.0 onwards).

Development environment set up: Windows 7 operating system, install Java 1.8.0_121,Eclipse4.7.0, download Apache-jena-3.6.0. Zip, set up the Eclipse environment, import the Jena system and call the RDF API programming provided by the Jena system to implement Knowledge Convergence by the operations of Union, Intersection and Difference (source code omitted). The interface design is showed in figure III.

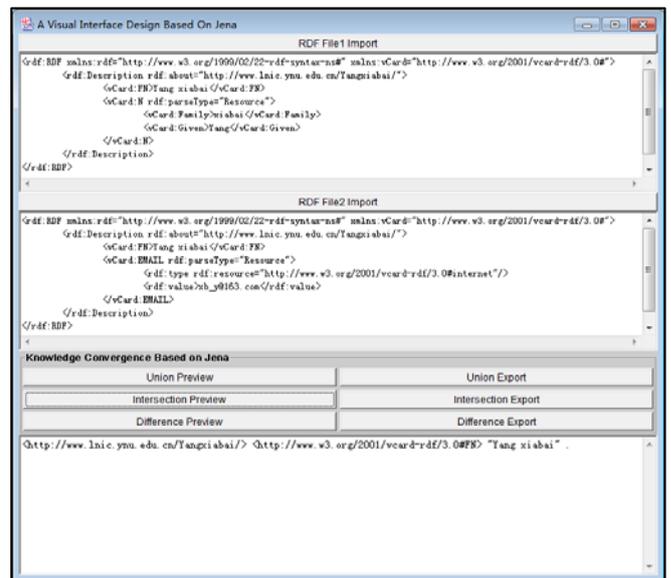


FIGURE III. A VISUAL INTERFACE DESIGN AND INTERSECTION PREVIEW (WITH THE OUTPUT FORMAT "N-TRIPLE")

Figure 3 shows the interface design of the Knowledge Convergence and the results of the Intersection Operation (N-TRIPLE parameter display). The command button "RDF File1 Import" and "RDF File2 Import" respectively implements the import of first RDF file and the second RDF file from the storage medium. The contents of the two RDF files are shown in Textbox1 and Textbox2 respectively. The "Union Preview", "Intersection Preview" and "Difference Preview" command button respectively implement the Union, Intersection and Difference Operation between RDF file1 with RDF file2 and show the results in the Textbox3. The "Union Export", "Intersection Export" and "Difference Export" command buttons respectively implement the Union, Intersection and Difference Operation between RDF file1 with RDF file2, and export the results to the storage medium.

V. PARSERS AND VALIDATION RESULTS

"The W3C RDF Validator" is an RDF Validator provided by the website (<https://www.w3.org/RDF/Validator/>). The two original files (file1.rdf, file2.rdf) and the three RDF files (Union.rdf, Intersection.rdf, Difference.rdf) which were generated from the program in figure 3 are uploaded to the web site for parsing and validation respectively. As the results shown from figure IV to figure VIII, each figure shows the corresponding Triple and Graph of the RDF file.

Figure IV is the original File1.rdf document. Figure V is the original file2.rdf document. Figure VI is the result of Union Operation between fine1.rdf with fine2.rdf. Figure VII is the result of Intersection Operation between fine1.rdf with fine2.rdf. Figure VIII is the result of Difference Operation between fine1.rdf with fine2.rdf.

The validation shows that the results obtained from the Union, Intersection and Difference Operation of Knowledge Convergence between RDF file1.rdf with RDF file2.rdf are correct.

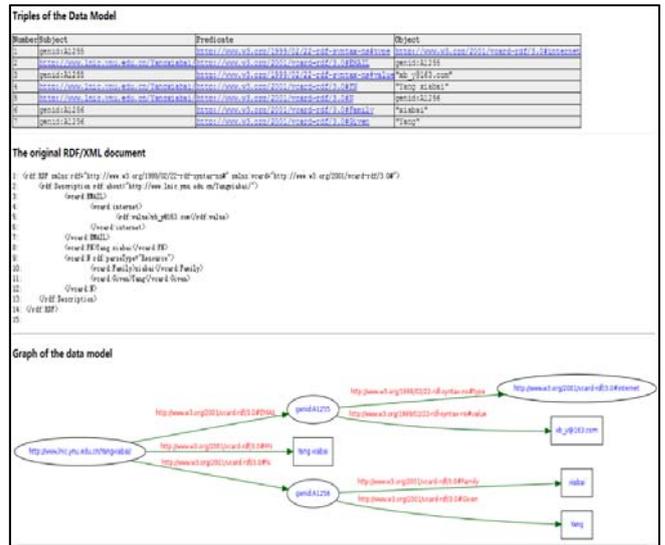


FIGURE VI. THE UNION OPERATION RESULT

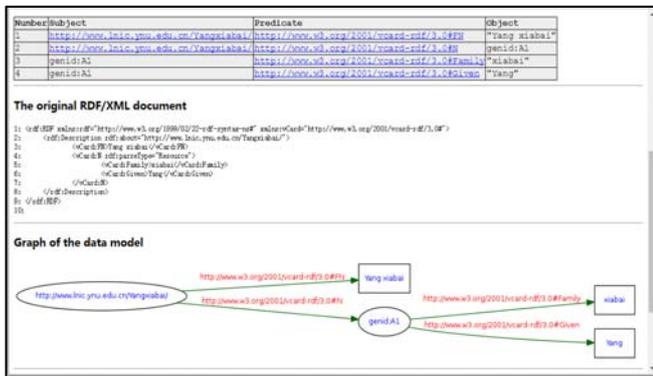


FIGURE IV. THE ORIGINAL FILE1.RDF DOCUMENT

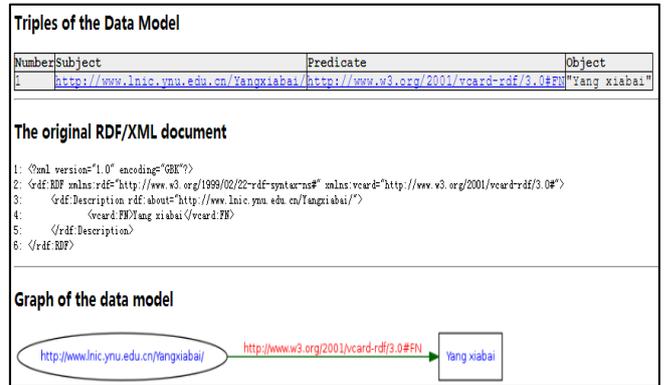


FIGURE VII. THE INTERSECTION OPERATION RESULT

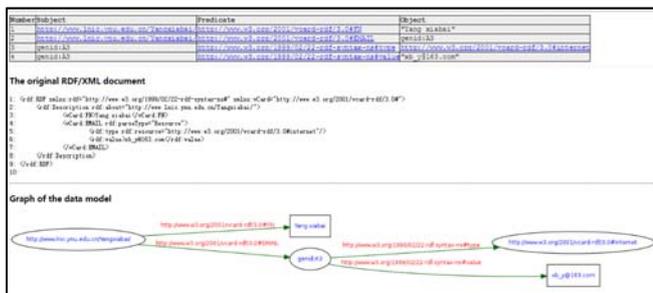


FIGURE V. THE ORIGINAL FILE2.RDF DOCUMENT

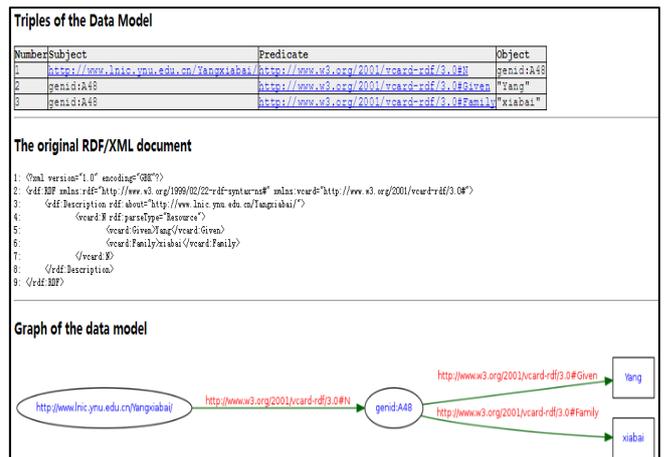


FIGURE VIII. THE DIFFERENCE OPERATION RESULT

REFERENCES

- [1] Berners-Lee, T.; Hendler, J.; Lassila, O., "The Semantic Web", Scientific American, May 2001.
- [2] http://jena.apache.org/about_jena/architecture.html, "Jena architecture overview", January 2018.