International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 2 Issue 11, November - 2013

Publishing & Querying Linked Data – Case Study Student Profiles

Abhijeet Verma (200701004)

DAIICT

Supervisor
Prof. PM Jat

Abstract: In this workwe have published the linked data of student profiles and implemented the web interface to the Sparql Query that allow us to Query the linked data through the sesame server and get the desired output.Later we will also discuss the basis of selecting certain vocabularies over others and the advantages of using RDF over RDBMS.

***Keywords*—RDF, SPARQL Query, URI, Vocabulary, License.**

I. INTRODUCTION

Querying Linked Datasets (Dataset is a set of collection of triples) help us to extract data from the database containingcollection of rdf file which is store in Sesame Server. Rdf file is collection of triples which stores triples in the format subject, object and predicate. We will extract data from rdf file by writing query in the Sparql Language. By querying the linked Dataset in the web interface to the Sparql query we will get the desired output.In the context of linked data we basically differentiate between: Linked data publisher- a person or organization exposing structured data as rdf on the web, Linked data consumer- these might be machine, for example using a semantic indexer or a query engine or, as well as humans. We publish the Resumes of DAIICT University students and in our case linked data consumer is the web interface to the Sparql query.

In the section 1 we will describe how we linked the data and in section 1.1 what are the advantage of linked data, section 2 we will tell what vocabulary (global namespace) we have used, in section 2.1we explain the selection of vocabulary

1. How we link the Data

First we find all the terms that are in the student resume document such as (college name school name, CPI, percentage) then we find the global namespace (URLs)for every term from the swoogle. Swoogle is the semantic web search engine. After finding the Global namespace for every term we provide every namespace with identifier so that we don't have to write the whole url while writing the query or making rdf document. For linking the data, we have to make RDF file, which may be written in RDF/XML or N3 format.

1.1 Advantage of RDF over RDBMS

.

A simple and uniform standard data model.

RDF database systems that all share the same well-specified and W3C-standardized data model at their base hence these data model are interoperable and standardize (A common syntax for all applicationsrdf/xml,n3). The RDF model is essentially the canonicalization of a (directed) graph and has all the advantage (and generality) of structuring information using graphs wherein the arcs start with subject URIs, are labeled with predicate URIs, and end up pointing to object URIs or scalar values

Flexibility

RDF's ability to assign any predicate/object pair to any subject makes it an ideal format for a freeform database. For example, if my address book lists each person's name, address, phone number and e-mail address, and I want to add a single new field for one

person, it's just one more triple to add, with no schema revision to contend with. (The W3C does have a specification for RDF schema available, but RDF schemas are more commonly used to provide additional information about data, such as data types and domains, than to constrain the entered data If I add a triple saying that Nick Charles has a dog Name value of "Asta", displaying Nick's information will then show the dog's name along with Nick's address and other information, even if no one else in my address book ever had a dogName property assigned. It means that information can be easily extended as application requirements evolve. Hence, Due to Flexibility RDF can be easily extended.

Distributed Data

Having two files of RDF triples written in the n-triples syntax, or in the related Notation 3 , append one file to the other, and you've just merged two databases. This ease of data aggregation makes RDF a popular format for storing distributed data that is collected whenever some application needs an updated set. Basically it helps in meaningful compositions of data from different applications. Hence, due to distributed data RDF can easily be merged.

If the RDF database solution *A* was easy to get going with but eventually for some reason hits a brick wall, just switch to RDF database solution *B* or *C* or any other of the many available interoperable solutions

What vocabularies we have used

In this section we will provide you with all the vocabularies (global namespace) that we have used for different terms.

Term	Identifier	Vocubalary
Name, mail box,project,technical elective	name, mbox, pastProject, Interest	http://xmlns.com/foaf/0.1/
CPI	cpi	http://www.burningbird.net/resume_schema#
Branch Name	Bname	http://xmlns.com/wordnet/1.6
Graduation	Graduation	http://ebiquity.umbc.edu/ontology/person.owl
Number of Members in Project	TS	http://www.w3.org/2003/01/geo/wgs84_pos
Id	ID	http://rdfs.org/sioc/ns
city	CITY	http://www.w3.org/2000/10/swap/pim/contact
Phone	ContactLocation	http://www.w3.org/2000/10/swap/pim/contact"
Address	add	http://www.w3.org/2001/vcard-rdf/3.0
College name	Cname	http://ontoworld.org/wiki/Special:URIResolver/Relation
(12th percentage),	per	http://www.owlontologies.com/nullontology.owl
10th percentage	perc	http://example.org/ontologies/VDO
10th School	Scname	http://lsdis.cs.uga.edu/projects/semdis/opus#School
12thSchool	Sname	http://www.isi.edu/webscripter/bibtex.o.daml
ExpertiseArea	XA	http://www.owl-ontologies.com/resume.owl
Programming Language	PL	http://owl.mindswap.org/2003/ont/owlweb.rdf
Programming Tools	tool	http://www.daml.org/tools/tools-ont
Internship	RI	http://pauillac.inria.fr/cdrom_a_graver/ftp/ocomma/comma.rdfs
Mentor	mentorof	http://purl.org/vocab/relationship/
Duration	duration	http://www.aktors.org/ontology/support
Project description	PDESC	http://usefulinc.com/ns/doap
Achievements and Awards	AA	http://xmlns.com/wordnet/1.6/
Positions of Responsibility	PR	http://veggente.berlios.de/ns/RIMOntology
Interests	hobby	http://www.daml.org/validator/examples/ont2.daml/
Date of Birth	DOB	http://http://www.w3.org/2001/vcard-rdf/3.0/

Web Interface to the SPARQL Query

Enter your SPARQL query in the textbox below.

Sample Queries:

- WHAT ARE THE NAMES AND MAILBOX OF STUDENTS
- WHAT ARE THE NAMES AND SCHOOL NAME OF STUDENTS

HELP

```
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
SELECT ?name ?mbox
WHERE {
?x foaf:name ?name .
?x foaf:mbox ?mbox
}
```

XML Execute

Thu Apr 21 2011 21:59:25 GMT+0530 (India Standard Time)

Figure2: Web Interface to SPARQL Query

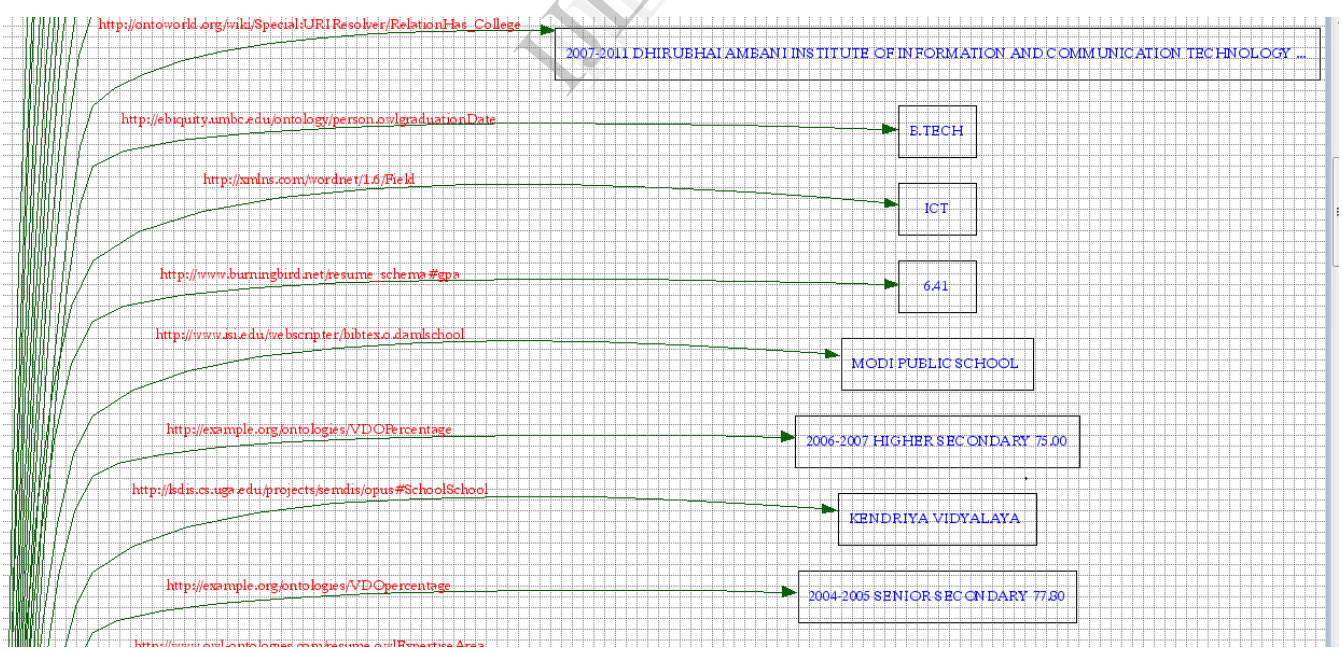

Figure3: Generated Sample RDF graph

Figure1 : Sample of RDF file

```
@prefix perc: <http://example.org/ontologies/VDO> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix tool: <http://www.daml.org/tools/tools-ont> .

<http://www.200701012RESUME.com/>     a foaf:Description;
     <http://ebiquity.umbc.edu/ontology/person.owlgraduationDate> "B.TECH"
     <http://example.org/ontologies/VDOpercentage> "2003-2004 SENIOR SECON
     DOB:dateOfBirth "1988-12-27";
     :SchoolSchool "Maharaja Sawai Man Singh Vidyalaya, Jaipur \"Standard
```

2.1 Selection of Vocabulary

Before selecting vocabulary we must saw that URLs must be *dereference able*, meaning that a description of the concept can be retrieved from the Web. We have use HTTP URIs for all attribute that we have in our Resume. The http:// scheme is the only URI scheme that is widely supported in today's tools and infrastructure. A set of well-known vocabularies have evolved in the Semantic Web communityand some of these vocabularieswhich we have used are:

Friend-of-a-Friend (FOAF)

Dublin Core (DC)

Semantically-Interlinked Online Communities (SIOC)

We also select these vocabulary on the basis of their associated license making sure that data are accessible under open access license and by looking their metadata(a set of data that describe and give information about other data). Creative Commons is the license which is associated with all these vocabulary and before finally selecting vocabulary we saw the vocabulary metadata. In meta data we find that what is the number of other link that particular vocabulary is linked to. We select that vocabulary which has the maximum number of link

Conclusion

As a data model, RDF has advantages over any other approach. We can represent, describe, combine, extend and adapt data and their organizational schema flexibly and at will. We can explore and analyze in ways not easily available with other models.

And, importantly, we can do all of this without the need to change what already exists.

References :

[1]Linked Data Tutorial http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/#hownae

[2] Describing Linked Datasets Keith Alexander,RichardCyganiak,Michael Hausenblasand Jun Zhao

[3]Describing Copyright in RDFhttp://creativecommons.org/ns

[4]Creative Commons Rights Expression Languagehttp://wiki.creativecommons.org/CC_REL

[5]Learn to Create Websiteshttp://www.w3schools.com

[6]SPARQL Query Language for RDFhttp://www.w3.org/TR/rdf-sparql-query/

[7]An Introduction to the Resource Description Frameworkhttp://www.dlib.org/dlib/may98/miller/05miller.html

[8]The RDF Advantages Pagehttp://www.w3.org/RDF/advantages.html

[9]How RDF Databases Differ from Other NoSQL Solutionshttp://blog.datagraph.org/2010/04/rdf-nosql-diff

[10] Advantages and Myths of RDF

http://www.mkbergman.com/483/advantages-and-myths-of-rdf/

[11] What is Linked data http://www.w3.org/egov/wiki/Tutorial/Linked_Data