Evaluation of LinDA and Comparison with other tools for visualization of Linked Data

List of Alternate Tools Specification

Lab Semantic Web Technologies – WS14/15

Authors:

Mentor:
Dr. Fabrizio Orlandi

Rose-Mary Owusuaa Mensah

Nishananth Baskaran

Ugochukwu Chimbo Ejikeme

Nikhil Patra

Date: 28th November, 2014

Introduction

This document outlines potential datasets which can be used in our lab project to evaluate LinDA and compare with other visualization tools for Linked data. Various datasets have been selected based on their domain (for example; energy, geology, biology), the number of triples and the data format (eg; nt, RDF/XML, csv) in which the triples are represented. These datasets have been described below

1. Linked Clean Energy Data (reegle.info)

Description: Comprehensive set of linked clean energy data including policy and regulatory country profiles, key stakeholders (organisation profiles), project outcome documents and a thesaurus on renewables, energy efficiency and climate change

Website: http://datahub.io/dataset/clean-energy-data-reegle

Triples: 334,049

Data format: RDF/XML

2. Bio2RDF: Wikipathways

Description: WikiPathways is a resource providing an open and public collection of pathway maps created in a Wiki like style.

Website: http://datahub.io/dataset/bio2rdf-wikipathways

Triples: 514,397

Data format: nquads, RDF

3. British Geological Survey (BGS) OpenGeoScience

Description: It contains Geology of the UK, BGS Lexicon of Named Rock Units

Website: http://data.bgs.ac.uk/downloads.html
Triples: Approx 840,000
Data format: N triples

4. DBpedia

Description: Data set containing extracted data from Wikipedia. About 2.6 million concepts described by 247 million triples, including abstracts in 14 different languages
Website: http://wiki.dbpedia.org/Downloads32
Triples: 247 million
Data format: nt, csv

5. BIO2RDF

Description: Various bio- and gene- related datasets
Website: http://download.bio2rdf.org/current/release.html
Triples: 10+ billion triples
Data format: n triples, n3/turtle, Json, csv

6. Project dataset

Description: A number of datasets on camera, cars, films, etc presented in csv format
Website: http://perso.telecom-paristech.fr/~eagan/class/as2013/inf229/labs/datasets
Data format: csv

Table 1 Summarised list of data sets

Dataset	Size	Data format	Source
Linked Clean Energy Data	334049	RDF/XML	http://datahub.io/dataset/clean-energy-data-reegle
Bio2RDF: Wikipathways	514,397	nquads, RDF	http://datahub.io/dataset/bio2rdf-wikipathways
British Geological Survey (BGS)	840,000	ntriples	http://data.bgs.ac.uk/downloads.html
DBpedia	247 million	nt, csv	http://wiki.dbpedia.org/Downloads32
BIO2RDF	10+ billion	n triples, n3/turtle, Json, csv	http://download.bio2rdf.org/current/release.html
Eurostat RDF datasets	68,000	RDF	http://datahub.io/dataset/eurostat-rdf
Semantic web dog food	242 million Triples	RDF	http://www.rdfhdt.org/datasets/
UNESCO Thesaurus	69766 triples	Turtle RDF/XML	http://skos.um.es/unescothes/downloads.php
Project dataset	Datasets on cameras, cars, countries, cereals, speed dating	csv	http://perso.telecom-paristech.fr/~eagan/class/as2013/inf229/labs/datasets