Prototype implementation on DBpedia KB

Rifat

Monday, November 5, 2017

In this implementation we presents, Dbpedia 10 class quality analysis report based on four quality characteristics: (i) Persistency (ii) Historical Persistency (iii) Consistency (iv) Completeness. We divided the report based on each quality characteristics. Also we presented DBpedia growth analysis measure.

(i)   Class dbo Animal
(ii) Class dbo Artist
(iii) Class dbo Athlete
(iv) Class dbo Film
(v)   Class dbo MusicalWork
(vi) Class dbo Organisation
(vii) Class dbo Place
(viii) Class dbo Species
(ix) Class dbo Work

(X)   Class foaf Person

Quality Problem Report Overview

Input

In this implementation the datasets is presented in folder "~/ExperimentalData/DBpedia" where each class property count over DBpedia 11 releases present in a CSV file. Also, the entity count of 10 classes presented in "DBpedia10ClassEntityCount.csv" file.

```
location="~/ExperimentalData/DBpedia/"
```

The dataset presented in the folder is already processed and build in a intermediate data structure for the quality assessment approach. To run the rmd file only need to specify the folder location for input data.
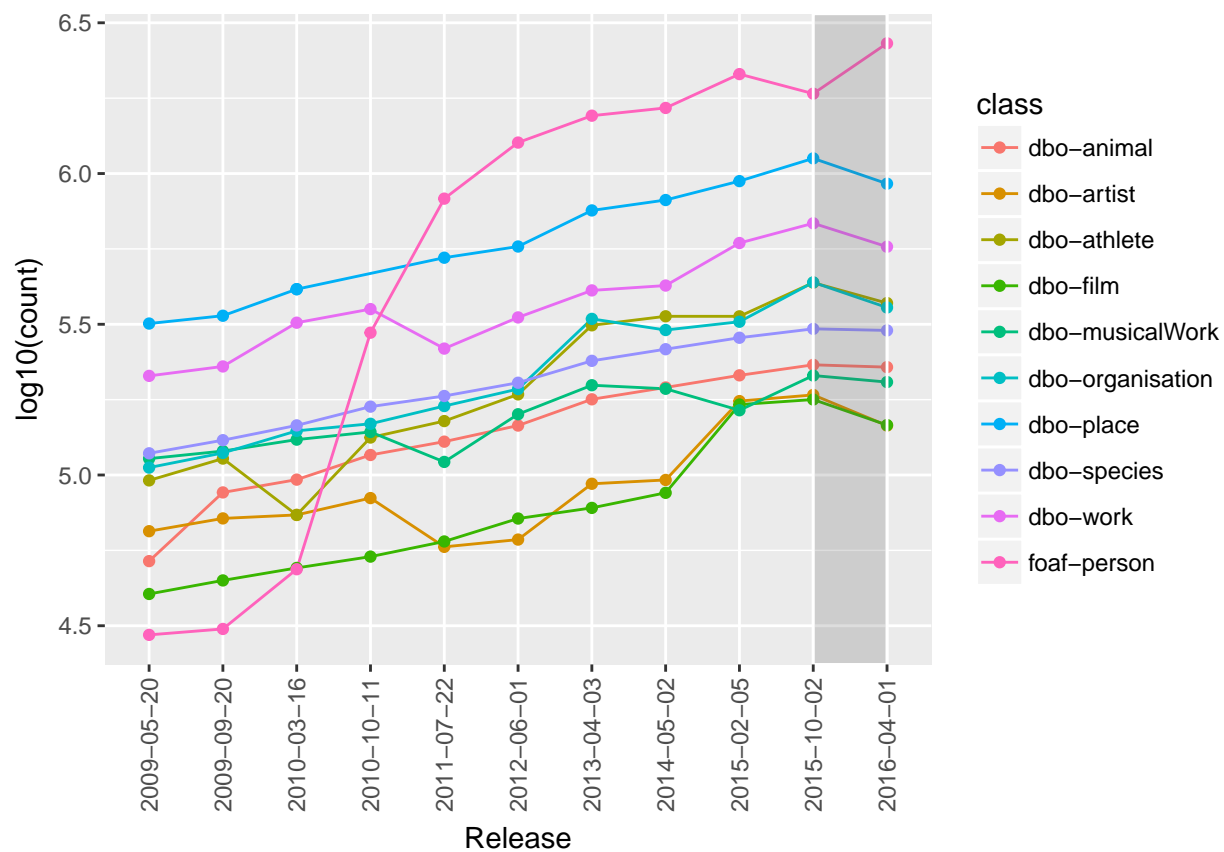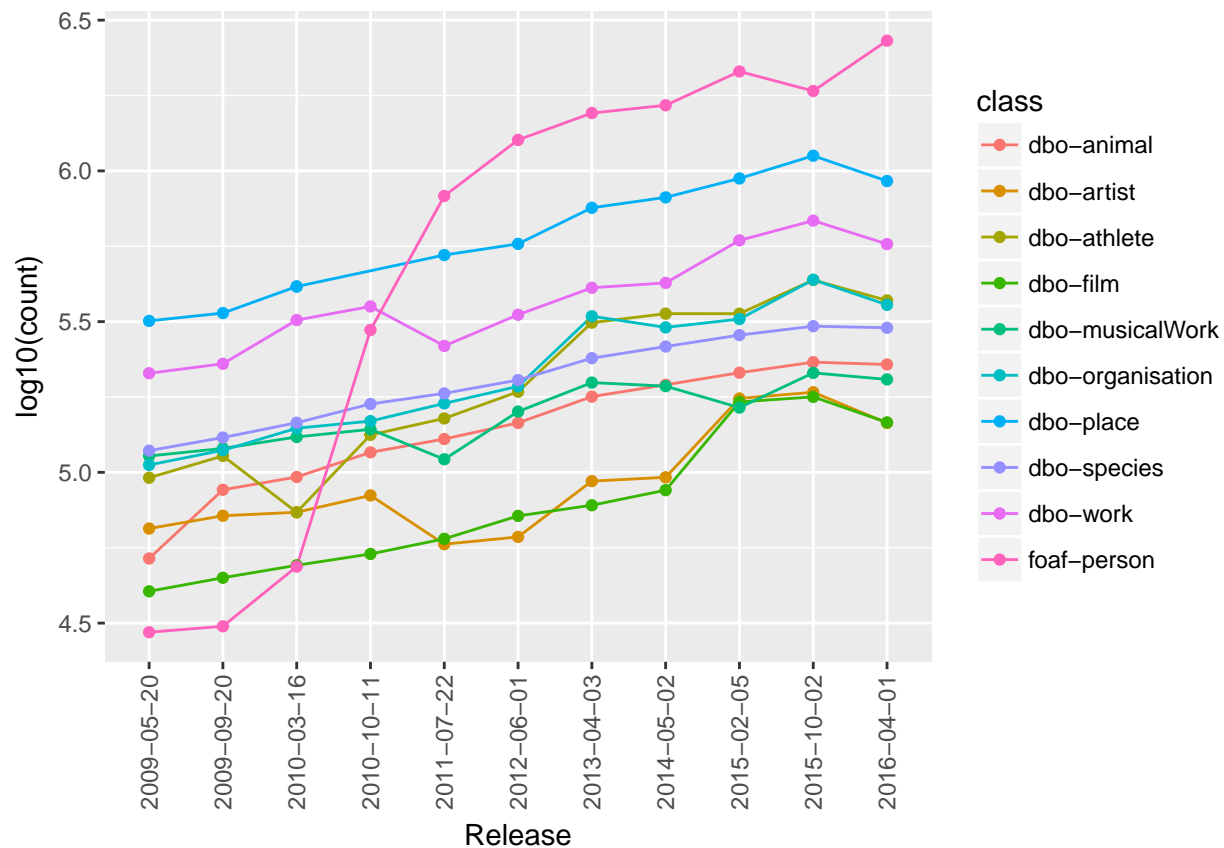
Results Structure

(i) Persistency: A line graph presented to visualize variation on last two DBpedia Release.

(ii) Historical Persistency: A table with this classes with persistency issues over the 11 Releases.

(iii) Consistency: Result of property values presented in a table. 1st table with incorrect property values and 2nd table with filtered property values

(iv) Completeness: Result of property values presented in a table with normalized property values and difference between two version of normalized property values.

(v) KB growth: A graph visualizing KB growth.

Experiment Results

Input: Set Entity file location such as -

```
location<- "/ExperimentalData/DBpedia/"
```

Over all Entity Analysis:

Class dbo Animal

[http://dbpedia.org/ontology/Animal]

Input: set file location such as -

```
location<- "/ExperimentalData/DBpedia/"
```

Persistency

Historical Persitency

Consistency

Completeness

KB growth

Class dbo Artist