Psychometry Ontology Report

Mine Ogretir

January 2019

1 Overview of the Project

Implementation of the project can be divided into four stages as shown in Figure 1.

1. **Gathering Artifact Data**: A user can enter data for a person/author via URL source. The interface is in the *ScoringAndCreatingLinkedData.ipynb* file.

2. **Converting Gathered Data as Linked Data**: The gathered data is converted to rdf data and stored in a DRF/XML file. The scripts are in the *ScoringAndCreatingLinkedData.ipynb* file.

3. **Inquiry the Data via SPARQL Queries**: Data can be queried via SPARQL. There is not a SPARQL query end point in python notebooks, however the RDF/XML file can be used via a SPARQL end point, like Apache Fuseki etc. The RDF/XML file with the highest number should be used in the *OwlData* folder.

4. **Analyzing the Data**: Obtained linked data is retrieved by SPARQL queries, and various graphs are obtained as output, which are:

   - radar chart of the person according to the mean of the all artifact scores,
   - ten basic value scores over time,
   - four higher order value scores over time,
   - statistical process control graphs of the value scores.

   The interface is in the *GraphsOfData.ipynb* file. The first four graphs are stored in the *graphs* folder.

2 Model

The diagram of the Ontology Model is shown in Figure 2. The model can be reached in the *onto* folder, the file name is *HumanPsychometry_2.owl*. The annotation of the model entities is not completed yet.

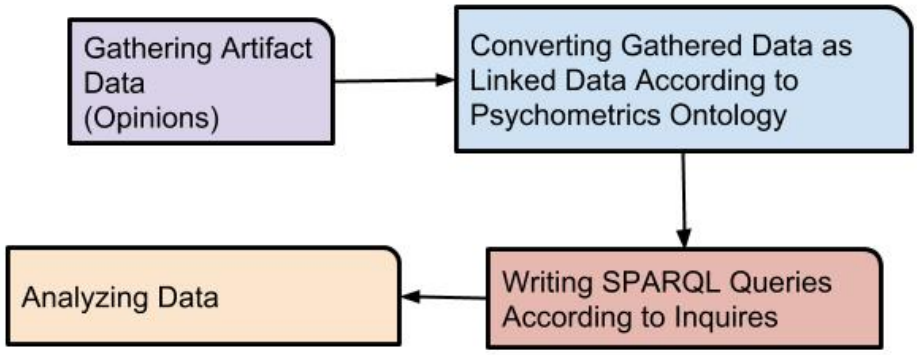

Figure 1: Pipeline of the Implementation

3 Instructions for the Usage of the Code

This section describes the usage of the notebooks. There are two notebooks for user interaction.

1. **ScoringAndCreatingLinkedData.ipynb** : This notebook allows the user to enter the data od an author, which are the name of the author, the text of the article, the value scores of the text according to SSNMF, date indicator etc.
   There are two important points:
   - Please contact with me for the API key.
   - The model file *pretrained_v3_t3_h10_1409.p* is big in size. So, please download it from **here**.
2. **GraphsOfData.ipynb** : Here, you can ask the graphs of a person. After you write the name of the author, you can execute the cells. Plotly graphs are saved in the graphs folder. Statistical process control charts are printed in the notebook. Some examples are in Figure: 3 and 3.

3.1 Loading Data

1. In the first step, the name of the author and the URL of the source page of the articles should be entered.
2. The URLs of the articles should be selected. There are two options for this:
   - Entering a URL to the text field and selecting the relevant links obtained from that URL,
   - Entering the URLs directly to a list of strings.

   The example is follows:
   Name: Abdulkadir Selvi
   Source URL: http://www.hurriyetdailynews.com/opinion/abdulkadir-selvi/

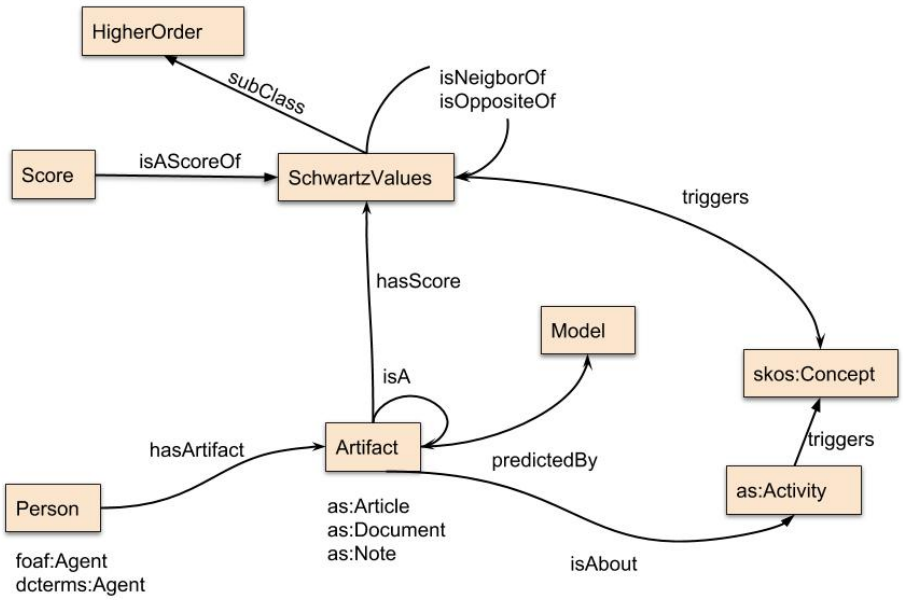

Figure 2: Diagram of Psychometrics Ontology

```
PREFIX psyc:   <http://www.semanticweb.org/mine/ontologies/2018/9/Humans_and_Psychometrics#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX as: <https://www.w3.org/ns/activitystreams#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
```

Figure 3: Used Prefixes

3. At the third step, the main text in the URLs are extracted and the scores of the texts are calculated and the data is converted to a dictionary.

4. At this step, the data dictionary is stored in a pickle file in order to converting it to linked data. The numbering is important. So, if you think you made a mistake, please delete the last file (corresponding file) in the PersonData folder.

5. The last step is to convert the data to linked data. The corresponding script gets the latest rdf file, and adds the unconverted person data to the new rdf file.

The details are also given in the notebooks.

3.2 Visualizing the Data on the Graphs

After you write the name and surname of the author, you can just execute the following scripts in the 'GraphsOfData' notebook. Plotly graphs are saved in the 'graphs' folder. They are interactive plots, you can choose which data to show.

```
SELECT  distinct  ?dateIndicator ?NamedValue_1 ?score1 ?NamedValue_2  ?score2
WHERE
{
  ?person foaf:surname 'Dyer' .
  ?person foaf:name ?name .
  ?person psyc:hasArtifact ?artifact .
  ?artifact dcterms:source ?source .
  ?artifact psyc:hasScore ?ScoreEntity1 .
  ?ScoreEntity1 psyc:isAScoreOf ?NamedValue_1 .
  ?artifact psyc:hasScore ?ScoreEntity2 .
  ?ScoreEntity2 psyc:isAScoreOf ?NamedValue_2  .
  ?NamedValue_1  psyc:isOppositeOf ?NamedValue_2  .
  ?ScoreEntity1 psyc:score ?score1 .
  ?ScoreEntity2 psyc:score ?score2 .
  ?artifact psyc:dateIndicator ?dateIndicator .
  FILTER ((?score1 > 50 && ?score2 < 50 ) || (?score2 > 50 && ?score1 < 50 ))
}
ORDER BY ?dateIndicator
```

Figure 4: An Example Query for Retrieving Scores, which are coherent with the assumption that the Values in polar coordinates have negative correlation. Text version is given in Appendix

4 Appendix

4.1 Example SPARQL Query

```
SELECT  distinct  ?dateIndicator ?NamedValue_1 ?score1 ?NamedValue_2  ?score2
WHERE
{
  ?person foaf:surname 'Dyer' .
  ?person foaf:name ?name .
  ?person psyc:hasArtifact ?artifact .
  ?artifact dcterms:source ?source .
  ?artifact psyc:hasScore ?ScoreEntity1 .
  ?ScoreEntity1 psyc:isAScoreOf ?NamedValue_1 .
  ?artifact psyc:hasScore ?ScoreEntity2 .
  ?ScoreEntity2 psyc:isAScoreOf ?NamedValue_2  .
  ?NamedValue_1  psyc:isOppositeOf ?NamedValue_2  .
  ?ScoreEntity1 psyc:score ?score1 .
  ?ScoreEntity2 psyc:score ?score2 .
  ?artifact psyc:dateIndicator ?dateIndicator .
  FILTER ((?score1 > 50 && ?score2 < 50 ) || (?score2 > 50 && ?score1 < 50 ))
}
ORDER BY ?dateIndicator
```

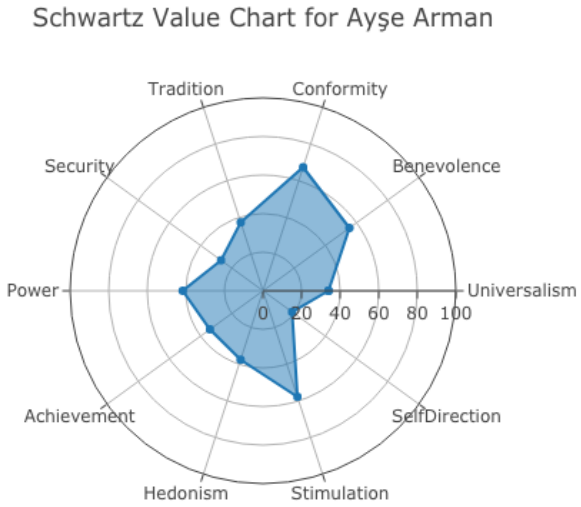

Figure 5: Radar Chart of Ayşe Arman

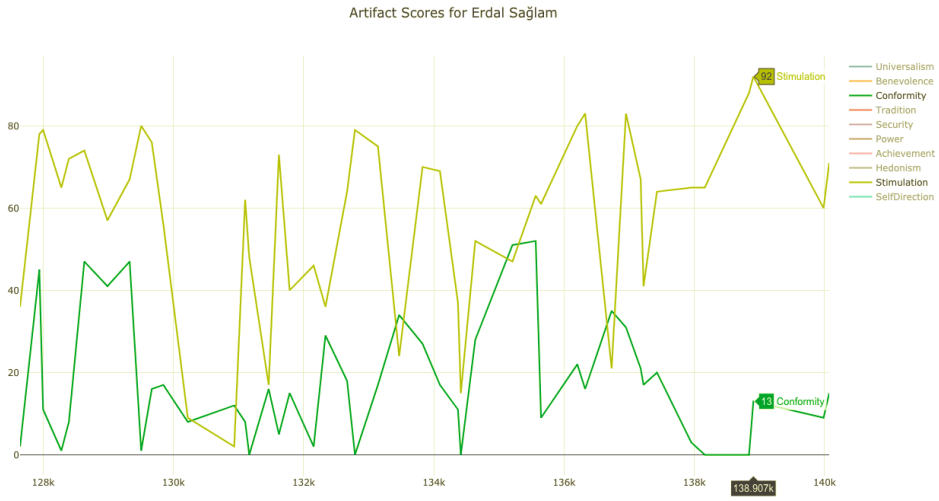

Figure 6: Partial Value Scores of Erdal Sağlam