

Linked Data quality assessment and fusion

Assumptions:

- multiple descriptions should point to the same URI
- two property values for same attribute should have been assigned a unique URI with two property values

Quality aspects: accuracy, timeliness, completeness, relevancy, objectivity, believability, understandability, consistency, conciseness, availability, and verifiability.

QualityAssessment

Assessment metrics: measures information quality dimensions. Each metric relies on a set of quality indicators and, calculates an assessment score from these dataQualityIndicators using a scoring function

- Data Quality Indicator: indicates data suitability
- Scoring function: assessment of a dataQualityIndicator. Evaluated by the user in the process of deciding suitability
- Aggregate metric: A user specified aggregate assessment metric build out of individual assessment metrics. Visualize it as a tree, the values at leaves are combined till we have a single value

Input

assessmentMetric: user defined assessmentMetric (relies on a set of quality indicators)

aggregatedMetric: Multiple AssessmentMetric

dataQualityIndicator (used by scoring functions)

dateInput (user input date)

listOfGraphs: list of graphs separated by a space

Output

score

There is a defined list of scoring functions. However, new ones can be added via xml file

```
Function qualityAssessmentModule(String assessmentMetric, String [] aggregatedMetric, String []  
dataQualityIndicator, Date dateInput, String [] listOfGraphs){  
{
```

```

switch(assessmentMetric)
{
case recency:
{
switch(scoringFunction)
{
case timeCloseness:
// Measures the distance between the input date from the provenance graph
// to the current date. The most recent data gets the value close to 1.

// output score will be a value for sieve:recency property score = timeCloseness(dateInput, currentDate); //
between 0 to 1
break;
}
break;
}

case reputation:
{
switch(scoringFunction)
{
case preference:
// assigns decreasing scores to each graph URI provided in the configuration
score = preference(listOfGraphs);
}
}

return score;
}

```

```

Function timeCloseness(dateInput, currentDate){
//measures distance between the input date from the provenance graph to the current date. Output is closer
to 1 if it is more recent.
}

```

```

Function preference(listOfGraphs){
// do stuff here
// Preference in descending priority order of appearance in the list parameter

}

```

Data Fusion

- keepUpToDate - elections
- average - election votes

- max - attendance
- pick most frequent - number of deaths

Conflict-ignorance:

- for locations and dates of events (e.g. Wembley, London; 2pm, 24th May)