universität innsbruck

Department of Computer Science

STI INNSBRUCK

SE Research Seminar: Knowledge Graphs Knowledge Assessment

By Penz Manuel, Rasmusen Sven, Zaric Marko

Dimensions & Metrics

- **Accessibility**
  - **Provisioning of public endpoint**
    - Weight = 0.4
    - 1 If SPARQL and REST API
    - 0.75 If SPARQL OR REST API
    - 0.5 If REST API OFFLINE DATA DUMP
    - 0 Otherwise
  - **Retrievable format**
    - Weight = 0.4
    - 1 If RDF export available
    - 0.75 If JSON export available
    - 0.5 If semi-structured data available
    - 0 Otherwise
  - **Content negotiation**
    - Weight = 0.2
    - 1 If content negotiation is supported
    - 0 Otherwise

developer preferences, one developer might prefer json over rdf, oder umgekehrt

Dimensions & Metrics

"Why is having SPARQL endpoint better than having a REST API?"

- **Accessibility**
  - **Provisioning of public endpoint**
    - Weight = 0.4
    - 1 If SPARQL and REST API
    - 0.75 If SPARQL
    - 0.5 If REST API
    - 0 Otherwise
  - **Retrievable format**
    - Weight = 0.4
    - 1 If RDF export available
    - 0.75 If JSON export available
    - 0.5 If semi-structured data available
    - 0 Otherwise
  - **Content negotiation**
    - Weight = 0.2
    - 1 If content negotiation is supported
    - 0 Otherwise

Dimensions & Metrics

- **Completeness**
  - **Data completeness**
    - Weight = 0.4
    - $$m = \frac{number\ of\ values\ from\ required\ classes\ \&\ properties}{number\ of\ total\ values\ from\ required\ classes\ \&\ properties}$$

Average over subset, remove "required" also take into account all existing props & classes

  - **Population completeness**
    - Weight = 0.4
    - $$m = \frac{number\ of\ objects\ per\ domain\ represented\ in\ the\ data\ source}{total\ number\ objects\ per\ domain}$$

  - **Interlinking completeness**
    - Weight = 0.2
    - $$m = \frac{number\ of\ instances\ are\ interlinked}{total\ number\ of\ instances}$$

redefine our meaning of interlinking

Dimensions & Metrics

- Rename to “Instance completeness”
- Why only **required** properties considered?

- **Completeness**
  - **Data completeness**
    - Weight = 0.4
    - $$m = \frac{number\ of\ values\ from\ required\ classes\ \&\ properties}{number\ of\ total\ values\ from\ required\ classes\ \&\ properties}$$
  - **Population completeness**
    - Weight = 0.4
    - $$m = \frac{number\ of\ objects\ per\ domain\ represented\ in\ the\ data\ source}{total\ number\ objects\ per\ domain}$$
  - **Interlinking completeness**
    - Weight = 0.2
    - $$m = \frac{number\ of\ instances\ are\ interlinked}{total\ number\ of\ instances}$$

Remove & Adjust Weights

Dimensions & Metrics

- **Accuracy**
  - **Semantic validity**
    - Weight = 1.0
    - $$m = \frac{|\{ o \mid (s,p,o) \in r \wedge o \in L \wedge semValid(o) \}|}{|\{o \mid (s,p,o) \in r \wedge o \in L \}|}$$

Dimensions & Metrics

- **Accuracy**
  - **Semantic validity**
    - Weight = 0.5
    - $m = \frac{|\{ o \mid (s,p,o) \in r \wedge o \in L \wedge semValid(o) \}|}{|\{ o \mid (s,p,o) \in r \wedge o \in L \}|}$

Say we do "formal semantic validity"

  - **Syntactic validity**
    - Weight = 0.5
    - m = $\frac{|\{o|(s,p,o) \in r \wedge o \in L \wedge synValid(o)\}|}{|\{o|(s,p,o) \in r \wedge o \in L\}|}$

Check datatype
Check string distance (example: hemming distance)

Say how we want to do our assessment. Eg. "we will take a subset for this metric...."

Thank You!

We are open for feedback and questions.