KNOWDIVE

KDI : **Knowledge and Data Integration**

Evaluation

Theory of evaluation

Fausto Giunchiglia

Contents

Contents

Motivation: Why we do evaluation?

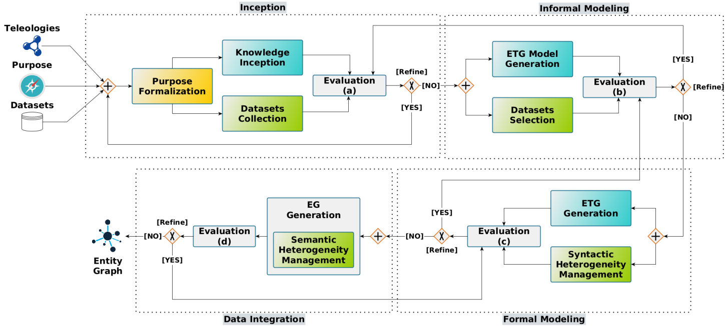
Evaluation for general purpose

Model evaluation is the task of measuring the quality of a model. It is essential for a wide adoption of model, both in the Semantic Web and in other semantically enabled technologies. As models are being developed and reused, the need to address quality issues becomes vital factor as having a true understanding of a model helps future data publisher to choose models based on ‘fitness for use’ [1].

[1] Joseph Juran and A Blanton Godfrey, Quality handbook, Republished McGraw-Hill (1999), 173–178.

Motivation: Why we do evaluation?

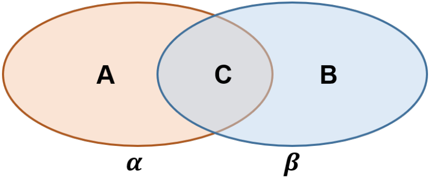
Evaluation in iTelos



- Without reasonable evaluation, it is impossible to obtain qualified entity graph in the final phase.
- In iTelos, we apply stepwise evaluation aims to refine the results in each step. This also helps to minimize the cost of mistakes.

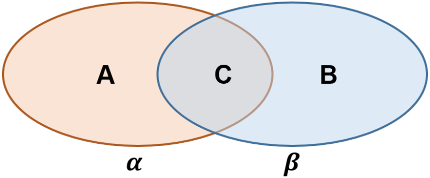
Contents

Metrics Definitions



- Coverage (*Cov*)
- Extensiveness (*Ext*)
- Sparsity (*Spr*)
- Cue Validity (*Cue*)

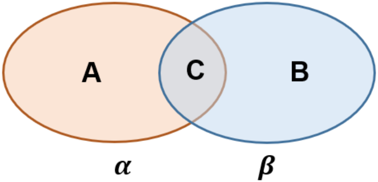
Metric definitions: Coverage

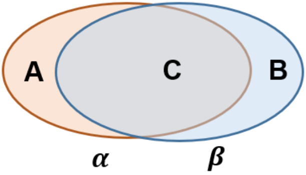


The Coverage is computed as the ration between the intersection of α and β and the whole α sets:

$$Cov = (\alpha \cap \beta)/\alpha = C/\alpha \tag{1}$$

Metric definitions: Coverage (extreme cases)





Cov = (α ∩ β)/α = C/α

Cov $\simeq 0$

Cov $\simeq 1$

Metric definitions: Coverage

About the Coverage: $Cov = (\alpha \cap \beta)/\alpha = C/\alpha$

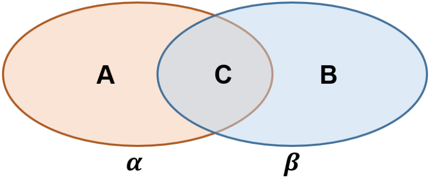
Values are always within the interval [0,1].

High values of Coverage mean that the reference schema is appropriate for the domain.

For low values of Coverage, we can have two possibilities.

- The reference schema is not appropriate for the domain and maybe a further lookup should be performed.
- The domain targeted by the knowledge graph is mostly unexplored.

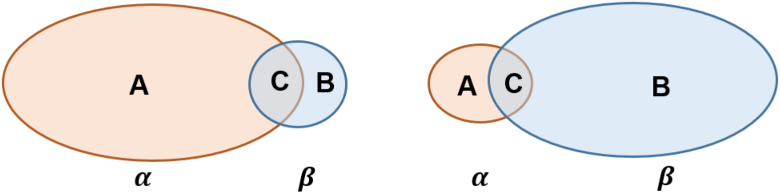
Metric definitions: Extensiveness



The Extensiveness is computed as the proportional amount of knowledge provided by β with respect to the whole knowledge defined in the graph:

$$Ext = (\beta - (\alpha \cap \beta))/((\alpha + \beta) - (\alpha \cap \beta) = B/(\alpha + \beta) \qquad (2)$$

Metric definitions: Extensiveness (extreme cases)



Ext = (β – (α ∩ β))/((α + β) – (α ∩ β)) = B/(α + β)

$\mathrm{Ext} \simeq 0$ $\mathrm{Ext} \simeq 1$

Metric definitions: Extensiveness

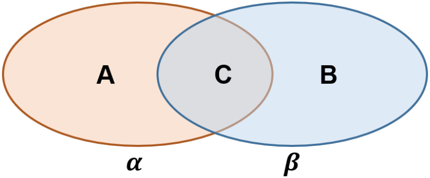
About the Coverage: $Ext = (\beta - (\alpha \cap \beta))/((\alpha + \beta) - (\alpha \cap \beta)$

Values are always within the interval [0,1].

High values of Extensiveness mean that the contribution of the created knowledge graph is predominant with respect to the content of reference schema.

Low values of Extensiveness mean that the contribution of the created knowledge graph is limited.

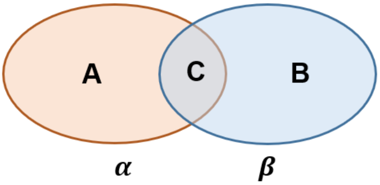
Metric definitions: Sparsity

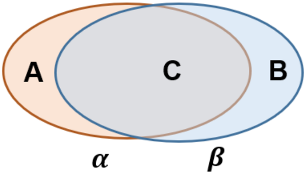


The Sparsity is computed as the sum among the percentage of α not defined in β and vice-versa:

$$Spr = ((\alpha + \beta) - 2(\alpha \cap \beta))/((\alpha \cap \beta) - (\alpha + \beta)) = (A + B)/(\alpha + \beta) \quad (3)$$

Metric definitions: Sparsity (extreme cases)





$$\mathrm{Spr} = ((\alpha + \beta) - 2(\alpha \cap \beta))/((\alpha + \beta) - (\alpha \cap \beta)) = (A + B)/(\alpha + \beta)$$

$\mathrm{Spr} \simeq 1$

$\mathrm{Spr} \simeq 0$

Metric definitions: Sparsity

About the Coverage: $Spr = ((\alpha + \beta) - 2(\alpha \cap \beta))/((\alpha \cap \beta) - (\alpha + \beta))$

Values are always within the interval [0,1].

Useful metric for measuring the differences between specific type of elements (e.g. datatype properties).

High values of Sparsity mean that there is an important difference between the considered type of elements defined in α and the ones defined in β.

Low values of Sparsity mean that there is a good match between the considered type of elements defined in α and the ones defined in β.

Metric definitions: Cue validity

Cue is a set of metrics to measure the quality of the Etype/ETG. By applying Cue, we focus on:

- Shareability and unity [1], we measure if the Etype/ETG is well-described by its features.
- Property richness [2], since we calculate the average number of properties that assigned to different Etypes.

[1] Giunchiglia, F. and Fumagalli, M., 2020, July. Entity type recognition–dealing with the diversity of knowledge. In Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (Vol. 17, No. 1, pp. 414-423).

[2] Tartir, S., Arpinar, I.B., Moore, M., Sheth, A.P. and Aleman-Meza, B., 2005. OntoQA: Metric-based ontology quality analysis.

Metric definitions: Cue validity

Cue is a set of metrics to measure the quality of the Etype/ETG.

Cue for Etype: $$Cue_e(e) = \sum_{i=1}^{|prop(e)|} Cue_p(p_i, e) \in [0, |prop(e)|]$$

Cue for ETG: $$Cue_k(K) = \sum_{i=1}^{|E_K|} Cue_e(e_i) \in [0, |prop(K)|]$$

Metric definitions: Cue validity

Cue for Etype: $$Cue_e(e) = \sum_{i=1}^{|prop(e)|} Cue_p(p_i, e) \in [0, |prop(e)|]$$

$$Cue_p(p,e) = \frac{PoE(p,e)}{|dom(p)|} \in [0,1] \qquad PoE(p,e) = \begin{cases} 1, if e \in dom(p) \\ 0, if e \notin dom(p) \end{cases}$$

- e represents an Etype. $Cue_e(e)$ represents the *cue* validity of the Etype e.
- $|prop(e)|$ is the number of properties associated with the specific entity type e.
- $Cue_p(p, e)$ returns 0 if p is not associated with e. Otherwise returns *1/n*, where n is the number of entity types in the domain of p. Cue_p takes the maximum value 1 if p has only one entity type.
- $|dom(p)|$ presents the cardinality of entity types that are the domain of the specific property p.
- $PoE(p, e)$ determines if the Etype e is in the domain of property p.

Metric definitions: Cue validity

Cue for ETG: $$Cue_k(K) = \sum_{i=1}^{|E_K|} Cue_e(e_i) \in [0, |prop(K)|]$$

- The $Cue_k(K)$ is calculated as a summation of the cue validity $Cue_e(e)$ of all the entity types e_i in a given ETG K,
- E_K presents the number of Etypes in a given ETG K.
- $|prop(K)|$ refers to the number of the properties in the ETG, as the maximization of $Cue_k(K)$.

Metric definitions: Cue validity

About $Cue_e(e)$ and $Cue_k(K)$:

- Values are always within the interval [0,1].
- It captures the idea, that is Etypes are properly described by more specific properties.
- High values of *Cue* mean that there are enough number of properties for specifically describing the target Etype/ETG, which makes the target Etype more likely belongs to contextual category.
- Low values of Cue mean that the target Etype/ETG is describe by few general properties, which makes the target Etype more likely belongs to common category.

Contents

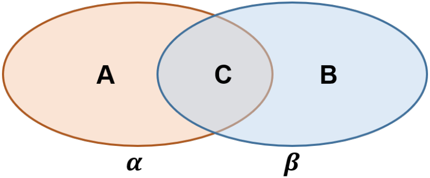
Categories of data

Common: involve resources associated to aspects which are common to all domains, also outside the domain of interest. Usually these resources. e.g., person, organization, event, location.

Core: involve resources associated to the more core aspects of the domain under consideration. They carry information about the most important aspects considered by the purpose.

Contextual: involve resources that carry specific, possibly unique, information from the domain of interest. These are the resources whose main goal is to extend added value.

Categories of data



Assume B is the collected data, we find data categories:

- Common: more likely appear in C
- Core: appear both in B and C
- Contextual: more likely appear in B

Categories of data

Thus, we should consider data categories during evaluation categories.

Table: *An example of metric values for each data category.*

	Common	Core	Contextual
Coverage	0.8	0.5 (0.8)	0.2
Extensiveness	0.5	0.5	0.5 (0.8)
Sparsity	0.2	0.5	0.8

KDI Knowledge and Data Integration

Fausto Giunchiglia

Evaluation

Theory of evaluation