



KDI ● **Knowledge and Data Integration**

iTelos - Evaluation

W3.L6.M3.T7

Contents

1 Purpose and Input Information

2 Metrics

Contents

1 Purpose and Input Information

2 Metrics

Purpose

- The Evaluation procedure puts the project in the center since it is the starting point of all activities.
- The evaluation does not aim to provide a judgement of the overall quality of the knowledge graph.
- The proposed metrics will provide a big picture about the modeling status in order to be exploited as a driver for the modeling process.

Input - General

- S is the set of axioms extracted from the reference schemata. In general, we can have more than one reference schema. Hence, all axioms (together with their possible alignments) are extracted and considered.
- E is the set containing the entities modeled within the set S .
- P is the set containing the object properties modeled within the set S .
- D is the set containing the datatype properties modeled within the set S .
- A is the set containing the annotations modeled within the set S .

Input - Inception phase

- CQ_E is the set containing the entities extracted from the Competency Questions/Queries defined by the Knowledge Engineers. The CQ_E set has the three subsets CQ_E^{CR} , CQ_E^{CM} , and CQ_E^{CN} representing respectively the Core (CR), Common (CM), and Contextual (CN) entities.
- CQ_P is the set containing the properties extracted from the Competency Questions/Queries defined by the Knowledge Engineers. The CQ_P set has the three subsets CQ_P^{CR} , CQ_P^{CM} , and CQ_P^{CN} representing respectively the object properties of the Core (CR), Common (CM), and Contextual (CN) entities.
- D is the set containing the list of the fields extracted from the datasets identified by the Data Scientist.

Input - Informal Modeling phase

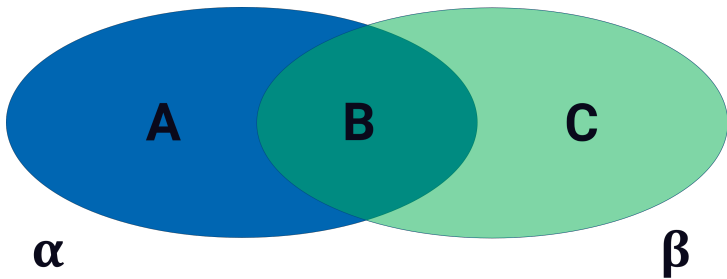
- M_E is the set containing the entities extracted from the model defined by the Knowledge Engineers. The M_E set has the three subsets M_E^{CR} , M_E^{CM} , and M_E^{CN} representing respectively the object properties of the Core (CR), Common (CM), and Contextual (CN) entities.
- M_P is the set containing the properties extracted from the model defined by the Knowledge Engineers. The M_P set has the three subsets M_P^{CR} , M_P^{CM} , and M_P^{CN} representing respectively the object properties of the Core (CR), Common (CM), and Contextual (CN) entities.
- D is the set containing the list of the fields extracted from the datasets identified by the Data Scientist.

Contents

1 Purpose and Input Information

2 Metrics

Metrics



- Coverage (Cov)
- Flexibility (Flx)
- Extensiveness (Ext)
- Sparsity (Spr)

Metrics



α

AGROVOC

Food
Nutrient
Pasta
Meet
Vegetable
Carbs
Protein
Lipid
EconomicActivity
Agriculture
Fishing
Feeding

Pasta isA Food
Meet isA Food
Vegetable isA Food
Carbs isA Nutrient
Protein isA Nutrient
Lipid isA Nutrient
Agriculture isA EconomicActivity
Fishing isA EconomicActivity
Feeding isA EconomicActivity



β

FOOD AND DIET

Food	hasFood
Nutrient	hasNutrient
Pasta	amount
Meet	hasId
Vegetable	
Carbs	Pasta isA Food
Protein	Meet isA Food
Lipid	Vegetable isA Food
	Carbs isA Nutrient
	Protein isA Nutrient
	Lipid isA Nutrient
Person	
ConsumedFood	
Recipe	
Diet	Food hasNutrient Nutrient
DietRule	ConsumedFood hasFood Food
	ConsumedFood amount <double>
	Recipe hasFood Food
	Recipe amount <double>
	Diet hasId <string>
	DietRule hasFood Food
	DietRule amount <double>

Metrics



A

EconomicActivity
Agriculture
Fishing
Feeding

Agriculture isA EconomicActivity
Fishing isA EconomicActivity
Feeding isA EconomicActivity

B

Food
Nutrient
Pasta
Meet
Vegetable
Carbs
Protein
Lipid

Pasta isA Food
Meet isA Food
Vegetable isA Food
Carbs isA Nutrient
Protein isA Nutrient
Lipid isA Nutrient

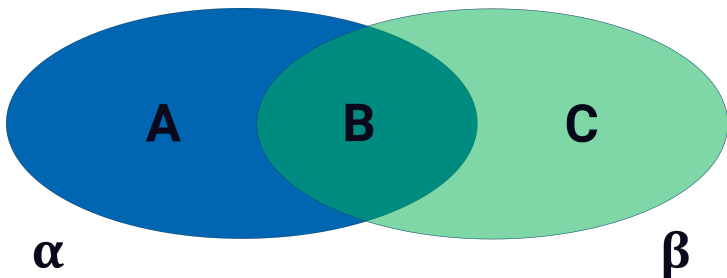
C

Person
ConsumedFood
Recipe
Diet
DietRule

hasFood
hasNutrient
amount
hasId

Food hasNutrient Nutrient
ConsumedFood hasFood Food
ConsumedFood amount <double>
Recipe hasFood Food
Recipe amount <double>
Diet hasId <string>
DietRule hasFood Food
DietRule amount <double>

Coverage



The **Coverage** is computed as the ration between the intersection of α and β and the whole α sets: $Cov = B/\alpha$

Coverage



A

EconomicActivity
Agriculture
Fishing
Feeding

Agriculture isA EconomicActivity
Fishing isA EconomicActivity
Feeding isA EconomicActivity

B

Food
Nutrient
Pasta
Meet
Vegetable
Carbs
Protein
Lipid

Pasta isA Food
Meet isA Food
Vegetable isA Food
Carbs isA Nutrient
Protein isA Nutrient
Lipid isA Nutrient

C

Person
ConsumedFood
Recipe
Diet
DietRule

hasFood
hasNutrient
amount
hasId

Food hasNutrient Nutrient
ConsumedFood hasFood Food
ConsumedFood amount <double>
Recipe hasFood Food
Recipe amount <double>
Diet hasId <string>
DietRule hasFood Food
DietRule amount <double>

$$\text{Coverage} = B/\alpha = 14/21 = 0.67$$

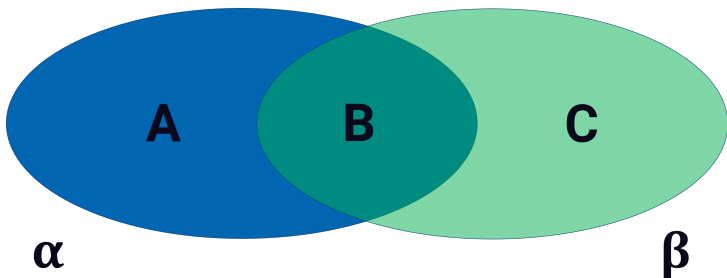
Coverage

The **Coverage** is computed as the ration between the intersection of α and β and the whole α sets: $Cov = B/\alpha$

About the **Coverage**:

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

Flexibility



The **Flexibility** is computed as the ration between the part of β not defined in α and the whole α sets: $Flx = C/\alpha$

Flexibility



A

EconomicActivity
Agriculture
Fishing
Feeding

Agriculture isA EconomicActivity
Fishing isA EconomicActivity
Feeding isA EconomicActivity

B

Food
Nutrient
Pasta
Meet
Vegetable
Carbs
Protein
Lipid

Pasta isA Food
Meet isA Food
Vegetable isA Food
Carbs isA Nutrient
Protein isA Nutrient
Lipid isA Nutrient

C

Person
ConsumedFood
Recipe
Diet
DietRule

hasFood
hasNutrient
amount
hasId

Food hasNutrient Nutrient
ConsumedFood hasFood Food
ConsumedFood amount <double>
Recipe hasFood Food
Recipe amount <double>
Diet hasId <string>
DietRule hasFood Food
DietRule amount <double>

$$\text{Flexibility} = C/\alpha = 17/21 = 0.81$$

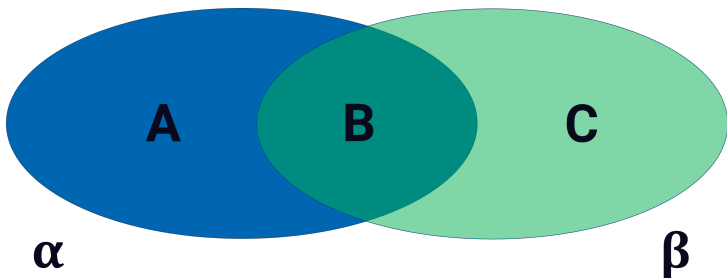
Flexibility

The **Flexibility** is computed as the ration between the part of β not defined in α and the whole α sets: $Flx = C/\alpha$

About the **Flexibility**:

- Any values within the interval $[0, +\infty]$.
- High values of **Flexibility** mean that the reference schema has been extended significantly.
- Low values of **Flexibility** mean that the reference schema has been extended by a small amount.

Extensiveness



The **Extensiveness** is computed as the proportional amount of knowledge provided by β with respect to the whole knowledge defined in the graph: $Ext = C/(\alpha + \beta)$

Extensiveness



A

EconomicActivity
Agriculture
Fishing
Feeding

Agriculture isA EconomicActivity
Fishing isA EconomicActivity
Feeding isA EconomicActivity

B

Food
Nutrient
Pasta
Meet
Vegetable
Carbs
Protein
Lipid

Pasta isA Food
Meet isA Food
Vegetable isA Food
Carbs isA Nutrient
Protein isA Nutrient
Lipid isA Nutrient

C

Person
ConsumedFood
Recipe
Diet
DietRule

hasFood
hasNutrient
amount
hasId

Food hasNutrient Nutrient
ConsumedFood hasFood Food
ConsumedFood amount <double>
Recipe hasFood Food
Recipe amount <double>
Diet hasId <string>
DietRule hasFood Food
DietRule amount <double>

$$\text{Extensiveness} = C/(\alpha+\beta) = 17/38 = 0.45$$

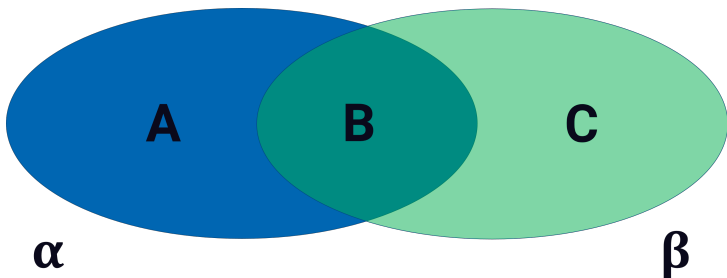
Extensiveness

The **Extensiveness** is computed as the proportional amount of knowledge provided by β with respect to the whole knowledge defined in the graph: $Ext = C/(\alpha + \beta)$

About the **Extensiveness**:

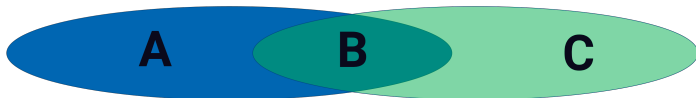
- Values are always within the interval $[0,1]$.
- High values of **Extensiveness** mean that the contribution of the created knowledge graph is predominant with respect to the content of reference schema.
- Low values of **Extensiveness** mean that the contribution of the created knowledge graph is limited.

Sparsity



The **Sparsity** is computed as the sum among the percentage of α not defined in β and vice-versa: $Spr = (A + C) / (\alpha + \beta)$

Sparsity



A

EconomicActivity
Agriculture
Fishing
Feeding

Agriculture isA EconomicActivity
Fishing isA EconomicActivity
Feeding isA EconomicActivity

B

Food
Nutrient
Pasta
Meet
Vegetable
Carbs
Protein
Lipid

Pasta isA Food
Meet isA Food
Vegetable isA Food
Carbs isA Nutrient
Protein isA Nutrient
Lipid isA Nutrient

C

Person
ConsumedFood
Recipe
Diet
DietRule

hasFood
hasNutrient
amount
hasId

Food hasNutrient Nutrient
ConsumedFood hasFood Food
ConsumedFood amount <double>
Recipe hasFood Food
Recipe amount <double>
Diet hasId <string>
DietRule hasFood Food
DietRule amount <double>

$$\text{Sparsity} = (A+C)/(\alpha+\beta) = 24/38 = 0.63$$

Sparsity

The **Sparsity** is computed as the sum among the percentage of α not defined in β and vice-versa: $Spr = (A + C)/(\alpha + \beta)$

About the **Sparsity**:

- 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 .

 **W3.L6.M3.T7** **iTelos - Evaluation**