KDI : Knowledge and Data Integration

iTelos - Evaluation

W3.L6.M3.T7

Contents

Contents

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

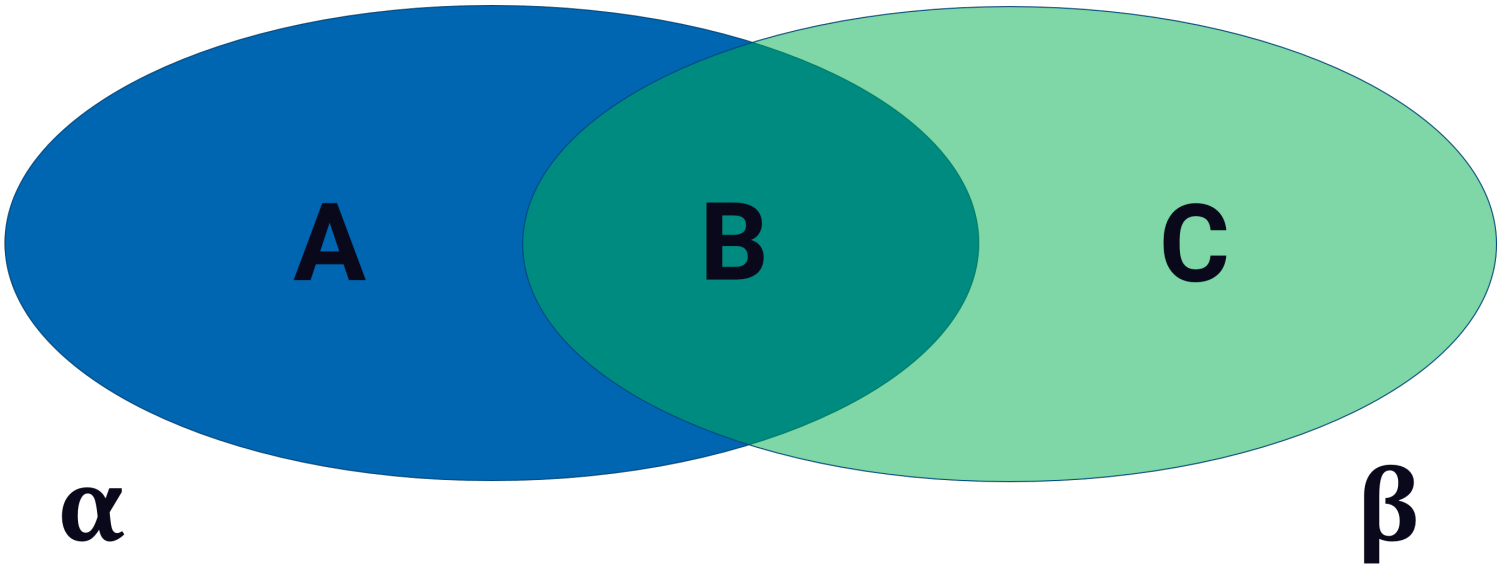
- M_E is the set containing the entities extracted from the model defined by the Knowledge Engineers. The M_E set has the three subsets M_E^{CR}, M_E^{CM}, and M_E^{CN} representing respectively the object properties of the Core (CR), Common (CM), and Contextual (CN) entities.

- 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.

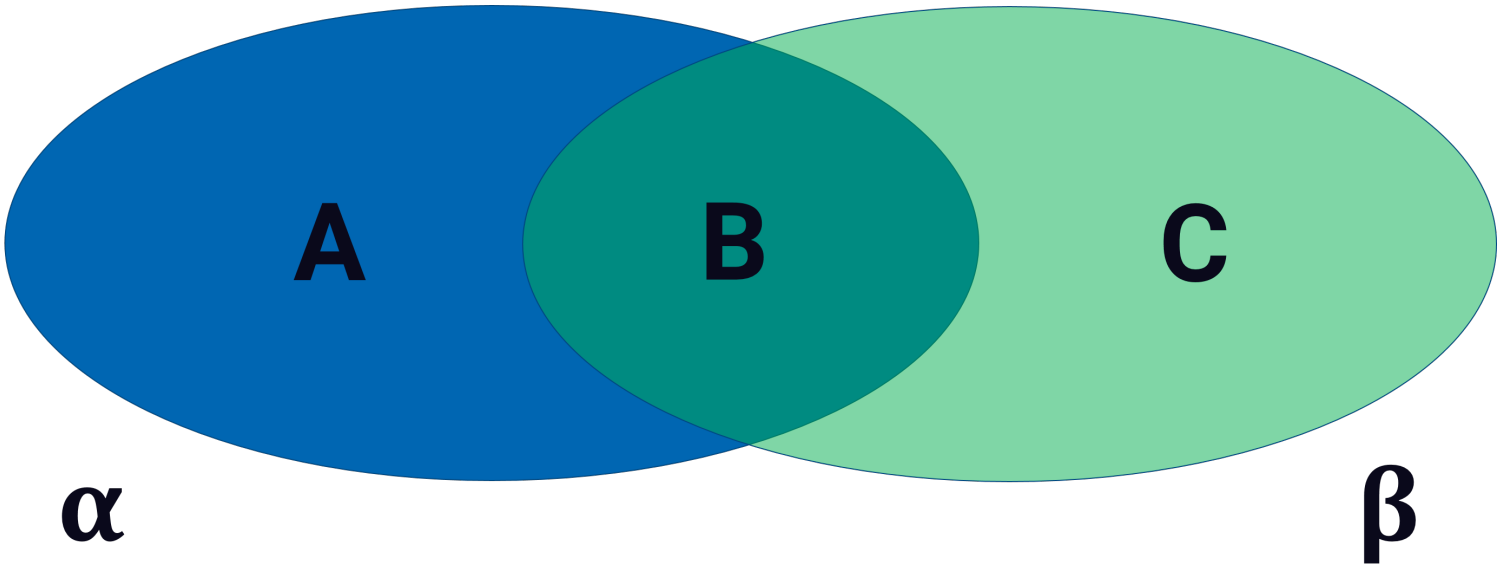
Contents

Metrics



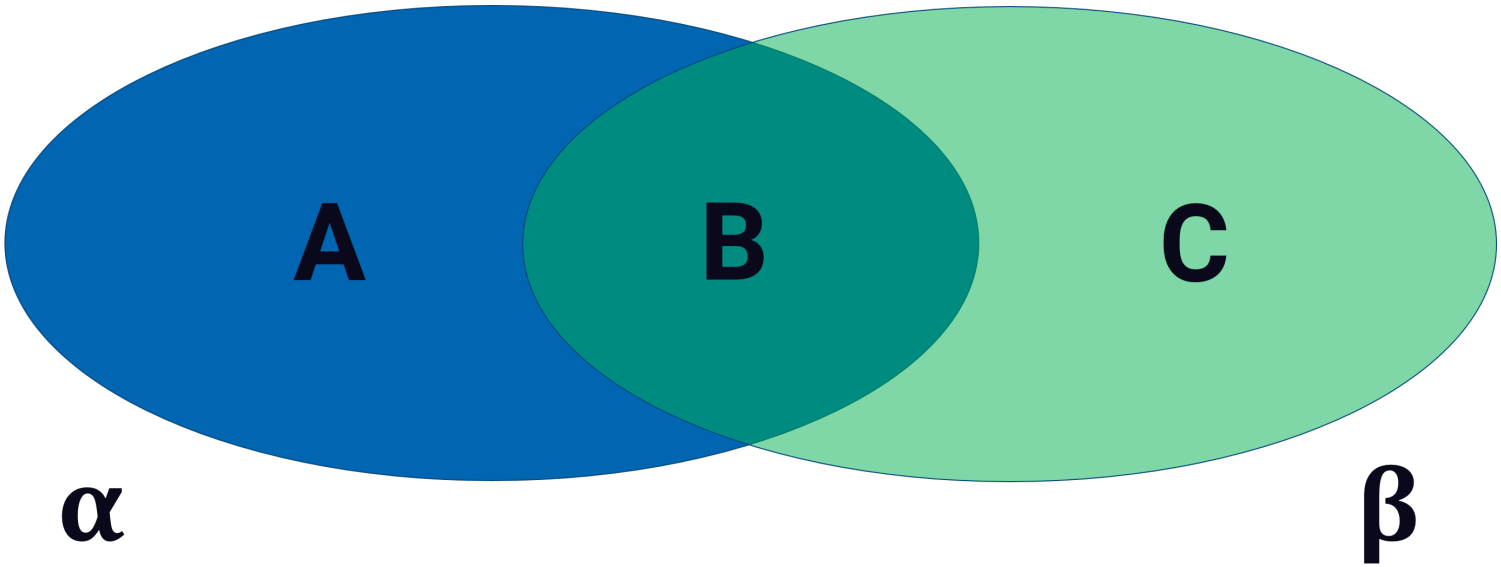
- Coverage (Cov)
- Flexibility (Flx)
- Extensiveness (Ext)
- Sparsity (Spr)

Coverage



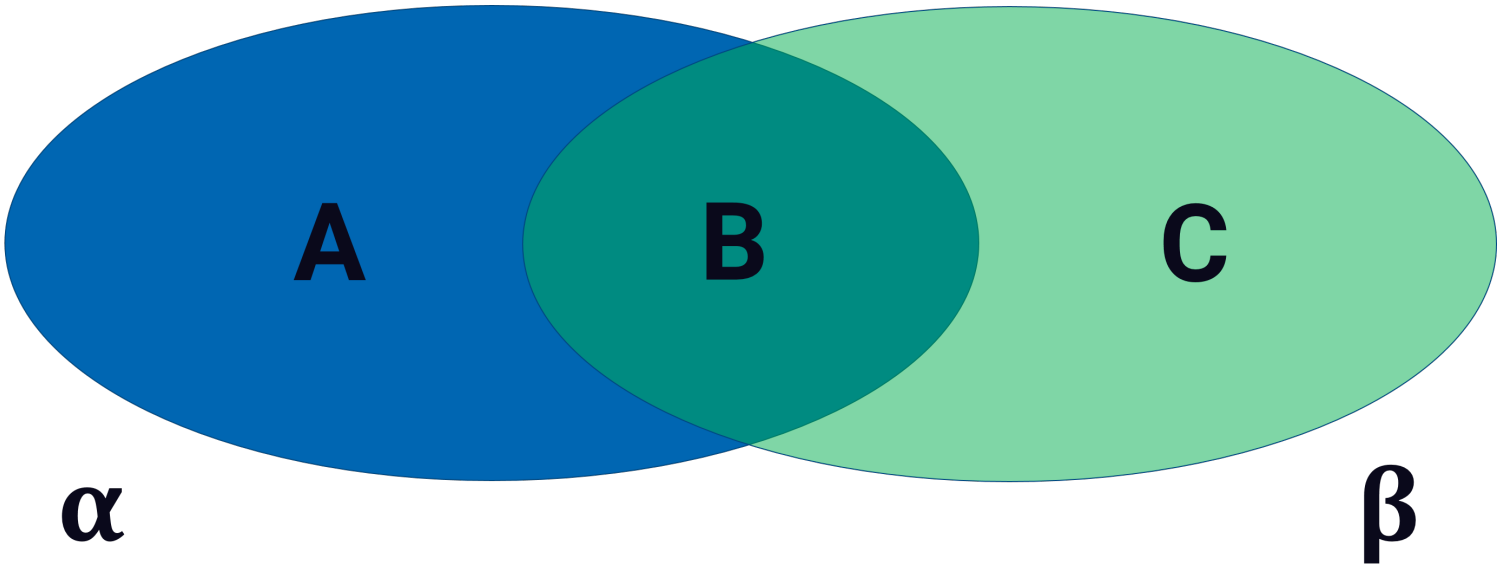
The **Coverage** is computed as the ration between the intersection of α and β and the whole α sets: $Cov = B/\alpha$

Flexibility



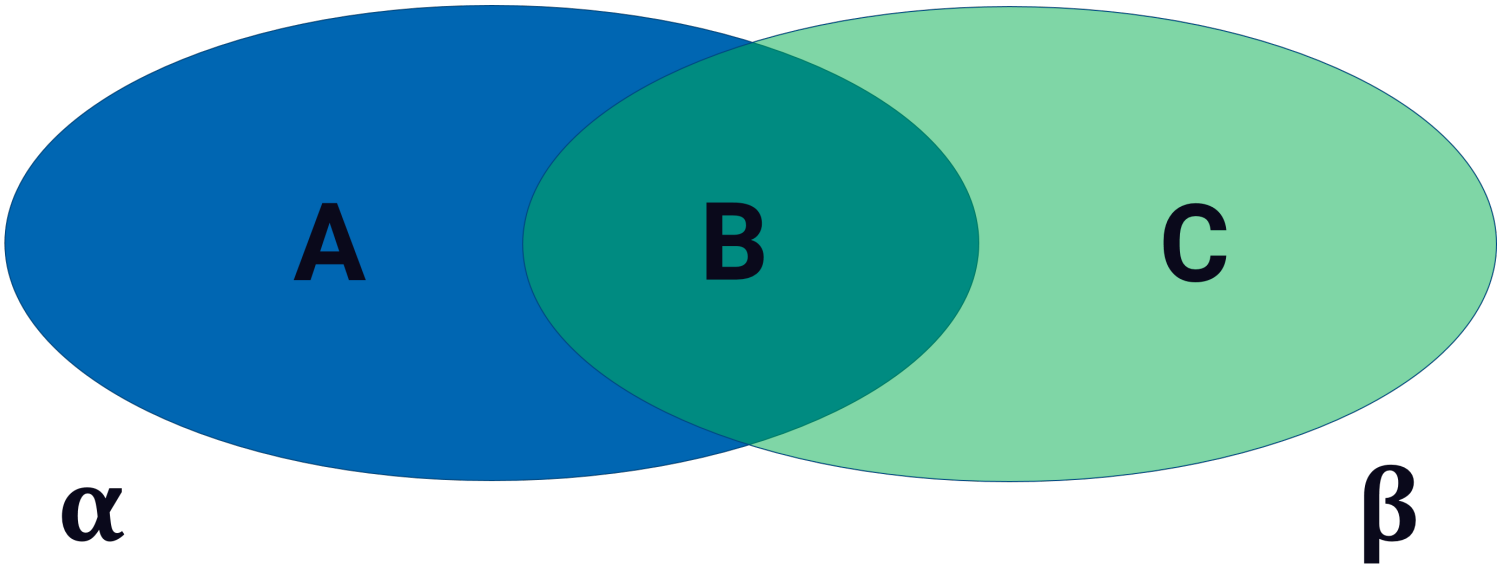
The **Flexibility** is computed as the ration between the part of β not defined in α and the whole α sets: $Flx = C/\alpha$

Extensiveness



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

Sparsity



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

KDI : Knowledge and Data Integration

W3.L6.M3.T7

iTelos - Evaluation