KDI **Knowledge and Data Integration**

Eliciting Concepts from Competency Questions (CQs)
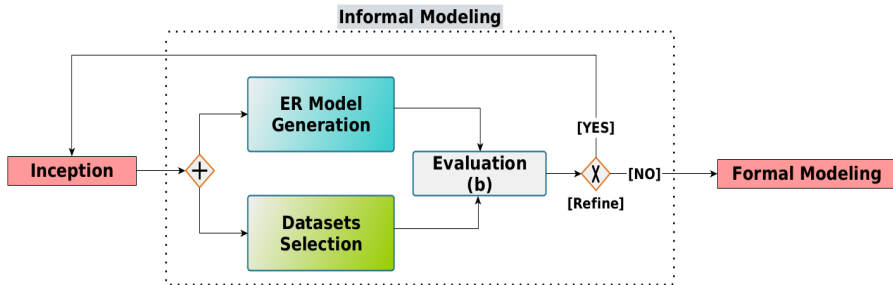iTelos Informal Modeling Phase

Fausto Giunchiglia, Mayukh Bagchi

Contents

Informal Modeling phase

Informal Modeling objective

Informal Modeling is the second iTelos phase

Inputs:

Outputs:

- Classified Competency Question (CQ) list.
- Datasets.
- Reference teleologies.

- ER Model.

- Selected datasets.

Knowledge layer: the KE aims to generate a model for the ETG, which is as suitable as possible for the information in the datasets selected. To achieve such a result the current phase is divided in two internal activities:

- Finalize the CQs formalization (started in Inception phase).

- Generate the ETG model.

Data layer: Moreover, the DS selects, form the whole set of datasets collected, only the relevant elements, filter away useless resources.

Contents

ER Model Generation activity - overview

1. The knowledge layer Informal Modeling internal activity completes the CQs formalization process stated in the previous phase.

2. Starting from the classified CQs (received in input) the *ER model generation* activity aims to extract the concepts that identify the entity types and properties, required to represent the information within the datasets collected.

3. Then the ETypes and properties extracted are used to build the ER model defining the purpose specific **ontology** able to represent the information we need to integrate and exploit

---

Observations:

- The ER model produced in this phase aims to represent the purpose specific information extracted from the datasets. Nevertheless, although the ER model design follows the teleology theory, it is not yet *shareable*(and, as a consequence *reusable*) enough.

- Only in the next phase, the ER model will be used, with the reference teleologies available, to build the ETG by composing the ontology, designed in the ER model, with the Foundational Teleology, becoming so a concrete shareable **teleology**.

Competency Questions (CQs) - Recap

- Given the purpose and motivating scenario, a set of natural language (informal) Competency Questions (CQs) will arise.

- We can consider these CQs to be expressiveness requirements that are in the form of questions

- The ETG must be able to represent CQs using its terminology, and be able to characterise the answers to these questions using the axioms and definitions

- Ideally, the CQs should be defined in a stratified manner, with higher level questions requiring the solution of lower level questions

- It is not a well-designed ETG if all competency questions have *only* the form of *simple* lookup queries

CQs formalization - Informal Modeling

The CQs formalization process is composed by 5 steps executed in the first two iTelos phases:

Inception phase:

1. **Raw CQ** (= informal CQ as elicited from purpose/objectives)
2. **Kernel CQ** (= Raw CQ minus all the auxiliary/apparatus words, resulting in each term denoting a concept)
3. **Analysed CQ** (= each concept in Kernel CQ classified as common, core or contextual concepts)

Informal Modeling phase:

4. **Classified CQ** (= each concept in Analysed CQ further classified as Objects, Functions and Actions)
5. **Attributed CQ** (= each concept in Classified CQ enriched with required object properties and data properties)

CQs formalization - Classified CQ

- **4. Classified CQ** (= each concept in Analysed CQ further classified as Objects, Functions and Actions)

  - *For (1):*
    1. **COMMON:-** OBJECT: Person, Establishment
    2. **CORE:-** FUNCTION: Hostel, ACTION: Consierge
    3. **CONTEXTUAL:-** FUNCTION: MS Student

  - *For (2):*
    1. **COMMON:-** OBJECT: Person, Establishment
    2. **CORE:-** FUNCTION: Hostel
    3. **CONTEXTUAL:-** FUNCTION: PhD Student

Note: The terms used as example above, are the same as the example used in the first three steps of CQs formalization process, in the Inception phase, on food and accommodation facilities in Trentino (see Slides on Inception phase).

CQs formalization - Attributed CQ

- **5. Attributed CQ** (= each concept in Classified CQ is used to define the Entity Types (ETypes) and relative data and object properties. For example, For example,

    - *For (Person):*
        1. *fiscalCode*
        2. *birthDate*
        3. *hasPersonProperFunction*
        4. .....

    - *For (Establishment):*
        1. *address*
        2. *areaServed*
        3. *globalLocationNumber*
        4. .....

Note: The terms used as example above, are the same as the example used in the first three steps of CQs formalization process, in the Inception phase, on food and accommodation facilities in Trentino (see Slides on Inception phase).

Data and Knowledge layers alignment

> The Purpose formalization Process is **data driven** !

- The CQs, as well as the kernel concepts and ETypes and Properties, are produced following the five-steps formalization process which **always considers the information available on the datasets collected**. We cannot consider a kernel concept extracted from a CQ that cannot be covered by any data collected.

- The ETypes with their data and object properties will be used to create the ER model, that will be in turn modeled on data.

Notes

- Unlike the first two steps of the CQs formalization process that are leaded by the DE, the last two steps are **performed by the KE** which has the right knowledge representation experience to properly model the concepts, and relative properties, following the teleology theory.

- Practically the two formalization process steps, considered in the Informal Modeling phase, are performed with the support of a specific spreadsheet, called **ModelingSheet**.

- The *Datasets Selection* activity strongly considers the ER model produced to achieve its objective. It aims to select, from the whole set of datasets collected in the previous iTelos's phase, only those which carry useful information w.r.t. the initial Purpose.

Contents

Summary

- We learnt about the general methodology for eliciting concepts from CQs, finalizing the CQs formalization process stated in the Inception phase.

- Within the methodology, we learnt how to sequentially classify the concepts so that it maps to the foundational teleological distinctions in formal modelling phase

- We saw how our methodology can be applied in real-life knowledge modelling via a small CQ from the domain of facilities for food and accommodation in Trentino

- THANK YOU !!!

KDI : Knowledge and Data Integration

Fausto Giunchiglia, Mayukh Bagchi

Eliciting Concepts from Competency Questions (CQs)
iTelos Informal Modeling Phase