KGE - Knowledge Graph Engineering

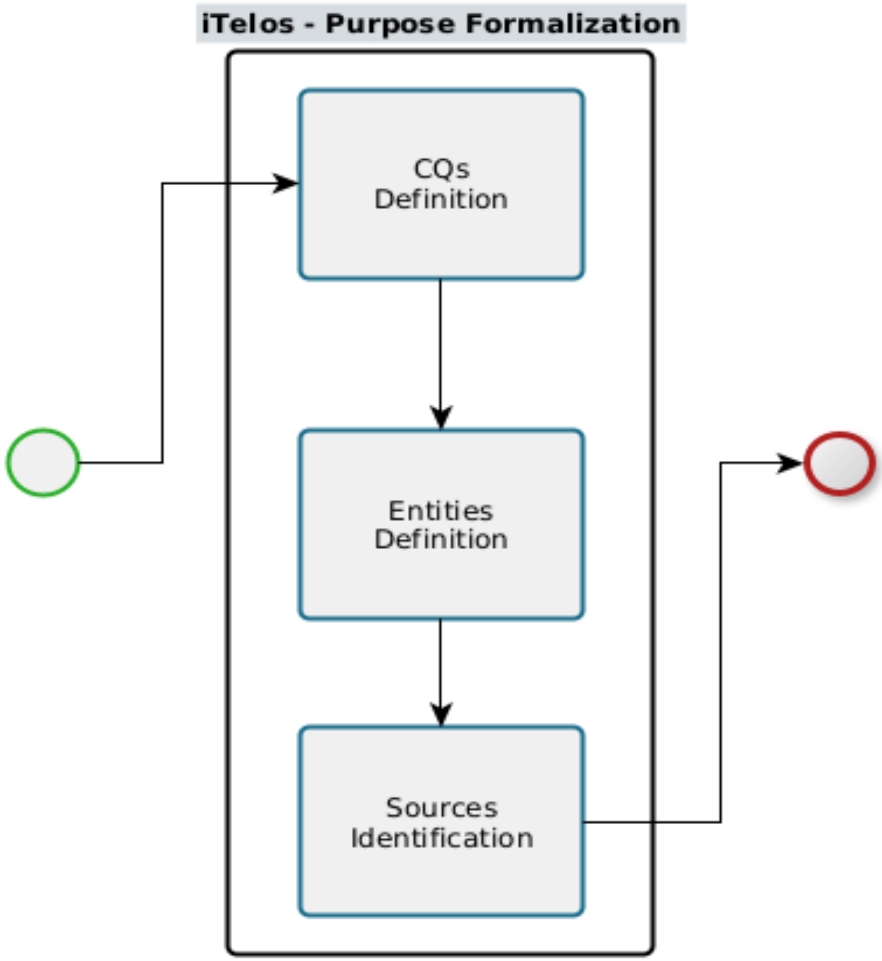
iTelos Methodology

Phase 0 - Purpose Formalization

Fausto Giunchiglia

Contents

Purpose Formalization



- **Input**: a natural language sentence representing the user's Purpose (plus optionally a list of already identified data sources).
- **Output**: a set of document in which Purpose's details are extracted and formalized (a first formalization step).
- **Objective**: to make explicit (in a more formal way) the functional requirements implicitly included in the input Purpose.

Contents

Purpose Formalization - Input

- **The Purpose**:

  "A person living in Trento, wants to easily access the best medical care she needs, by declaring specific symptoms."

- **Data source list**:
  - (data) Hospital A
  - (data) Hospital B
  - (data) OpenStreetMap
  - (knowledge) FHIR
  - (knowledge) Schema.org

Are these sources enough to satisfy the Purpose ?

We need to understand the Purpose better ...

Contents

Purpose Formalization - Activities

- The Purpose Formalization activities regard mainly **documentation** which helps to get the required information.
- The activities of the first methodology's phase, are thus executed by filling dedicated section on the **Project Report document**.

Purpose Formalization - CQs Definition

- Competency questions definition is the first activity in Purpose formalization phase.
- The objective is to **extract and explicate**, from the input Purpose, all the information we need to built the KG.
- What is a Competency Question ?
- They are the **functional requirements** of the KGE project.
- *"These requirements, which we call competency questions, are the basis for a rigorous characterization of the problems that the enterprise model is able to solve"* [1]

[1] Gruninger, M., Fox, M.S.: The role of competency questions in enterprise engineering. In: Benchmarking—Theory and practice, pp. 22–31. Springer (1995)

CQs Definition - DoI, Scenarios & Personas

- With this objective, CQs definition activity is composed by the following tasks:
  - **Domain of Interest (DoI)**: definition of the domain considered by the KGE project, specifying its boundaries in space and time.
  - **Scenarios**: definition of a set of scenarios, defined in the DoI previously specified, describing the possible use cases environments.
  - **Personas**: description of a set of actors/personas, having specific needs to be satisfied within the scenarios defined above.

Purpose Formalization
Competency Questions

- DoI, scenarios and personas, are used to define the CQs.
- A CQ is a natural language sentence, describing a possible use case that the final KG has to support.
- Within a CQ, usually, we can find a **persona** acting in a specific **scenario** within a **domain**, with the objective to **satisfy a specific need**.
- **Each CQ aims to describe a specific requirement over the final KG !**
- The set of **CQs should be as much as possible heterogeneous**, trying to specify all the possible requirements to be satisfied.

Purpose Formalization
Competency Questions - Examples

- Let's use the same example used above providing the Purpose and data sources list.
- **DoI**: The KGE project considers as geographical space, the municipality of Trento (Italy) over a period of time of two years (2020 and 2021)
- **Scenarios**:
  - Scenario 1: Trento city center, during the weekend.
  - Scenario 2: Trento city center, during the week.
  - Scenario 3: Trento suburbs, during winter, night time (from 20:00 to 05:00).
  - ...

Purpose Formalization
Competency Questions - Examples

- **Personas**:
  - Anna is a 27 years old student living in Trento, she loves camping in the mountains around the city.
  - Mario is a 67 years old farmer living out of Trento. He often goes in Trento to follow a specific medical treatment.
  - ...
- **CQs**:
  - CQ-1: Anna was injuried Saturday while camping with friends in mount Bondone (scenario 3). She needs emergency assistance.
  - CQ-2: Mario comes in Trento for a specific surgery to be executed in Santa Chiara Hospital (scenario 2).
  - ...

Purpose Formalization
Entities Definition

- The next step in Purpose Formalization phase, is to **extract**, from the CQs defined, all **those terms representing the most important entities** to be included in the final KG.
- The entity terms extracted will be the base elements used to build the KG's knowledge structure (ETG).
- In order to provide more information about the level "popularity" (regarding their possible re-use in other projects) which these entities will have in the final KG, in this activity **iTelos classifies the entity terms into Common, Core and Contextual** categories.

Purpose Formalization
Entities Definition - Example

- **Common entities**:
  - CQ-1: Location, ..
  - CQ-2: Location, Move, ..
- **Core entities**:
  - CQ-1: Mount, Injury, ..
  - CQ-2: Hospital, ..
- **Contextual entities**:
  - CQ-1: Emergency, ..
  - CQ-2: Surgery, ..

Note: The more details are reported in CQs, the easier will be to identify the entities terms, as well as to classify them.

Purpose Formalization
Sources Identification

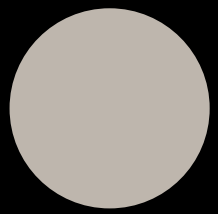
- The previous activities extracted several details from the initial user's Purpose.
- The last activity, in Purpose Formalization phase, aims at **exploiting such an information to integrate the list of data sources** (optionally) received in input for the project.
- Often, the data source list initially provided **is not sufficient** to cover all the Purpose's requirements.
- Of course, if the data source list is missing for the project, **this activity plays the crucial role to identify the sources required**.

Contents

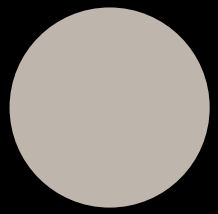
Purpose Formalization - In practice

- The Purpose formalization phase (as already mentioned) can be fully executed by filling the Project Report document.
- Nevertheless, using **iTelos application** the domain of interest as well as the personas can be defined using the application itself, thus producing a **semi-structured version** of such information.
- See the next lesson for an example of using the iTelos application.

KGE - Knowledge Graph Engineering



Fausto Giunchiglia



iTelos Methodology

Phase 0 - Purpose Formalization