Université Jean Monnet, Saint Etienne
Master in Computer Science
Cyber-Physical and Social Systems

Project in Virtual Communities

Pierre Maret

January-March 2020

Title: Elaborate a method for domain-specific wikibase instantiations

Communities of interest group people dealing with the same topic of interest whatever their various backgrounds. Wiki-bases are adapted tools for these communities because they offer a way for a) to edit, store and access information easily.

The difficulties in the design and implementation of such a system are several:

- initiate the base with a multidisciplinary consensual domain description,
- implement the means to extract and store the data,
- implement the means to query and exploit the data.

Today, different tools or components exist that make wiki-bases possible. However, no method exist to guide an interdisciplinary team to elaborate such a system.

Among the components that are available, Wikidata (http://wikidata.org) is the most well-known wiki database example. It targets general "public knowledge". The wikidata project is open source and belongs to the Wikimedia foundation. Wikidata uses Semantic Web concepts, i.e. RDF data format (which is graph based and uses triplets Subject-Property-Object).

QAnswer (http://qanswer.eu) is also a convenient component for querying wikibases. This technology enables querying in free text any RDF database. It avoid users to learn writing SPARQL queries.

The goal of a method to elaborate is to guide a group of interdisciplinary team members to create, fill and exploit a wikibase dedicated to a given domain. The method will consist of steps including for instance task descriptions (input, content and output), tasks' flow (parallel/sequential tasks), events and coordination points, etc. Recommendations, design tools and software programs can be associated a task or to the entire method. The method should lead the team to the concrete implementation and exploitation of the wikibase.

The method must

1. help the team to organize and structure the chosen domain and identify data sources.
2. help how to create the instance of the wikibase.
3. help implementing ways to insert data into the base (extracts from wikidata + other data sources). Data insertion can be manual and automated (extract/insert data from databases or from text).
4. help the team to implement a search engine to query data. Wikibases can be queried with text by means of the QAnswer technique (http://qanswer.eu)
5. help the team to implement in the background a monitoring system in order to observe the requests and needs of users of the wikibase for further improvements.

Notice that the method should also try to respond (even partially) to issues such as the evolutions/extensions of the description of the covered domain (step 1), the integration of new data sources and/or the implementing innovative data extraction methods (step 3), and the evolutions of the QAnswer technique.

For your project, you have first to discover the wikidata approach and its tools (you can start with these pages: https://www.wikidata.org/wiki/Special:SpecialPages, https://www.wikiba.se/, and search for others!). You must select an application domain and find examples of heterogeneous data sources for it (so that you can elaborate/test your method). You have also to learn how to implement QAnswer. Each task in the method must be documented. Build your own experience, and then describe the tasks.

You will progressively describe the method and create/collect software pieces relevant to support it. Every second week you will present (slides + document) the current version of your method and the remaining work yet identified.

Agenda

Sessions	Expected results presented during the session	Content of the session
16/01		Project initiation
23/01		Work on the project
30/01	Presentation of the method V1 Document for the method V1 ; Source code for V1; Remaining work	Presentation Work on the project
6/02		Work on the project
13/02	Presentation of V2 (revision and improvements of V1) Document for V2; Source code for V2; Remaining work	Presentation Work on the project
20/02		Work on the project
27/02	Presentation of V3 (revision and improvements of V2) Document for V3 ; Source code for V3; Remaining work	Presentation Work on the project
12/03		Work on the project
19/03	Presentation of V4 (revision and improvements of V3) Document for V4 ; Source code for V4; Perspectives	Oral presentation
26/03		Written exam

For each presentation or document, you can propose, augment and revise:

- Identification of the different tasks or steps
- Experience in task executions (for the chosen example)
- Overall method process description
- Single step's description, with associated tools