Knowledge and the Web 2015/2016: Homework 2

October 22, 2015

This homework is a continuation of the task you have started during the third exercise session. The **goal** of the homework is to **design a data model** needed to answer a certain question using Semantic web technologies. The main knowledge base used for this homework is the Talk of Europe (ToE).

To successfully finish this homework you have to complete several parts:

1. **Think of a question** to answer, based on the content of the ToE knowledge base. Take the position of a data analyst, journalist, citizen, politician, researcher, etc. and think of a research question which you would like to answer. This research question should involve a non-trivial data analysis using information from ToE, enriched with further source(s). **Note:** This is not a trivial task, and we will refine these questions through several iterations. At the moment, focus on a topic that you find interesting, then take it from there in the next homeworks.

2. **Find or invent a knowledge base** that contains the answer you want to find. It can be an existing one (DBpedia, Yago, Freebase), or you can create an imaginary/fake one that contains an answer you need.

3. **model the necessary data** from both knowledge bases in `RDFS` using `Protege` (exercise session 2). **Note:** if you are inventing a new knowledge base, be sure to define its partial vocabulary (ontology)

Practicalities

The **deadline** for the homework is Tuesday **November 3**, 2015, 23:59h. The homework should be completed in **groups of 4**. Further through the course, you will be working on your projects in the same groups. The submission should contain:

1. Protege file modelling your data model (`RDFS`)

2. a report containing:

   - list of the members of the group
   - explanation of your question
   - description of the vocabularies you use. You only have to explain the parts you need, both from the ToE and the other knowledge base of choice
   - Note any issues you expect (such as availability of data, data quality issues, ...) and a first plan for how to address them.

Only one member should submit the homework for the entire team. The homeworks should be submitted by e-mail to Bettina Berendt. Feel free to ask any questions about this assignment in class, on Toledo or by e-mail to Sebastijan Dumancic (be sure to put [KaW] in the subject line).