Self-reflection report

Knowledge and Data Engineering — Group Project

Aishwarya Agarwal

My role was to map and uplift data into RDF Triplestore. To do the mapping, first I did some data cleaning like eliminating the unwanted columns from the dataset, splitting a column where required and replaced all the '-' values with 0 for players' performance score attributes. I also added data for the City where each of the Stadium is located in csv. After that I did the R2RML mapping using JUMA initially and later using Text Editor. I converted the data into RDF triples using the R2RML engine provided on Blackboard and uplifted the triples into the Triplestore. Along with data mapping and uplifting I helped in Ontology creation and created few object properties that are Transitive, Symmetric or Inverse in nature. I also contributed to SPARQL query creation and created R2RML documentation.

Strength/Weakness: For this Group Project, our group was well organised and we all worked well together. The approach we followed was to divide work not necessarily according to role and complete our tasks and then discuss with each other to improve on our Ontology, mapping etc. This approach helped all of us to remain in sync with all the areas of the project. My strength was that after getting familiar with creating R2RML mapping manually, I was able to incorporate the changes quickly in the mapping with reference to all the modifications done in our Ontology model and uplift data into Triplestore. This project helped me gain practical knowledge in data uplifting, mapping and ontology creation. While we were able to complete almost all the requirements for the project but initially we created a very basic Ontology model and had to run many iterations of data mapping and uplifting to fulfil all the requirements which was a time consuming process. Also since we were not very familiar with tools like JUMA, Protégé and GraphDB, we had invested a lot of time in exploring these and still were not able to utilize them to their full capacity which would have helped us in enhancing our Ontology model and mappings.

Boris Flesch

Generally speaking, we handled this project as a united team and were always working together during our many meetings that we all systematically assisted. Therefore, even though leads on specific tasks have been assigned to each of us, we have mainly been working together on the different tasks (i.e. finding datasets, creating the ontology, creating the triplestore, etc.).
As "Lead on UI and Query Design", my major contribution to this project is the development of the Application Query Interface (i.e. web-application) that can be used to query our triplestore. I also worked on many SPARQL queries along with other team members.

A strength for this project was my previous experience with web-application development. Even though it was my first time using Python Flask (which seemed convenient with SPARQLWrapper library), I had a good knowledge of web-development languages and Bootstrap framework for UI. I also tried to be as proactive as possible and remained available throughout the whole project, as all of us did.

However, I have sometimes been focusing on specific facets of the project without considering other components enough (i.e. ontology, R2RML mapping, triplestore, etc.). Thus, it took me some time to get used to the interactions between all the components that were implied; which is also due to a lack of experience in this domain of data engineering.

Sherwin Mascarenhas

My main role was working on the Ontology and incorporating existing Ontology to our Ontology. Finalizing the Ontology was mostly a group effort, but I looked at the extra bit of verifying if some of the properties and classes we have created worked well with the rest of the ontology. I also succeeded in bringing in existing ontology by searching the web for popular Onotlofies and vocabularies whose entities could be reused and then went ahead and searched for specific classes that worked well with our ontology all along seeing that it did not break the existing ontology design. I also took charge of the technical report and worked on some of the queries.

Strengths - As a group, we worked really hard, but most importantly we worked very well together. Everyone understood his or her responsibilities and took up the task willingly. Everybody was here to learn and can confidently say that we have finished the project doing more than just learning a lot. I believe I put in all the effort I could in doing this project and doing it well. I was able to push for answers to questions we were stuck with by raising questions with the TA and by posting on the clinic.

Weaknesses - There weren't a lot of weaknesses with the group. The only problem with the group was not being able to collaborate with each other in person. My weakness was not being able to map all the classes with the DBpedia. I also believe that I wasn't able to put as much time on other aspects of the project like manual r2rml mapping as I could due to my disability.

Niejun Yin

My main role in the team is in charge of Data selection.
At the beginning of the project, I searched database resources, provided them for discussion, and participated in the evaluation of databases provided by other members. Finally, we obtained suitable datasets. Then I participated in the ontology design, including trying to reuse existing vocabulary and ontologies, creating some specific properties. Trying to use juma for mapping design, participating in the design of the questions and Sparql statements.

strengths:
Participate in each stage of the project, learn a lot from the processing and our teammates, provide some ideas for linking the dataset and reuse existing ontologies and other steps.
weaknesses:
When the first time we selected and evaluated the database, the mapping process was not fully considered. The database I provided was mostly numerical and inconvenient for expansion.

Rui Xu

My main contribution is in competency questions design and csv data processing, such as converting the date format in csv to xsd:dataTime and the unification of the team name of the competition. I also contribute to the ontology design part and the query part.

Strengths
All initial questions can be answered from the combined datasets, saving time for our query part. After designing the questions, there are several issues related to the date column, because the SPARQL statement will involve the comparison of dates. I noticed that the date format in our csv cannot be compared correctly when querying in SPARQL. So I changed them to the xsd:dataTime format, and used the xsd:dataTime function in SPARQL to complete the date comparison, which made our ontology more in line with the specification.

Weaknesses
When designing problems, I associate SPARQL queries more with queries in relational databases. In fact, SPARQL query RDF data has some other features. After group discussions, we added some questions, such as querying external data that is linked to our data.