**Title:** Building an entity store on top of a triple store

**Reference person:** Uladzimir Kharkevich ([uladzimir.kharkevich@disi.unitn.it](mailto:uladzimir.kharkevich@disi.unitn.it))

**Introduction**

A ***triple-store***[[1]](#footnote-1) is a generic database for the storage and retrieval of ***triples***: atomic facts in the form of ***subject-predicate-object***.

Many applications use triple-stores for storing, processing, and querying facts about the ***real world entities***. For instance, one might need to store information about some entities (e.g., people, locations, and organizations), their attributes (e.g., age, name, and description) and relations between them (e.g., people work in the organization which is located at some place). Also at the query time, one might need to search for entities which satisfy some specific criteria (e.g., people of a given age, or people who work at some place, or organization which is located in some country) or one might search for some facts about entities (e.g., how many people are working in some company).

The described above tasks are entity-centric and they require translation to more generic (triple-centric) standards and implementation. The exercise suggested in this proposal is to see how an ***entity management system***[[2]](#footnote-2) (i.e., the system specifically designed to work directly with entity types and entities and not with triples) can be implemented by building intermediate entity management layer on top of available general purpose semantic web standards and tools (RDF(S)/OWL/SPARQL and triple stores).

**Goal of the project**

The goal of this project is to design entity storage (as a layer) on top of a triple store. The possible steps are:

1. To familiarize oneself with the notions of entity types, entities, attribute definitions, attributes.
2. To familiarize oneself with editing and storing RDF(S)/OWL data and querying it with SPARQL.
3. Design the RDF-based storage for entity types, entities, attribute definitions and attributes.
4. To familiarize oneself with the syntax of HQL[[3]](#footnote-3) and adapt it to be entity centric (e.g., entity types instead of classes and entities instead of objects).
5. Write mappings from the resulting entity centric HQL to the SPARQL queries against designed RDF-based entity storage.

The work has to be described in a final report providing (a) the description of the designed RDF-based entity storage, (b) the description of an entity centric (HQL-like) query language, (c) and mappings of its queries to the SPARQL. In addition, the report should provide (d) examples of 3-5 entity types and 10-20 entities as well as their RDF representations, (e) 3-5 entity centric queries with their translations to SPARQL.

**Goal of the thesis**

As a result of the proposed project above we will have a preliminary design for an entity storage implementation on top of a triple store (and its feasibility will be studied). The goal of master thesis will be:

1. To finalize and implement the designed architecture.

2. Extend entity centric query language to reuse the out of the box OWL/RDF(S) reasoning capabilities.

3. To integrate semantic (NLP/concept) search on string attributes.

**Requirements**

The candidate must study fast and be innovative ;). The knowledge of semantic web technologies is a plus.

**List of the background material to read**

Entity types will be discussed during the lectures of the course.

**RDF(S)**: http://www.w3.org/TR/rdf-primer/

**OWL**: http://www.w3.org/TR/owl2-primer/

**SPARQL**: http://www.w3.org/TR/rdf-sparql-query/

**HQL**: http://docs.jboss.org/hibernate/core/3.3/reference/en/html/queryhql.html

1. http://en.wikipedia.org/wiki/Triplestore [↑](#footnote-ref-1)
2. Details about entities, entity types, and entity management will be covered during lectures. [↑](#footnote-ref-2)
3. http://docs.jboss.org/hibernate/core/3.3/reference/en/html/queryhql.html [↑](#footnote-ref-3)