

Creation of an Index of Lemmatized Labels

Saeedeh Shekarpour, Dr. Axel-C. Ngonga Ngomo



Problem Statement

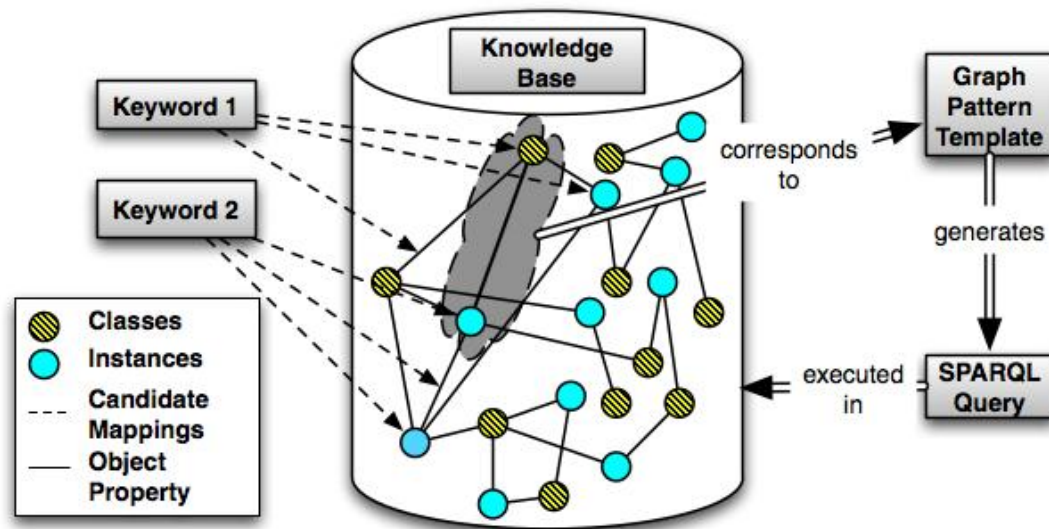
- ▶ Keyword-based search is a convenient way to query knowledge bases.
- ▶ SINA is a semantic search engine aiming to retrieve information from RDF knowledge bases in a manner close to question answering systems.



Search

Problem Statement

- ▶ One challenge of querying knowledge bases is finding the correct anchor points (i.e. classes, properties and instances).
- ▶ The first step of recognizing the relevant anchor points is string matching.



Problem Statement

- ▶ To increase the success of string matching we need to lemmatize the **rdfs:labels** or even maybe all literals contained in the knowledge base.
- ▶ For example , for the labels **director, directing, directed by**, the **lemma direct** should be found.
- ▶ If lemmatization is done on the fly (on the time of search process) , it will degrade the performance.

Object

- ▶ We need a tool which by choosing a graph of the underlying knowledge base:
 1. Lemmatizes all labels (literals) of the associating graph.
 2. Stores lemmas in the knowledge base (using a defined property)
 3. Make an index (Configuring Solr)over the new labels.
- 