ITI8600 Homework 2: Knowledge Representation

Group nr. 10:

Carl Maask
Antonio Gianola

Base task:

Making a program that is capable of reading a plain text file in English and printing the information read and understood in a logic-based formalism. The output should be resented in RDF.

Code:

The code is available at (https://gitlab.cs.ttu.ee/camaas/iti8600hw2).

Solution:

We started by creating the program and grabbing some example text to go with it. The sentences given to the program are the following:

- “Puise is a village and it is north of Matsalu bay”
- "Estonia is a county and Tallinn is in Estonia."
- "Tallinn is south of Puise and it is a city."
- "Milano is far from Tallinn."
- ”Rome is bigger than Tartu.”

At first we initialize the Spacy module with the Spacy English language model. We decided to use Spacy due to its’ simplicity and many functionalities. We also felt that it has plenty of documentation on usage.

We start our analysis process by converting pronouns in the text with their “correct” values. This means we replace pronouns by their values within the text. In the case of the given text, it converts the pronouns and changes two sentences:

- “Puise is a village and **Puise** is north of Matsalu bay”

- “Tallinn is south of Puise and **Tallinn** is a city."

After doing that we use a method to split the sentences to make the computing of the sentences simpler. The initially presented text looks like this after splitting:

- “Puise is a village.”
- “Puise is north of Matsalu bay.”
- "Estonia is a county.”
- “Tallinn is in Estonia."
- “Tallinn is south of Puise.”
- “Tallinn is a city."
- “Milano is far from Tallinn."
- “Rome is bigger than Tartu.”

Then we use a NLTK (Natural Language Toolkit) functionality to form a tree from the given nodes. After that we use a method to find the URIs of the words. For this purpose we crafted our own small method to find URIs for different types of words. For nouns we used DBpedia to return their wikipedia URL and for adjectives, adverbs and verbs we return the corresponding Conceptnet page. After finding and attaching the URIs, we eventually write it all into RDF format into a file.

Required downloads:

- Spacy (https://spacy.io/docs/usage/)
- Spacy english language model (https://spacy.io/docs/usage/models)
- Natural Language Toolkit (http://www.nltk.org/)