Knowledge and the Web: Homework 3

Katrien Laenen Gust Verbruggen Ward Schodts

November 12, 2015

1. Which European politicians have *a high chance* of receiving a Nobel Prize?

We emphasise *a high chance* because that could use some clarification. How do we compare probabilities of receiving Nobel Prizes between arbitrary politicians?

First, we build a profile of a politician. This should not only cover their performance in the European Parliament, other achievements and activities are probably far more interesting. Where are they from? What subjects have they worked on before or talked about in public? Have they received other awards? Perhaps they've written a book or performed research activities?

Next, we attempt to build an analogous profile for Nobel Prize winners. Even though prizes in Peace or perhaps Literature and Economic Sciences are far more likely to be won by a politician, we'll build the model for all winners.

The final step is comparing the profiles we built. Do we find particular similarities between a specific Nobel Prize winner and politician? Or, for example, we could try learning a model for a Nobel Prize winner trough machine learning and classify the politicians using this model.

Knowledge bases

Aside from the ToE knowledge base, we need to gather as much information as possible about the people we'll build profiles for.

DBpedia (http://wiki.dbpedia.org/) is always a good starting point. Coming straight from Wikipedia, it is updated daily and should thus be up-to-date all the time. For many projects, Wikipedia is *the* go-to data source, because it contains complete information about relevant topics to our subject.

Although it can be edited by anyone, data correctness is verified thoroughly by a team of content moderators.

Nobelprize.org (http://data.nobelprize.org/) is the most complete data source concerning nobel prize winners. From 1901 on, it contains all nobel prizes that have been won. Moreover, it is maintained by the official organisation responsible for the Nobel Prize and should thus be complete and correct. It makes use of the FOAF and DBpedia ontologies and more importantly, laureates have a `owl:sameAs` relation with DBpedia entries.

2. Which characteristics of a politician influence the amount of TV appearances he makes?

There is one specific characteristic that we really want to research. Namely: "Does the country of origin of the politician influence his TV appearences?".

3. Which politicians have a high chance of being murdered?

European Union Open Data Portal (https://open-data.europa.eu/) contains datasets about homicides and other criminal activities e.g. homicide offences. This data necessary to determine if location influences the chance of being murdered.

Freebase (https://www.freebase.com/) This is a knowledge base specialized among others in well-known people. Because most politicians are well-known this is a interesting knowledge base. Furthermore in Freebase is the property *cause_of_death* which we can use to find murders or homicides among the politicians.

DBpedia (http://wiki.dbpedia.org/) For this question we need also some background information. As already pointed out, DBpedia is a good source for this.