Multilingual syntactic analysis of persons/entities / Pertsona edo entitateen analisi sintaktiko eleanitza

Proposer(s) / Proposatzailea(k): Koldo Gojenola / Aitziber Atutxa

Contact / Kontaktua: koldo.gojenola@ehu.eus

Description / Deskribapena

The main idea of this project is to perform a syntactic analysis of the appearances of people (politician, scientist, sportsman/woman, …) in different media (newspapers, twitter, …) in several languages. This will allow to find new insights on the accomplishments of these entities, but also their different treatment according to language, media group, etc.

Universal Dependency tools, like the multilingual parser UDPipe (http://ufal.mff.cuni.cz/udpipe), could be used for the project.

Goals / Helburuak

These are the main objectives:

- Obtain a corpus of texts, taken from the web, that mentions persons (politician, scientist, sportsman/woman, …) or entities in several languages and media (newspapers, blogs, …)

- Process the text by multilingual syntactic analyzers

- Study the syntactic patterns depending on each entity, language and media

- For computer scientists, an automatic tool that allows the selection of different entities, languages and media and presents the patterns could be developed

Requirements / Betebeharrak

Depending on the student background, the project could be adapted. For example:

- Computer Science background: development of an automatic tool that, given a person/entity, obtains a multilingual syntactic analysis of this element in different media

- Linguistics background: selection of entities to be processed, media, and a linguistic analysis of the obtained results.

Framework / Esparrua

Tasks and plan / Atazak eta plana

- Specification of the work: requirements, hypotheses to be tested, tools, and corpora
- Analysis of the texts
- Analysis of patterns regarding entity type, language and media. Different criteria:
  - Associated verbs
  - Type of relation: subject, object, …
  - Associated complements and types
  - …
- Results and conclusions