Bootstrapping semantic lexicons for low-resource languages

Proposer(s) / Proposatzailea(k): Jeremy Barnes

Contact / Kontaktua: jeremy.barnes@ehu.eus

Description / Deskribapena

For many tasks, such as emotion detection or abstractness prediction, rich features coming from semantic lexicons are often more informative than trying to use deep learning to extract features from raw text. However, compiling these lexicons by hand requires large amounts of work, and most of the world's languages do not currently have access.

Goals / Helburuak

This thesis will contribute by exploring bootstrapping methods to automatically generate semantic lexicons for low-resource languages. Specifically, it would be of interest to compare hand-annotation, machine-translation, and cross-lingual embedding approaches. The precise details and scope of the thesis will be further decided in agreement between the supervisor and the candidate.

Requirements / Betebeharrak

The project presupposes a good balance of technical and linguistic expertise. Good programming skills, experience with machine learning and a solid background in NLP are relevant qualifications. Knowledge of several foreign languages would also be an advantage. Please contact the supervisor to discuss further details.