WordNets from scratch

Proposer(s) / Proposatzailea(k): names / izenak

German Rigau

Contact / Kontaktua: german.rigau@ehu.eus

Description / Deskribapena

Building large-scale knowledge resources such as local wordnets is very time consuming and requires large research groups and long periods of development (Gonzalez-Agirre et al. 2012). Lately, several new approaches have been devised towards its automatic development. For instance, Watset (Ustanov et al. 2017) have been used for the automatic induction of English and Russian synsets, and (Noraset et al. 2017; Gadetsky et al. 2018) propose different systems for automatically creating English definitions.

Goals / Helburuak

Try different alternative software available on English, Spanish, Basque, etc. for creating new wordnets from scratch.

Requirements / Betebeharrak

Basic knowledge of Linux command-line interface:

- execution of programs through the command line
- handling of text files

Framework / Esparrua

There are different systems to try and evaluate on different languages. We will need to decide the systems, corpora, gold-standards and metrics for its empirical evaluation.

Tasks and plan / Atazak eta plana

- Apply Watset (Ustanov et al. 2017) to Spanish, Basque, etc.
- Test the results using the wordnets already integrated into the MCR
- Try alternative systems (Noraset et al. 2017; Gadetsky et al. 2018) to generate word sense definitions automatically for Spanish, Basque, etc.
- Test the generated definitions with respect to reference ones.

References

Gonzalez-Agirre A., Laparra E. and Rigau G. *Multilingual Central Repository version 3.0*. 8th international conference on Language Resources and Evaluation (LREC'12). Istambul, Turkey. 2012.

Gadetsky A., Yakubovskiy I. and Vetrov D.. Conditional Generators of Words Definitions. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics ACL* 2018.

Noraset, Thanapon, Chen Liang, Larry Birnbaum, and Doug Downey. "Definition Modeling: Learning to Define Word Embeddings in Natural Language." In *AAAI*, pp. 3259-3266. 2017.

Ustalov, Dmitry, Alexander Panchenko, and Chris Biemann. Watset: Automatic Induction of Synsets from a Graph of Synonyms. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, pp. 1579-1590. 2017.