Automatic WordNet Domains

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, ask2Transformer (Sains and Rigau 2021) exploits different pre-trained Language Models for assigning domain labels to WordNet synsets without any kind of supervision.

Goals / Helburuak

Try different alternative approaches, knowledge sources of domain information and neural language models for assigning automatically domain labels to WordNet synsets.

Requirements / Betebeharrak

Basic knowledge of Linux command-line interface:

- execution of programs through the command line, python, etc.
- handling of text files

Framework / Esparrua

There are different alternative approaches, knowledge sources and neural language models for assigning automatically domain labels to WordNet synsets.. We will need to decide the knowledge sources, language models, gold-standards, etc. to improve the existing labelling.

Tasks and plan / Atazak eta plana

- to apply this approach to other sources of domain information such as WordNet topics and WordNet Domains.
- to explore the cross-lingual capabilities of pretrained Language Models for domain labelling of non-English wordnets and other lexi-cal resources.
- to explore the utility of these findings in the Word Sense Disambiguation task.

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.

Sains O. and Rigau G. Ask2Transformers: Zero-Shot Domain labelling with Pre-trained Language Models. Proceedings of the 11th Global WordNet Conference (GWC 2021). Pretoria, South Africa. 2021.