

Master Thesis title / MAL izenburua

Towards the Enrichment of Basque WordNet with a Sentiment Lexicon

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

Jon Alkorta eta Itziar Gonzalez-Dios

Contact / Kontaktua: email

jon.alkorta@ehu.eus, itziar.gonzalezd@ehu.eus

Description / Deskribapena

WordNet is a lexical database of semantic relations between words. WordNet links words into semantic relations including synonyms, hyponyms, and meronyms and is used in automatic text analysis and artificial intelligence applications.

The Basque WordNet (Agirre et al., 2002) is also developed. In the latest distribution (2016), BWN had 30 263 synsets, 40 420 variants for nouns, 9 469 for verbs and 148 for adjectives. There were no variants for adverbs. As far as the size of BWN is concerned, its size is limited in comparison with wordnet in other languages, especially in the case of adjectives.

Goals / Helburuak

The objective of this task is to develop and evaluate a methodology to enrich the Basque WordNet using a sentiment lexicon, especially focusing on the adjectives.

Requirements / Betebeharrak

- Advanced skills in Basque language

Framework / Esparrua

- Language Resources

- WordNet

- Sentiment analysis

Tasks and plan / Atazak eta plana

1- Study the latest advances in Wordnet and sentiment analysis.

- <http://nmis.isti.cnr.it/sebastiani/Publications/LREC10.pdf> and <https://github.com/aesuli/SentiWordNet/blob/master/papers/SWNFeedback.pdf>.
- <http://www.ix.eus/sites/default/files/dokumentuak/3417/eacl2014-qwn-ppv.pdf>.

2- Design a methodology to enrich the Wordnet using the sentiment lexicon.

3- Enrich the Wordnet with the developed methodology.

4- Evaluate the designed methodology.

5- Test/evaluate in a NLP application (sentiment analysis, WSD...)

6- Propose improvements for the future.

7- Write up the report

References / Erreferentziak

Agirre, E., Ansa, O., Arregi, X., Arriola, J. M., de Ilarraza, A. D., Pociello, E., & Uria, L. (2002, January). Methodological issues in the building of the Basque WordNet: quantitative and qualitative analysis. In *Proceedings of the first International WordNet Conference in Mysore, India* (pp. 21-25).

Alkorta, J., Gojenola, K., & Iruskietia, M. (2018). SentiTegi: Semi-manually Created Semantic Oriented Basque Lexicon for Sentiment Analysis. *Computación y Sistemas*, 22(4).

Gonzalez-Dios, I., & Alkorta, J. (2020, May). Exploring the Enrichment of Basque WordNet with a Sentiment Lexicon. In *Proceedings of the LREC 2020 Workshop on Multimodal Wordnets (MMW2020)* (pp. 20-24).