Open semantic analysis: The case of word level semantics in Danish

The present research is motivated by the need for accessible and efficient tools for automated semantic analysis in Danish. We are interested in tools that are completely open, so they can be used by a critical public, in public administration, non-governmental organizations and businesses. We describe data-driven models for Danish semantic relatedness, word intrusion and sentiment prediction. Open Danish corpora were assembled and unsupervised learning implemented for explicit semantic analysis and with Gensim’s Word2vec model. We evaluate the performance of the two models on three different annotated word datasets. We test the semantic representations’ alignment with single word sentiment using supervised learning. We find that logistic regression and large random forests perform well with Word2vec features.

General information
Publication status: Published
Organisations: Department of Applied Mathematics and Computer Science, Cognitive Systems, Copenhagen Center for Health Technology
Contributors: Nielsen, F. Å., Hansen, L. K.
Number of pages: 5
Publication date: 2017

Host publication information
Title of host publication: Proceedings of 8th Language and Technology Conference
Source: PublicationPreSubmission
Source ID: 140579168
Research output: Chapter in Book/Report/Conference proceeding > Article in proceedings – Annual report year: 2017 > Research > peer-review