Natural Language Processing

The Natural Language Processing group at the University of Szeged has been involved in human language technology research since 1998, and by now, it has become one of the leading workshops of Hungarian computational linguistics. Both computer scientists and linguists enrich the team with their knowledge, moreover, MSc and PhD students are also involved in research activities. The team has gained expertise in the fields of information extraction, implementing basic language processing toolkits and creating language resources. The Group is primarily engaged in processing Hungarian and English texts and its general objective is to develop language-independent or easily adaptable technologies. With the creation of the manually annotated Szeged Corpus and TreeBank, as well as the Hungarian WordNet, SzegedNE and other corpora it has become possible to apply machine learning based methods for the syntactic and semantic analysis of Hungarian texts, which is one of the strengths of the group. They have also implemented novel solutions for the morphological and syntactic parsing of morphologically rich languages and they have also published seminal papers on computational semantics, i.e. uncertainty detection and multiword expressions. They have developed tools for basic linguistic processing of Hungarian, for named entity recognition and for keyphrase extraction, which can all be easily integrated into large-scale systems and are optimizable for the specific needs of the given application. Currently, the group’s research activities focus on the processing of non-canonical texts (e.g. social media texts) and on the implementation of a syntactic parser for Hungarian, among others.

The group has more than 100 international publications and has participated in major language technology projects funded by the Hungarian Government. The group has active international collaborations with leading European institutes. The team has also delivered several R&D solutions for industrial partners. The group has organized the annual Hungarian Computational Linguistics Conferences since 2003, the Shared Task of the CoNLL-2010 and the Fourth Global WordNet Conference (GWC2008). Members of the team also offer courses on language technology for university students at all levels.

Our website: [http://rgai.inf.u-szeged.hu/nlp](http://rgai.inf.u-szeged.hu/nlp)


Links: