

Semantic frame matching, and the automatic evaluation of an Information Extraction system

Richard Farkas¹, Kinga Konczer², György Szarvas¹

¹ Hungarian Academy of Sciences, University of Szeged;
Research Group on Artificial Intelligence,
6720 Szeged, Aradi vertanuk tere 1., Hungary,
{rfarkas, szarvas}@inf.u-szeged.hu

² University of Szeged, Hungary
kinga.konczer@hungary.org

Abstract: The Frametagger is a semantic pattern matching software designed to identify the actors in short business news, developed by the Human Language Technology Group of the University of Szeged. The module is based on the semantic frames and tables developed in the NKFP 2/017/2001 project by the Research Institute for Linguistics, and extended by us later on, and is the final module of the Information Extraction Toolchain developed in Szeged. In this paper, we introduce a Benchmark algorithm as well, which is made to give a realistic insight how the development, or replacement of each module in the toolchain effects the results/accuracy of the whole Information Extraction System.

Semantic frame matching and evaluation

The goal of Information extraction is collecting and marking relevant information in documents. Systems in practice usually concentrate on the identification of the relevant actors in texts (in spite of a more general semantic role labeling task, where the goal is to identify the syntactic structure for each verb), without doing much detailed syntactic or semantic analysis.

In our system we applied shallow parsing and a semantic frame set for identifying relevant actors. The frames describe events by giving syntactic and semantic constraints of sentence constituents that play relevant roles in that event. In our case a pattern matching is done on parsed texts and each frame's target word and other roles.

The benchmark algorithm we developed compares the results of the system to a some gold standard files containing the hand-made annotation of semantic roles concerning the same frame set the IE toolchain uses. We give not only an accuracy relative to manual annotation, but also try to guess (in case of differences between the gold standard and the toolchain results) which module's errors misled the whole system to find other, or no fit at all. By now the toolchain works with 70.25% accuracy according to the benchmark, with the majority of errors caused by different NP structure annotation, or matching constituent that play indifferent roles but are ambiguous to a role in the frame we use.