UBY – A Large-Scale Unified Lexical-Semantic Resource

Iryna Gurevych
Technical University Darmstadt
Email: gurevych@ukp.informatik.tu-darmstadt.de

Judith Eckle-Kohler
Technical University Darmstadt
Email: eckle-kohler@ukp.informatik.tu-darmstadt.de

Silvana Hartmann ✎
Technical University Darmstadt
Email: hartmann@ukp.informatik.tu-darmstadt.de

Michael Matuschek
Technical University Darmstadt
Email: matuschek@ukp.informatik.tu-darmstadt.de

Christian M. Meyer
Technical University Darmstadt
Email: meyer@ukp.informatik.tu-darmstadt.de

Tri Duc Nghiem
Technical University Darmstadt
Email: nghiem@ukp.informatik.tu-darmstadt.de

We present UBY, a large-scale lexical-semantic resource combining a wide range of information from expert-constructed and collaboratively created resources for English and German. It currently contains nine resources in two languages: English WordNet, Wiktionary, Wikipedia, FrameNet and VerbNet, German Wikipedia, Wiktionary, and GermaNet, and the multilingual OmegaWiki.

The main contributions of our work can be summarised as follows. First, we define a standardised format for modelling the heterogeneous information coming from the various lexical-semantic resources (LSRs) and languages included in UBY. For this purpose, we employ the ISO standard Lexical Markup Framework and Data Categories selected from ISOCat. In this way, all types of information provided by the LSRs in UBY are easily accessible on a fine-grained level. Further, this standardised format facilitates the extension of UBY with new languages and resources. This is different from previous efforts in combining LSRs which usually targeted particular applications and thus focused on aligning specific types of information only.

Second, UBY contains nine pairwise sense alignments between resources. Through these alignments, we provide access to the complementary information for a word sense in different resources. For example, if one looks up a particular verb sense in UBY, one has simultaneous access to the sense in WordNet and to the corresponding sense in FrameNet.

Third, UBY is freely available and we have developed an easy-to-use Java API which provides unified access to all types of information contained in UBY. This facilitates the utilization of UBY for a variety of NLP tasks.