
Pattern-based extraction of taxonomic relations from German forum data

Julia Bettinger
Universität Stuttgart

julia.bettinger@ims.uni-stuttgart.de

Ulrich Heid
Universität Hildesheim

heidul@uni-hildesheim.de

Tag
Datum
Zeit
Raum

Pattern-based extraction of hypernyms and hyponyms has been used since Hearst's (1992) seminal work in applications for relation extraction (cf. e.g. Snow et al. 2005) or ontology construction (cf. Maynard et al. 2009). For German, Granitzer et al. 2009 presented a similar system. More recently, Zouaq et al. (2012) used pos-tagged, lemmatized and dependency-parsed data along with the pattern-based approach.

We analyze the performance of a refined pos-shape-based version of four Hearst-style hyperonymy patterns on German data from online forum texts (27 million words, domain of do-it-yourself instructions). We compare the output with that of patterns based on predicate-argument structures of German verbs of class membership (*gehören zu*, *zählen zu*).

We evaluate the approaches intrinsically (do the patterns find correct hypernym/hyponym structures?) and wrt the relevance of the result data for an ontological modelling of the domain. As we do not have a gold standard for the targeted relation type yet, only precision can be measured: we get 76% good pairs and over 90% of these are domain-relevant.

In a practical application, we combine these data with hyperonymy relations extracted from compounds and multiword patterns.

References: • Granitzer, M. et al. (2009): Taxonomy extraction from german encyclopedic texts. In: *Proceedings of the Malaysian Joint Conference on Artificial Intelligence 2009* • Hearst, M.A. (1992): Automatic acquisition of hyponyms from large text corpora. In: *Proceedings of the 14th Conference on Computational Linguistics - Volume 2*, COLING '92, pages 539–545, Stroudsburg, PA, USA, 1992. Association for Computational Linguistics. Doi: 10.3115/992133.992154. • Maynard D. et al. (2009): Sprat: a tool for automatic semantic pattern-based ontology population. In: *International conference for digital libraries and the semantic web, Trento, Italy* • Snow R. et al. (2005): Learning syntactic patterns for automatic hypernym discovery. In: L. K. Saul, Y. Weiss, and L. Bottou, *Advances in Neural Information Processing Systems 17*, pages 1297–1304. MIT Press • Zouaq A. et al. (2012): Linguistic patterns for information extraction in ontocmaps. In: *Proceedings of the 3rd International Conference on Ontology Patterns-Volume 929*, pages 61–72. CEUR-WS. Org