Statistical Methods for Identifying and Linking Linguistic Terminology

Christian Lang
IDS Mannheim
lang@ids-mannheim.de

Roman Schneider
IDS Mannheim
schneider@ids-mannheim.de

Terminological vocabularies play a central role in the organization and retrieval of scientific texts. But their compilation is often rather cumbersome. This seems especially true for long-established scientific fields with various theoretical and historical branches, where the use of terminology within documents from different origins is sometimes far from being consistent.

A manual compilation and organization of key terms for a certain scientific domain is time consuming and bound to be subjective. That is why recent developments in the context of Natural Language Processing (NLP) and Digital Humanities (DH) automate the finding – and sometimes even the rough classification – through statistical and linguistic. Against this background, we present a novel approach for the computation of a terminological knowledge base. Our data basis is the grammatical online information system GRAMMIS (Schneider/Schwinn 2014): a highly popular specialist hypertext resource that brings together text-oriented, lexicographical, and bibliographic information about German grammar. We combine information from these semantic markups with linguistic and statistical methods to extract grammatical terminology (Suchowolec et al. 2017). We compare the precision and recall performance of an array of well-established statistical methods (i.a. Weirdness, PageRank and C-value) against a human annotated standard.

References: • Schneider, R./Schwinn, H. (2014): Hypertext, Wissensnetz und Datenbank: Die Web-Informationssysteme grammis und ProGr@mm. In: Ansichten und Einsichten. 50 Jahre Institut für Deutsche Sprache. IDS Mannheim, 337–346. • Suchowolec, K./Lang, C./ Schneider, R./ Schwinn, H. (2017): Shifting Complexity from Text to Data Model. Adding Machine-Oriented Features to a Human-Oriented Terminology Resource. In: Language, Data, and Knowledge. Lecture Notes in Artificial Intelligence. Springer International Publishing, 203–212