



Disambiguation of Biomedical Text.

Dr. Mark Stevenson

Lecturer and EPSRC Advanced Research Fellow (2006-2011) at the
Natural Language Processing group of Sheffield University.

Sala de Grados • 17 de mayo de 2010 • 11: 00
entrada libre hasta completar el aforo

resumen:

Like text in other domains, biomedical documents contain a range of terms with more than one possible meaning. These ambiguities form a significant obstacle to the automatic processing of these texts. Previous approaches to resolving this problem have made use of a variety of knowledge sources including the context in which the ambiguous term is used and domain-specific resources (such as UMLS). We compare a range of knowledge sources which have been previously used and introduce a novel one: MeSH terms. The best performance is obtained using linguistic features in combination with MeSH terms. Performance exceeds previously reported results on a standard test set. Our approach is supervised and therefore relies on annotated training examples. A novel approach to automatically acquiring additional training data, based on the relevance feedback technique from Information Retrieval, is presented. Applying this method to generate additional training examples is shown to lead to a further increase in performance.

sobre Mark Stevenson:

Dr. Mark Stevenson is a lecturer and EPSRC Advanced Research Fellow (2006-2011) at the Natural Language Processing group of Sheffield University. His research interests include lexical semantics, word sense disambiguation, semantic similarity, information extraction and text retrieval. His PhD explored the application of a diverse set of knowledge sources to the word sense disambiguation problem and his thesis was published as a monograph by CSLI Publications. Other publications include two edited volumes and over sixty papers in journals, collected volumes and international conferences. He is a member of the EPSRC Peer Review College, the editorial board of the journal Computational Linguistics and regularly reviews for the leading journals and conferences in his field. He has previously worked for Reuters Ltd. in London where he led projects on the application of language technology to a variety of business problems. While at Reuters he was involved in the release of two corpora of newswire articles which have since been widely used by the research community. In 2001-2 he was an inaugural Reuters Foundation Visiting Fellow at the Center for the Study of Language and Information (CSLI), Stanford University where he worked with several Silicon Valley companies. In 1999 he was a Short Term Research Fellow at the Knowledge Management Group of British Telecoms research.