

Generating Topic Chains and Topic Views: Experiments Using GermaNet

Irene Cramer (with Marc Finthammer and Angelika Storrer)
Technical University of Dortmund

Section: Detecting Thematic Structures: Methods, Resources, Applications

Abstract:

In the project HyTex we investigate strategies, based on text-grammatical features, for the conversion of linear text documents into documents publishable in a hypertext environment. One subtask of this research is the analysis of topic-based linking strategies based on lexical chains and on topic chains.

In our presentation we shortly describe the lexical chainer GLexi implemented for German corpora, which uses GermaNet as well as Google co-occurrence counts as a resource for the calculation of semantic relatedness/similarity. The GLexi chainer also includes an extensive pre-processing component in order to deal with German morphology (noun inflection and noun compounding). We will illustrate how we use GLexi for the construction of topic chains and topic views: topic chains are partial text representations based on a selection of so called topic anchors. Topic views visualize the topic structure of documents with the goal to support the user's orientation and navigation in a hypertext system. Finally, we will report on the results of evaluation studies that compared the performance of the topic chains and views with manually annotated data.