

The Wiki Way to Corpus Analysis: Aspects of Automatic Text Classification

Alexander Mehler

Fakultät für Linguistik und Literaturwissenschaft,
Universität Bielefeld, 33615 Bielefeld, Germany

Abstract. We analyze four different types of document networks with respect to their small world characteristics. These characteristics allow distinguishing wiki-based systems from citation and more traditional text networks augmented by hyperlinks. The study provides evidence that a more appropriate network model is needed which better reflects the specifics of wiki systems. It puts emphasize on their topological differences as a result of wiki-related linking compared to other networks. It also emphasizes the consequences of this analysis for text classification and develops an integrative model of content *and* structure-based classification.

Key words: text networks, graph model, structure-based classification, content-based classification