Journal of Statistical Software, Vol 25 (2008)

Text Mining Infrastructure in R

*Ingo Feinerer, Kurt Hornik, David Meyer*

Abstract

During the last decade text mining has become a widely used discipline utilizing statistical and machine learning methods. We present the **tm** package which provides a framework for text mining applications within R. We give a survey on text mining facilities in R and explain how typical application tasks can be carried out using our framework. We present techniques for count-based analysis methods, text clustering, text classification and string kernels.

https://www.jstatsoft.org/article/view/v025i05