

# HPI Kolloquium

## 04.12.2014, 16:00 Uhr

Hasso-Plattner-Institut, Vorlesungsgebäude, Auditorium 1  
Campus Griebnitzsee, 14482 Potsdam

## “Fun With Words – Text Mining on the Web”

Dr. Ralf Krestel

University of Potsdam

### Abstract

The World Wide Web has grown from being a tool for nerds to an important factor in the social, cultural, political, and economic world. In this Web, unstructured, textual data plays a predominant role. Besides user-generated content, online news articles, social media, professional Web pages or Internet encyclopedias, you will also find specialized text collections, such as parliament speeches, legal documents, patent applications, or scientific papers freely available in the Web. Many application areas can benefit from analyzing this data, e.g. journalism, health care, or scientometrics. For such application areas the analysis of single, isolated corpora can only be a starting point. The major challenge in this context is to process and analyze big textual data, including data streams from heterogeneous sources across genre boundaries in near real-time. This will enable the discovery of relationships between corpora and documents, as well as between entities extracted from these different corpora and allow insights otherwise impossible to gain. By developing algorithms, combining, and adapting methods from the area of big data, text mining, and Web science, we aim to address this challenging problem step by step.

### Short Biography

Dr. Ralf Krestel is a research fellow at the HPI graduate school. His research interests are text mining, information retrieval, recommender systems, and natural language processing. He studied at the University of Karlsruhe and Concordia University in Montreal before he obtained his Ph.D. from the University of Hannover while working at the L3S research center. Afterwards he was a postdoctoral scholar for two years at the University of California, Irvin prior of joining HPI. He published more than 30 peer-reviewed articles and is reviewer for different journals and conferences.

Host: Prof. Dr. Felix Naumann