

SIGIR Workshop on Semantic Matching in Information Retrieval

Title: Six Tweets per Second

Speaker: Maarten de Rijke, University of Amsterdam

Abstract:

People increasingly live their lives in an online setting. This observation is fundamental to a growing body of work that aims at describing, understanding and exploiting the abundance of online personal information. The strong potential of social media platforms for understanding people's behavior has already played an important role in scientific literature and as we increasingly integrate social media in our lives, it becomes possible to explore more personal aspects of our lives. For example, Twitter can be used to analyze public health and Facebook has been used to predict personality.

In the talk I will report on ongoing work on real-time analysis of self-reported music listening behavior on Twitter. Roughly six tweets per second report about the music that someone is listening to. That's about 500K tweets per day, 180M tweets per year: an interesting signal about music listening behavior. Our analysis of this signal starts by recognizing artists and songs, a basic entity linking task that is made challenging by the highly dynamic nature of the domain. We then proceed to analyze music listening behavior along four dimensions: who is listening, when are we listening, what are we listening to, and how are we listening to music. Along the way we attempt to understand our Twitter-based music listening signal by matching it to other signals, for instance based on sales or on real-world events or on activities of listeners.

Based on joint work with Manos Tsagkias (University of Amsterdam), Wouter Weerkamp (904Labs), Edgar Meij (Yahoo! Labs), Guido van Bruggen (University of Amsterdam).