

Social media has become an integral part of society. Omnipresent mobile devices allow for immediate sharing of experiences. Experiences can be about brands and other entities. For social media analysts a collection of posts mentioning a brand can serve as a magnifying glass on the prevalent opinion towards a brand: The overall estimation of a brand's reputation is increasingly based on the aggregation of a brand's reputation polarity in social media posts. This polarity of reputation is currently annotated manually. However, with the dramatic increase of social media, this is no longer feasible.

This thesis aims to facilitate and automate parts of the process to estimate the reputation of a brand. We motivate this by performing user studies with expert social media analysts. We analyse three resulting datasets: a questionnaire, log data of a manual annotation interface, and videos of annotating experts following the think-aloud protocol. We find the online and offline authority of the user posting online influences the annotation decision most. This online and offline authority is therefore a strong indicator for reputation polarity of this posting. Additionally, experts welcome automation of information retrieval and filtering tasks. For both, information retrieval and filtering, as well as for several indicators, the reputation analysts' background information proves vital.

Based on the indicators used for manual annotation, we proceed with the development of algorithms for the automatic estimation of reputation polarity. Unlike earlier, static evaluation scenarios, we follow a dynamic scenario, which mimics the daily workflow of social media analysts. Our algorithms are successful because we distinguish between reputation and sentiment.

The second part of this thesis is motivated by the analysts' desire for automation of retrieval and filtering of new media. For information retrieval, we present two improvements to existing algorithms. The first improvement is based on burst identification in time series of pseudo-relevant documents. We sample terms in those bursts for query modeling, thus improving effectiveness on news and blog corpora. Secondly, recency is an important aspect of relevance in social media. Inspired by memory models from cognitive science, we point out new (cognitive) document priors based on those models. We show that those priors are more effective, efficient, and plausible than commonly used temporal priors.

Background knowledge is essential for information filtering. Additionally, topics around an entity are dynamic. We find that for a consistently strong performance of filtering algorithms, the expertise of social media analysts is needed. The algorithms for filtering presented in this thesis are based on active learning: For documents where the algorithm cannot classify with high certainty, manual labels are requested from the analyst. Using intuitions about bursts and cognitive priors for sampling difficult to classify documents, we need very little help from annotators to reach high effectiveness.

We conclude that many aspects of the annotation of reputation can be automated – using in particular time series analysis, memory models, and low-impact help from expert social media analysts.

ISBN 978-94-6182-540-7



MARIA-HENDRIKE PEETZ TIME-AWARE ONLINE REPUTATION ANALYSIS

# TIME-AWARE ONLINE REPUTATION ANALYSIS

## MARIA-HENDRIKE PEETZ