

## AIM - A Social Media Monitoring System for Quality Engineering

In the last few years the World Wide Web has dramatically changed the way people are communicating with each other. The growing availability of Social Media Systems like Internet fora, weblogs and social networks ensure that the Internet is today, what it was originally designed for: A technical platform in which all users are able to interact with each other. Nowadays, there are billions of user comments available discussing all aspects of life and the data source is still growing.

This thesis investigates, whether it is possible to use this growing amount of freely provided user comments to extract quality related information. The concept is based on the observation that customers are not only posting marketing relevant information. They also publish product oriented content including positive and negative experiences. It is assumed that this information represents a valuable data source for quality analyses: The original voices of the customers promise to specify a more exact and more concrete definition of "quality" than the one that is available to manufacturers or market researchers today.

However, the huge amount of unstructured user comments makes their evaluation very complex. It is impossible for an analysis protagonist to manually investigate the provided customer feedback. Therefore, Social Media specific algorithms have to be developed to collect, pre-process and finally analyze the data. This has been done by the Social Media monitoring system AIM (Automotive Internet Mining) that is the subject of this thesis. It investigates how manufacturers, products, product features and related opinions are discussed in order to estimate the overall product quality from the customers' point of view. AIM is able to track different types of data sources using a flexible multi-agent based crawler architecture. In contrast to classical web crawlers, the multi-agent based crawler supports individual crawling policies to minimize the download of irrelevant web pages. In addition, an unsupervised wrapper induction algorithm is introduced to automatically generate content extraction parameters which are specific for the crawled Social Media systems. The extracted user comments are analyzed by different content analysis algorithms to gain a deeper insight into the discussed topics and opinions. Hereby, three different topic types are supported depending on the analysis needs.

- The creation of highly reliable analysis results is realized by using a special context-aware taxonomy-based classification system.
- Fast ad-hoc analyses are applied on top of classical fulltext search capabilities.
- Finally, AIM supports the detection of blind-spots by using a new fuzzified hierarchical clustering algorithm. It generates topical clusters while supporting multiple topics within each user comment.

All three topic types are treated in a unified way to enable an analysis protagonist to apply all methods simultaneously and in exchange. The systematically processed user comments are visualized within an easy and flexible interactive analysis frontend. Special abstraction techniques support the investigation of thousands of user comments with minimal time efforts. Hereby, specifically created indices show the relevancy and customer satisfaction of a given topic.