

UCSMA - Understanding the Customer with Social Media Analyses

Background

The massive adoption of social media provides new ways for individuals to express their opinions online. This creates an enormous and ever-changing archive of open source intelligence.



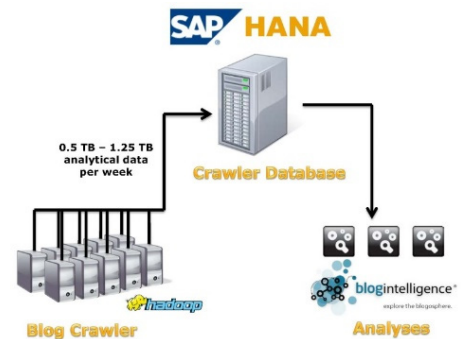
However, it is increasingly difficult - if not impossible - for the average internet user and social media enthusiast to make use of this intelligence, due to thousands of new users and an almost uncountable number of new posts adding up to the world-wide collective on a daily basis.

Mining, analyzing, modeling and presenting this immense data collection is of central interest. This enables the user to detect technical trends, political atmospheric pictures or news articles about a specific topic.

Description

Within this bachelor project we focus on real world use cases from different customers like Standard Bank South Africa, AMG and others for the development of a new social media analyses tool.

The early detection of technical problems of car manufactures to avoid product recalls is one use case. We like to come over this issue by combining social media analyses with analyses of internal service reports from their retailers.



The base is the huge amount of data provided by social media all over the world. The following topics are part of this project:

- Massive parallel crawling of social media sources on multi-core machines
- Continuously identifying, downloading and updating web sources
- Detailed knowledge about the structure and dynamics of social media channels
- Algorithms for parsing and extracting semantic information
- Analyses like Ranking, Clustering, Dimensional Reduction or Importance Indices
- Visualization of interconnections and exploration of big unstructured data

You will be able to work with a 150 core Hadoop Cluster and a 6TB SAP HANA database. The Jobs for this cluster will be written with Java and MapReduce. Further, the analyses is based on SQL, SQL Script and C++. The results of these analyses are retrieved via the XSEngine by using a RESTful API with JavaScript.

We will especially focus on nice and easy-to-use visualizations with advanced web technologies like D3 (d3js.org) for graphs and Harp (harpjs.com) for CoffeScript, Stylus, etc.

Contact

Fachgebiet: Internet-Technologien und -Systeme

- Fachgebietsleiter: Prof. Dr. Christoph Meinel
- Ansprechpartner: Patrick Hennig, Philipp Berger
- Projektpartner: SAP, Standard Bank, AMG

References

- [1] BlogIntelligence: blog-intelligence.com