

Semantic-based Learning for Trend Mining in Text Collections

Olga Streibel

Networked Information Systems, Free University Berlin,
Königin-Luise-Str.24-26 , 14195 Berlin, Germany
streibel@inf.fu-berlin.de
<http://www.ag-nbi.de>

Determination and early detection of emerging trends can be retrieved from numeric data as well as from texts [2]. However, the use of text collections in the process of trend detection requires new analysis techniques. Although many interesting approaches have been developed in the field of Trend Mining on texts, defined as Emerging Trend Detection in Text Mining[1], they are still lacking the integration of expert knowledge in the process of trend recognition. Such knowledge is crucial for the proper trend detection in various domains, e.g. in medical diagnosis, opinion mining, financial markets.

Our research addresses trend detection in text collections and we are concentrating on the development of a novel, knowledge-based learning approach for the automatic detection of trends in text streams. At this stage of our work, we are concerning ourselves with two main research issues: representation of trend knowledge and knowledge-integrating learning approach for trend detection.

Knowledge about emerging trends is hard to define and relies on the knowledge of given domain including experts' experience in detecting trends for this domain. Knowledge of given domain can be expressed in language, thus in formal definition of domain concepts and relations between these concepts. Semantic Web offers ontology as a technology for knowledge formalization. However, the classic Ontology approach has been created under assumption of hierarchical, static relations describing knowledge and therefore can be successfully applied in domains of taxonomic characteristic (i.e. life sciences). Knowledge about trends is more intuitive, dynamic, context- and time-dependent, subjective. We assume that formalization of trend knowledge requires for novel lightweight formalization techniques, like *extreme tagging*[4]¹, that allow for collective capturing non-hierarchical relations and enhanced semantics. For this reason, we are concentrating on the utilization of *Extreme Tagging Systems* approach in order to gather and formalize knowledge needed for trend detection.

Regarding chosen knowledge representation paradigm and different learning approaches from Artificial Intelligence[3], we are elaborating on the proper definition of a knowledge-integrating learning approach for trend detection in texts. Considering several statistical learning methods, we aim at applying the appropriate method to the financial market texts in order to enable trend detection in this domain.

¹ <http://www.corporate-semantic-web.de/extreme-tagging.html>

References

1. April Kontostathis, Leon Galitsky, William M. Pottenger, Soma Roy, and Daniel J. Phelps. *A Survey of Emerging Trend Detection in Textual Data Mining*. Springer-Verlag, 2003.
2. Victor Lavrenko, Matt Schmill, Dawn Lawrie, Paul Ogilvie, David Jensen, and James Allan. Mining of concurrent text and time series. In *In proceedings of the 6th ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining Workshop on Text Mining*, pages 37–44, 2000.
3. Stuart Russel Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall International, 2003.
4. Vlad Tanasescu and Olga Streibel. Extreme tagging: Emergent semantics through the tagging of tags. In *ESOE*, pages 84–94, 2007.