

Determination and early detection of emerging trends can be retrieved from numeric data as well as from texts [1] Since mining trends by analyzing text streams can enhance the trend analysis based only on numeric data, the use of qualitative information in the process of trend recognition 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 [2], they are still lacking the integration of expert knowledge in the process of trend recognition and such knowledge is crucial for the proper trend detection in various domains, i.e. in financial markets. For this reason we have concentrated our research on the trend recognition and aim to develop a novel approach for recognition of emerging trends from texts. Thus, our work contributes to the integration of formalized expert knowledge to the process of trend recognition. Furthermore, since Semantic Web technologies enable the appropriate and advantageous formalization of knowledge we strive to apply them to enhance the quality of machine learning methods utilized for trend recognition.

The key objective of our research is to develop a semantic based learning method for trend recognition in hybrid information systems; information systems consisting of both qualitative and quantitative data. Using the financial market as an example of a hybrid information system an adequate trend recognition method has to be developed for the recognition and analysis of temporal changing patterns in texts that can effect a trend in the system. Numeric data, i.e. stock values, can then be used in order to evaluate recognized trends.

We aim to apply the novel knowledge acquisition techniques based on lightweight annotations in social environments such as *extreme tagging*, i.e. the tagging of tags controlled by social processes, in order to generate a formalized description of expert knowledge regarding trends. An extreme tagging principle as described in [3] allows easy retrieval of the expert knowledge. Relying on the weighted experts' associations that are created through tagging the tags, we aim to use two powerful aspects of Extreme Tagging System (ETS) approach: collective knowledge and serendipity effect through *knowledge paths* [3]. As for analyzing a text stream there is a need for an appropriate knowledge base, called a *TrendWordNet*, which should be extracted from ETS tool. Currently, our main research questions are how to structure the output of the ETS tool in order to create an appropriate set of trend vocabulary for a given domain? How to structure this vocabulary in TrendWordNet and combine TrendWordNet with a clustering method in order to keep its structure "up-to-date"? How to deal with the time-component in this vocabulary?

Considering these questions, the primary aim of this research considers a combination of the classic learning method and the knowledge base created using Semantic Web technologies. Utilizing the ETS-tool to an expert group of a financial market analysis, we aim to gather expert knowledge that can be then formalized with our methods. Furthermore, applying this knowledge to the method for text stream analysis, we expect to discover trends that are based on collective knowledge, not only on statistics.

---

[1] Victor Lavrenko and Matt Schmill and Dawn Lawrie and Paul Ogilvie and David Jensen and James Allan, *Mining of Concurrent Text and Time Series* In proceedings of the 6 th ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining Workshop on Text Mining, pages 37-44, 2000

[2] April Kontostathis and Leon Galitsky and William M. Pottenger and Soma

Roy and Daniel J. Phelps *A Survey of Emerging Trend Detection in Textual Data Mining* in A Comprehensive Survey of Text Mining, Springer Verlag, 2003

[3] Vlad Tanasescu, Olga Streibel, *Extreme Tagging: Emergent Semantics through the Tagging of Tags* in International Workshop on Emergent Semantics and Ontology Evolution, November 2007, Busan Korea

---