

Graph Transformation-based Opinion Mining in Web Queries

Gábor Imre and Gergely Mezei

In our research we evaluate the possible uses of opinion mining algorithms in a web-based graph transformation system. Text analysis algorithms, such as opinion mining [1], can be applied within graph transformations to provide effective analyzing and processing capabilities for interpreting public web data.

The web provides tremendous amounts of data available in various formats. Some data providers publish their data in a semi-structured format, where plain text and formally defined data is mixed. A typical example is the Stack Exchange API¹, where huge amounts of Q&A forum data is available. The data is semi-structured: it has a formal structure for tags, scores, connections between questions and answers, etc.; yet the text of the questions and answers is available as plain text. Text mining algorithms can be used to process the plain text, while graph transformations are strong in processing structured data. We examine the possibilities and requirements of applying opinion mining algorithms as parts of graph transformations.

Graph transformations are widely used in various industrial and scientific tasks in Model-Driven Architecture [2]. Graph transformations over typed attributed graphs [3] are especially strong in practical representation of object oriented structures, in addition they have strong mathematical background. Combining graph transformations with text analysis algorithms can lead to powerful analysis applications.

We have constructed a case study that demonstrates the usage of opinion mining through graph transformations over semi-structured data. We analyze the questions and answers available at Stack Overflow², the world's greatest programming Q&A site. Some of the most important Linux distributions are compared based on the question and answer plain texts. The sentiment of a question or answer is extracted by a simple, dictionary-based opinion mining algorithm. If the user opinions are generally more positive, then the regarding distribution is considered more comfortable. This simple heuristics can be used to compare similar technologies based on the general user opinions.

Acknowledgements

This work was partially supported by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013) organized by VIKING Zrt. Balatonfüred; and by the Hungarian Government, managed by the National Development Agency, and financed by the Research and Technology Innovation Fund (grant no.: KMR 12-1-2012-0441).

References

- [1] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and trends in information retrieval*, vol. 2, pp. 1-135, 2008.
- [2] "The Model-Driven Architecture", <http://www.omg.org/mda/>, OMG, Needham, MA, 2002
- [3] H. Ehrig, U. Prange and G. Taentzer, *Fundamental theory for typed attributed graph transformation*, Springer Berlin Heidelberg, 2004.

¹<https://api.stackexchange.com/docs/>

²<https://stackoverflow.com>