



Opinion Mining in the Semantic Web using Search Engines

Ioannis Torakis, Vasilis Stavrou

{john.torakis@gmail.com, stavrouv}@aueb.gr

Information Security and Critical Infrastructure Protection (INFOSEC) Laboratory
Dept. of Informatics, Athens University of Economics & Business (AUEB), Greece



Introduction

- Search Engine Results about political and generally opinion-based issues tend to generate and feedback people opinions.
- Visualize and measure the opinions that a Search Engine possesses about different subjects.

Methodology

- Semantic Document Clustering of Search Engine Results.
- Sentiment Analysis of the Search Result Topics.
- Visualization of the expressed sentiment.
- Text Visualization Methods for every result (Word Clouds, etc.).

Case study

- Opinion Mining about "Grexit" and "European Union".
- Analysis of Search Results from related Search Terms.
- Analysis of similar Search Terms from different Countries (Greece, Germany, USA, etc.).

Results of a "Grexit" cluster

- Formation of 3 results included in the cluster.
- Cluster contains information about "Grexit".
- Sentiment analysis of the processed content.

Grexit: Why even 'yes' could see Greece leave - CNBC
The path for Greece to leave the euro zone (if that is what eventually happens) is unlikely to be smooth. In the chart below, we outline some of the ...
<http://www.cnbc.com/2015/07/03/grexit-why-even-yes-could-see-greece-leave.html>

Grexit™ could happen by accident - WSJ
"Grexit" Could Happen by Accident Greek Crisis Testing Eurozone to the Limit
<http://www.wsj.com/articles/grexit-could-happen-by-accident-1421012401>

Will a Grexit Be the Lehman-Like Trigger of the Next ...
Around the Web. US economy contracts thanks to strong dollar and poor weather -as it happened. Spectre of 'Grexit' stalks G7 finance meeting. Putin far ...
http://www.huffingtonpost.com/joseph-e-stiglitz/grexit-global-financial-crisis_b_7531366.html

Figure 1: Clustered links

Tag Cloud



Figure 3: Tag Cloud

Sentiment pie

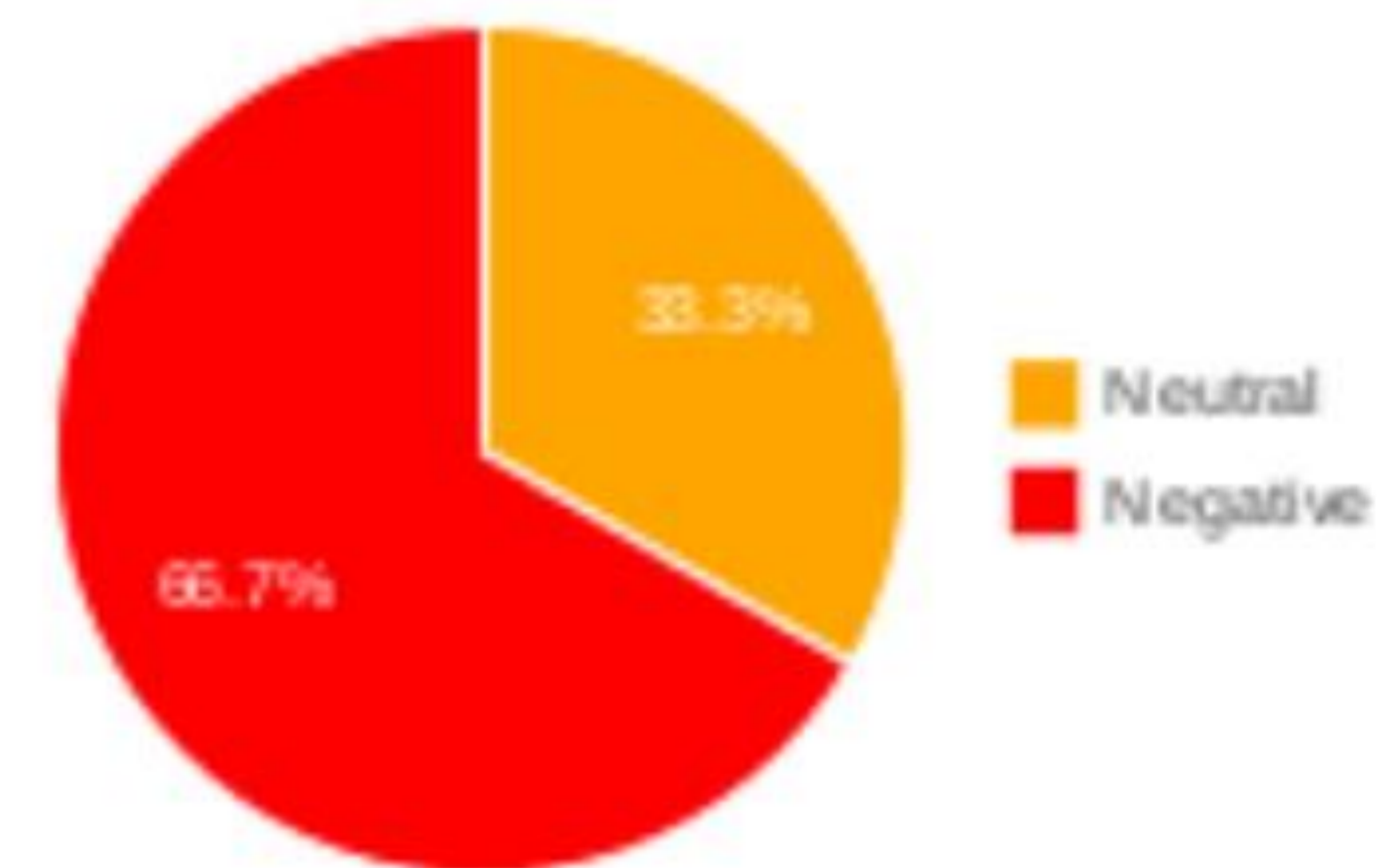


Figure 4: Sentiment expressed in the examined cluster

Bipolar sentiment analysis for each cluster result

- Bipolar sentiment analysis for each result of the aforementioned example.
- Results 1 and 2 tend to contain more negative content than result 0.
- Result 0 refers to the path of Greece to leave the euro zone.
- Result 1 refers to the case that Grexit would happen by accident.
- Result 2 refers to potential dangers caused by a Grexit.

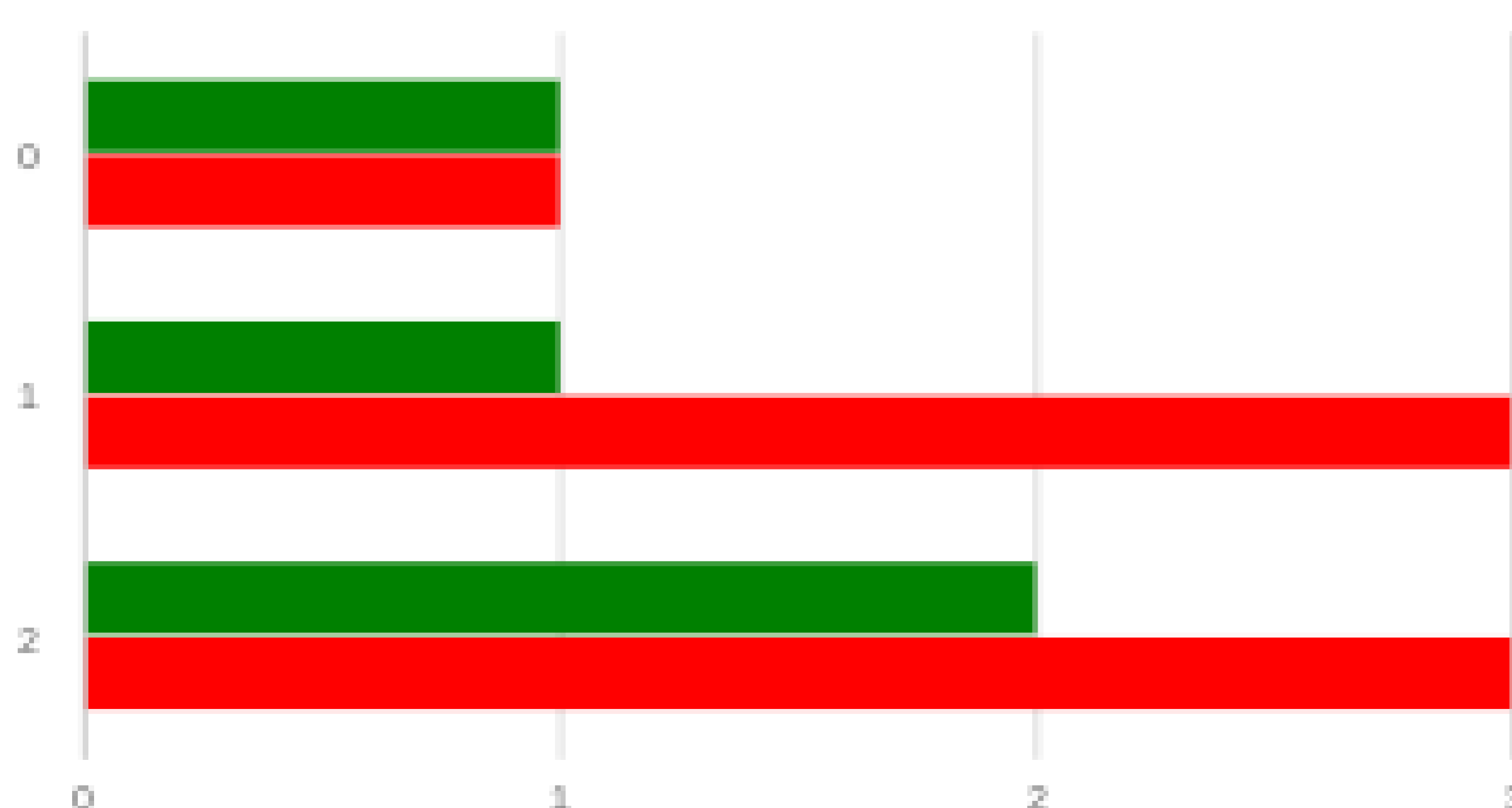


Figure 2: Sentiment ratings

Conclusions

- Result Topics and Sentiment Analysis scores differ vastly by country.
- Greece has the least negative "opinion" about "Grexit". Germany is the most negative.
- "European Union" tends to get negative sentiment scores in all regions checked.
- This topic combines well with similar experience and publications of our Lab.

References

- Page L., Brin S., Motwani R., Winograd T., "The PageRank citation ranking: bringing order to the Web", 1999.
- Ma W. Y., Xiao X., Xie X., "Classifying functions of web blocks based on linguistic features." U.S. Patent No. 7,895,148, 22 Feb. 2011.
- Osiński S., "An algorithm for clustering of web search results", Doctoral dissertation, Poznań University of Technology, Poland, 2003.
- Ilic M., Spalevic P., Veinovic M., "Suffix Tree Clustering-Data mining algorithm".
- Berners-Lee T., & Cailliau R., "WorldWideWeb: Proposal for a HyperText project", Retrieved on February, 26, 2008.
- Amichai-Hamburger, Y., Vinitzky, G., "Social network use and personality", 2010.
- Shaw, E., Ruby, K., Post, J., "The insider threat to information systems: The psychology of the dangerous insider", Security Awareness Bulletin, pp. 1-10, 1998.
- Gritzalis D., Stavrou V., Kandias M., Stergiopoulos G., "Insider Threat: Enhancing BPM through Social Media", in Proc. of the 6th IFIP International Conference on New Technologies, Mobility and Security (NMITS-2014), Springer, UAE, 2014.
- Gritzalis D., "Insider threat prevention through Open Source Intelligence based on Online Social Networks", Keynote address, 13th European Conference on Cyber Warfare and Security (ECCWS-2014), Greece, 2014.
- Gritzalis D., Kandias M., Stavrou V., Mitrou L., "History of Information: The case of Privacy and Security in Social Media", in Proc. of the History of Information Conference, Law Library Publications, Athens, 2014.
- Kandias M., Mitrou L., Stavrou V., Gritzalis D., "Which side are you on? A new Panopticon vs. privacy", in Proc. of the 10th International Conference on Security and Cryptography (SECRYPT-2013), pp. 98-110, Iceland, 2013.
- Kandias M., Galbogini K., Mitrou L., Gritzalis D., "Insiders trapped in the mirror reveal themselves in social media", in Proc. of the 7th International Conference on Network and System Security (NSS-2013), pp. 220-235, Springer (LNCS 7873), Spain, June 2013.
- Kandias M., Virvilis N., Gritzalis D., "The Insider Threat in Cloud Computing", in Proc. of the 6th International Conference on Critical Infrastructure Security (CRITIS-2011), pp. 93-103, Springer (LNCS 6983), United Kingdom, 2013.
- Kandias M., Stavrou V., Bozovic N., Mitrou L., Gritzalis D., "Can we trust this user? Predicting insider's attitude via YouTube usage profiling", in Proc. of 10th IEEE International Conference on Autonomic and Trusted Computing (ATC-2013), pp. 347-354, IEEE Press, Italy, 2013.
- Kandias M., Stavrou V., Bosovic N., Mitrou L., Gritzalis D., "Proactive insider threat detection through social media: The YouTube case", in Proc. of the 12th Workshop on Privacy in the Electronic Society (WPES-2013), pp. 261-266, ACM Press, Germany, 2013.
- Kandias M., Virvilis N., Gritzalis D., "The Insider Threat in Cloud Computing", in Proc. of the 6th International Workshop on Critical Infrastructure Security (CRITIS-2011), Bologna S., et al (Eds.), pp. 93-103, Springer (LNCS 6983), Switzerland, 2011.
- Kandias M., Mylonas A., Virvilis N., Theoharidou M., Gritzalis D., "An Insider Threat Prediction Model", in Proc. of the 7th International Conference on Trust, Privacy, and Security in Digital Business (TrustBus-2010), pp. 26-37, Springer (LNCS-6264), Spain, 2010.
- Mitrou L., Kandias M., Stavrou V., Gritzalis D., "Social media profiling: A Panopticon or Omnipoticon tool?", in Proc. of the 6th Conference of the Surveillance Studies Network, Spain, 2014.
- Stavrou V., Kandias M., Karoulas G., Gritzalis D., "Business Process Modeling for Insider threat monitoring and handling", in Proc. of the 11th International Conference on Trust, Privacy & Security in Digital Business (TRUSTBUS-2014), pp. 119-131, Springer (LNCS 8647), Germany, September 2014.