More than a Feeling: Insights and Information from a Sentiment Analysis Study

Leonidas Papachristopoulos

Hellenic Open University Distance Library and Information Center, Patras, Greece
lpapachristopoulos@eap.gr, orcid.org/0000-0002-4148-2689

Giannis Tsakonas

Library and Information Center, University of Patras, Patras, Greece
gtsak@upatras.gr, orcid.org/0000-0002-8786-9440

Abstract

Conferences have gained a significant role within the scholarly communication ecosystem as they contribute to the immediate dissemination of scientific results and, also, to the professional socialization. LIBER's Annual Conference constitutes one of the main channels of communication for the information professionals of the European research libraries. Each year, a parallel universe arises during the conference's activities in the social media, especially in the micro-blogging sphere of Twitter, where conference attendees express their opinions about trending topics, organizational aspects and conference talks. The current study proposes that the implementation of sentiment analysis to data generated on that sphere could be a useful decision-making tool for both the organizational committee and the host organization. By applying a sentiment analysis on 5,500 tweets regarding the LIBER 47th Annual Conference in Lille, France, we highlight the positive and negative aspects of the specific event as they were mentioned by Twitter users.

Keywords: Sentiment analysis; machine learning; Twitter; social web; conferences

1. Introduction

Academic conferences constitute venues where intellectual discussions emerge and frame the future of a scientific domain. This fermentation is made formally through the scheduled talks, presentations, workshops and other activities, as well as informally via the dynamic backchannel of communicating on the social media sphere, such as micro-blogging. It has been found that the Twitter activity for scholarly purposes is more intense during a conference than every other day (Ross, Terras, Warwick, & Welsh, 2011). The LIBER Annual Conference attendees seem to be 'hyper-active' in the platform of Twitter, as previous conferences have proved. Recent observations show that the Annual Conference hashtags, such as #Liber2017 for Patras, #liber2018 for Lille and #liber2019 for Dublin, have topped the national Twitter trends. Such an amount of data cannot be manually analyzed in an effective way, so every researcher who would be interested in detecting, classifying and correlating entities and statements would have to apply machine learning techniques. Sentiment analysis constitutes an emerging Natural Language Processing (NLP) method for detecting people's emotion in texts (Cambria, Schuller, Xia, & Havasi, 2013). In this way, any implementation can capture opinions, preferences and interest about products, events, marketing campaigns, etc.

The current study attempts, not to analyze Twitter's users behaviour for scholarly purposes, but to explore their use for "organizational enhancements" (Reinhardt, Ebner, Beham, & Costa, 2009). An effective exploitation of these spontaneous textual data would ameliorate decision making from two viewpoints. From the macro viewpoint, LIBER governance would consider the opinions posted on Twitter and redirect the spotlight of its activities, democratizing in this way the process of decision making and materializing one of its core values, inclusivity. Second, from the micro viewpoint, it would give each local conference organiser the opportunity to evaluate the opinions expressed, especially the ones regarding organizational issues of the events, allowing them to foresee areas of interest and increase the ability to better organise their event.

2. Background

Conferences serve as nodal points of information dissemination, professional advancement – scholars may use them as a starting point for their academic

career and others as an opportunity for skill development—and networking (Seidel, 2018). As Mahrt, Weller and Peters (2014) mention, in parallel with these formal communication channels, informal communication activities, like tweeting, are taking place. Tweets have been analysed from different perspectives, both quantitatively, such as the level of activity, participation and message distribution using qualitative (Borgmann et al., 2016) and quantitative methods (Mishori, Levy, & Donvan, 2014).

It is profound that these datasets need a different methodological approach in order to overcome the size and complexity issues; machine learning methods, such as sentiment analysis, can be an answer. Kimmons and Veletsianos (2016) focus on the tweeting activity of the American Educational Research Association annual conferences between 2014 and 2015, inferring that participants with different roles have different behaviour. The same research applied sentiment analysis to the conference's hashtags, concluding that academics generally exhibit positive sentiments. Desai et al. (2012) also tried to detect sentimental polarity at a medical conference's tweets and found that, in general, informative tweets had more negative sentiment scores than uninformative tweets. They also found that these were more negative during conferences than in the leading up period.

The use of sentiment analysis as a decision-making tool has been recognised lately from researchers coming from diverse fields. The detection of investors' sentiments would be a safe indicator about stock market performance (Wu, Zheng, & Olson, 2014), while other researchers imply that the implementation of such methods would be beneficial for the amelioration of supply management decisions (Wood, Reiners, & Srivastava, 2015). Additionally, big data analytics with sentiment analysis can improve brand managers' and consumers' decisions (Çalı & Balaman, 2019; Kauffmann et al., 2019).

In the library and information science field, sentiment analysis has been used lately for managerial purposes, such as performance evaluation. Papachristopoulos, Ampatzoglou, Seferli, Zafeiropoulou, and Petasis (2019) tried to overcome the handicap of the absence of a dedicated evaluation questionnaire of the Hellenic Open University Distance Library and Information Center by implementing sentiment analysis to user comments on the account of the library in social media platforms, while Moore (2017) applied sentiment analysis to comments that originated from the widely accepted library evaluation tool LibQUAL+.

To our knowledge, this is the first application of sentiment analysis to a library and information science conference's tweets that examines tweeting activity from a managerial perspective. Twitter comments could be used creatively, under the lens of sentiment analysis, as a tool for updating the topics agenda of the LIBER Conference Program Committee or as an organizational compass for the hosts of the conference.

3. Methodology

The proposed work implements sentiment analysis on 5,530 tweets regarding the LIBER 47th Annual Conference in Lille, France, which was held between 4–6 of July 2018 and was entitled "Research libraries as an Open Science hub: from strategy to action". Tweets were identified by the hashtag #liber2018 and were mined via the TAGS tool (v6.1.9),[1] published in the period 18/6/2018 to 9/7/2018. Before analysing, we detected and removed many intruding hashtags that were irrelevant to the scope of the conference, but they were taking advantage of the attention of the #Liber2018 hashtag.

For the purposes of our study, we used Semantria[2] a platform which enables researchers to handle data from diverse channels, including Twitter, Facebook and WordPress, in multiple languages. Our decision to choose this specific tool is justified by other studies that compare the performance of similar software (Serrano-Guerrero, Olivas, Romero, & Herrera-Viedma, 2015). According to these tests, Semantria was found having stable behaviour with low errors. The software detects sentiment in various levels, such as in paragraphs, sentences or proper nouns that represent an entity, implementing the necessary Part Of Speech (POS) tagging. Semantria is available both as cloud API and Excel plug-in. For the needs of the current study, we used the latter version, modifying the sentiment polarity scale. The scale for sentiment ranking ranged from "−2: completely negative" to "2: completely positive". Intervals at "−1.2: mostly negative", "−0.6: slightly negative", "0.6: slightly positive" and "1.2: mostly positive", covered the interim levels of sentiment intensity, while "0" responded to entirely neutral statements. Score assignments close to 0 are characterised as neutral. It has to be mentioned that negative and positive polarities have various nuances and they should not be considered as being one-dimensional, e.g. bad or good. A positive comment may reflect happiness, calmness, interest for something, etc., and,

accordingly, a negative may reflect sadness, anger, boredom and so on (Poria et al., 2012).

We configured the system to detect queries' sentiments. Queries were specific, preferred terms that were imported in order to be matched together (query co-occurrence) and have their sentimental polarity compared. The queries that were imported to the system were a set of broad relevant terms to the themes of the conference, such as 'Citizen Science', 'Copyright', 'FAIR', 'Infrastructures', 'Metrics', 'Open Access', 'Open Science', 'Research Data', 'Science' and 'Skills'.

4. Findings

4.1. Queries

The queries were detected in 1,876 tweets, which constitute 33.92% of the total number of tweets. From these 1,876 tweets, 64 (3.4%) were detected and defined as negative, 1,542 (82.19%) as neutral and 270 (14.39%) as positive. The sentiment analysis of our queries (see Table 1) indicated that the most negative comments were found on copyright-related commentary, while, on the other hand, the most positive ones commented on the FAIR principles. It was evident that the majority of the tweets were categorised as neutral, but also interesting was the fact that, although the 'Open Science' query gathered

Table 1. Queries polarities.

Query	Negative	Neutral	Positive	Total
Citizen Science	5	80	34	119
Copyright	40	118	12	170
FAIR	8	369	66	443
Infrastructures		120	27	147
Metrics		69	11	80
Open Access	10	178	37	225
Open Science	1	377	41	419
Research Data		6		6
Science		71	31	102
Skills		154	11	165
Total	**64**	**1,542**	**270**	**1,876**

Table 2. Co-occurrence* of queries.

Topic A	Topic B										
	Citizen Science	Copyright	FAIR	Infrastructures	Metrics	Open Access	Open Science	Research Data	Science	Skills	Total
Citizen Science	77		1	1			2				81
Copyright		57	10		5	2	4			2	80
FAIR	1	10	186	6	9		6	1	4	6	229
Infrastructures	1		6	73		3	9	2		4	98
Metrics		5	9		44					2	60
Open Access		2		3		97	1		2	4	109
Open Science	2	4	6	9		1	157			10	189
Research Data			1	2				5			8
Science			4			2			36	1	43
Skills		2	6	4	2	4	10		1	79	108
Total	**81**	**80**	**229**	**98**	**60**	**109**	**189**	**8**	**43**	**108**	**1,005**

* Co-occurrence of a topic with itself, shows the number of sole occurrences in a tweet, i.e. where it was the only topic that appears in the document.

many comments – it was the second more popular – , only one of the tweets was labeled as negative.

Additionally, we detected query co-occurrence. Co-occurrence happens when topic *a* and topic *b*, regardless of their sentiment polarity, appear in the same document, i.e. tweet. Table 2 shows that otherwise conceptually distant topics, such as FAIR, Metrics and Copyright, are likely to co-occur. This shows however that the LIBER community is concerned about the copyright and impact aspects of FAIR data management. Similarly, Open Science was associated with the human and technical dimensions that its advocates take into account, as the professional skills and the infrastructures were in the spot of discussion.

In NLP, phrases that were recognised as important were classified as themes. Semantria detected the most descriptive noun phrases and assigned to them sentiment polarities. Table 3 presents the top themes according to their polarities. As one can see, the theme of 'Open Access' has been associated with both positive and negative mentions, indicating that there is a range of reactions to the developments of the field.

4.2. Hashtags

Regarding the hashtag #Liber2018, there were only 19 negative comments from a total of 5,530 comments. The comments can be grouped in five categories: users' comments concerned with keynote speakers' impression, the experience of conference's events, the support that the organisational team offered, the general atmosphere and the performance of facilities. Table 4 provides indicative tweets about each one of these categories.

Table 3. Theme's polarities.

Most negative	Neutral	Most positive
Change proposal	Science ouverte	Open access
Open science	Ouvert aux publications	Universal software source code library
Opinion piece	Une obligation	Open data
Open access	Decide que nous	French ministry
Causing time loss	Research libraries	Ambitious plan

Table 4. Organisational comments.

Comment example	Description
@LIBERconference Kick off with the French Minister Dr. Vidal explaining elegantly the basics of #openscience #liber2018	Keynote speaker
#liber2018 conference dinner at a fantastic venue!	Social events
….the #liber2018 post-conference excursion today has been very interesting! it took us to the @VillaCavrois….	Social events
Superb evening hosted by #liber2018 at @PBALille—Manet, Agache and Brueghel among my favourites, plus wonderful Rodin sculptures. And great company of course.	Social events
…..we appreciate your #Liber2018 support and dedication so much!...	Organizational support
What a great atmosphere at our #liber2018 session for newcomers!	General atmosphere
Team Lille, what a great job! Absolutely wonderful conference #liber2018 Thank you so much!	General atmosphere
Anyone selling fans or ice cream could make a killing now! #meltinginlille #liber2018	Facilities
Oh #NMBS how I hate thee #lostconnection #delay #LIBER2018	Facilities

One of the most positive hashtags was #saveyourinternet. This was related to the DSM Directive and the risks for internet freedom that the Directive was introducing. We infer that that the hashtag was associated with acts of encouragement and fight for the aversion of the proposed articles. These articles had their respective hashtags, #article11 and #article13, been detected as negatively charged. The most positive hashtag was associated with the concept of innovation, namely #innovativedissemination, while a similar hashtag, this of #disruption, invoked negative feelings. Another positively charged hashtag was this of #collaboration, which is principal to the LIBER strategy.

We noticed that initially positively charged comments could be transferred to the other extreme and be re-charged negatively. For instance, #openscience was a positively expressed hashtag, but its co-occurrence with a mention of the complaint filing against Elsevier, a publisher criticised heavily for 'closed' practices, gave to it negative attributes and its sentiment was propagated through a much favourited retweet during the conference timeline. This shows that the hashtags can be interpreted differently in various contexts.

Two specific sessions, namely the Knowledge Cafe and the Library Innovation Award, were highly appreciated by the attendees, as their respective hashtags, #knowledgecafe and #libraryinnovationsaward, received positive comments. While it is safe to conclude that the positive comments of the latter are quite natural, due to the nature of the session, which is a rewarding one, the organisational complexity and intensity of both parts of the Knowledge Cafe seem to have been unnoticed by the participants and left them good impressions.

4.3. Language constructs

As almost two thirds of the tweets were in English, it was evident that tweeting in the context of an international conference dictated the use of English language, which was followed by the language of the host country, French, and German, which was spoken by a considerable amount of participants.

Additionally, we analysed the syntactic context of the tweets and detected negation and intensification terms. Negation refers to a term that assigns negative tone to the words of its context, whereas intensification refers to adverbs or adverbial phrases that give emphasis to their context. Participants use more intensifying terms for positive reasons, such as 'absolutely', 'admittedly', 'all', 'already', 'easily' and so on. On the other hand, negation is expressed with a limited set of words, including only no, not, nothing, while the term 'really' belonged to both sides.

Finally, we explored the use of emoticons, which are very popular in the Twitter platform, as these can condense sentiment information in a visual way. However, we found that they are rarely producing the desired outcome and most of the times they are placed in tweets with neutral sentiment content. The emoticons that are mostly used are portraying happy faces, see -) and :3, grinning face (:d), winking faces, see ;) and ;-) and heart (<3).

5. Conclusions

The current study attempted to shed light on what Reinhardt et al. (2009) called "organisational enhancements". Sentiment analysis can be used both as conference assessment and planning tool. Conference committees could

enhance the dialogue of the conference participants on topics that are loaded with both negative and positive polarities in order to give the necessary platform to the community to express and mature their opinions.

Our analysis indicated that 'Copyright' and 'Open Access' issues, although constantly present in the conference agenda, are so polarised that they have the potential to remain within the focal lens of LIBER's community. The copyright related tweets carried a lot of intense sentiments, as at the time of the conference there was a heated discussion about the new Directive on Digital Single Market. On the other hand, comments on Research Data were quite few, with rather lukewarm reactions, indicating that LIBER's community has a rather concrete opinion about this topic. Similarly, the Open Science issues have a concrete interest for the LIBER community, without though raising any high feelings. The co-occurrence of themes shows that several of the topics that LIBER has high in its agenda can be intertwined. The case of FAIR, Metrics and Copyright, which belong to different strategic directions, shows that there is margin for common work from the working teams of LIBER.

Furthermore, we must keep in mind that a conference is not only a venue of discussion, but also an experience. Such experience is affected by organisational issues which include the peripheral events, the facilities and equipment, the staff's and volunteers' support and the overall atmosphere. Every host organization should have in mind previous malfunctions and work towards their resolution that will conclude to positive testimonials. Any problem occurred may affect the image of the conference and, accordingly, have an impact on the retention of attendees (Lee & Back, 2008).

Acknowledgments

References

Borgmann, H., Woelm, J.-H., Merseburger, A., Nestler, T., Salem, J., Brandt, M. P., … Loeb, S. (2016). Qualitative Twitter analysis of participants, tweet strategies, and

tweet content at a major urologic conference. *Canadian Urological Association Journal = Journal de l'Association Des Urologues Du Canada*, *10*(1–2), 39–44. Retrieved December 28, 2019, from https://cuaj.ca/index.php/journal/article/view/3322/2524.

Çalı, S., & Balaman, Ş. Y. (2019). Improved decisions for marketing, supply and purchasing: Mining big data through an integration of sentiment analysis and intuitionistic fuzzy multi criteria assessment. *Computers & Industrial Engineering*, *129*, 315–332. https://doi.org/10.1016/J.CIE.2019.01.051.

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, *28*(2), 15–21. https://doi.org/10.1109/MIS.2013.30.

Desai, T., Shariff, A., Shariff, A., Kats, M., Fang, X., Christiano, C., & Ferris, M. (2012). Tweeting the meeting: An in-depth analysis of Twitter activity at Kidney Week 2011. *PLoS One*, *7*(7), e40253, 1–9. https://doi.org/10.1371/journal.pone.0040253.

Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2019). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*, article in press. https://doi.org/10.1016/J.INDMARMAN.2019.08.003.

Kimmons, R., & Veletsianos, G. (2016). Education scholars' evolving uses of twitter as a conference backchannel and social commentary platform. *British Journal of Educational Technology*, *47*(3), 445–464. https://doi.org/10.1111/bjet.12428.

Lee, J.-S., & Back, K.-J. (2008). Attendee-based brand equity. *Tourism Management*, *29*(2), 331–344. https://doi.org/10.1016/J.TOURMAN.2007.03.002.

Mahrt, M., Weller, K., & Peters, I. (2014). Twitter in scholarly communication. In K. Weller, A. Burns, J. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and Society*, Digital formations, vol. 89 (pp. 399–410). New York: Peter Lang Publishing.

Mishori, R., Levy, B., & Donvan, B. (2014). Twitter use at a family medicine conference: analyzing #STFM13. *Family Medicine*, *46*(8), 608–614. Retrieved December 28, 2019, from http://www.ncbi.nlm.nih.gov/pubmed/25163039.

Moore, M. T. (2017). Constructing a sentiment analysis model for LibQUAL+ comments. *Performance Measurement and Metrics*, *18*(1), 78–87. https://doi.org/10.1108/PMM-07-2016-0031.

Papachristopoulos, L., Ampatzoglou, P., Seferli, I., Zafeiropoulou, A., & Petasis, G. (2019). Introducing sentiment analysis for the evaluation of library's services effectiveness. *Qualitative and Quantitative Methods in Libraries*, *8*(1), 99–110. Retrieved December 28, 2019, from http://78.46.229.148/ojs/index.php/qqml/article/view/515.

Poria, S., Gelbukh, A., Cambria, E., Yang, P., Hussain, A., & Durrani, T. (2012). Merging SenticNet and WordNet-Affect emotion lists for sentiment analysis. In *2012*

IEEE 11th International Conference on Signal Processing (Vol. 2, pp. 1251–1255). https://doi.org/10.1109/ICoSP.2012.6491803.

Reinhardt, W., Ebner, M., Beham, G., & Costa, C. (2009). How people are using Twitter during conferences. In V. Hornung-Prähauser & M. Luckmann (Eds.), *Creativity and Innovation Competencies on the Web. Proceedings of the 5th EduMedia* (pp. 145–156). Salzburg: Salzburg Research. Retrieved December 28, 2019, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.148.1238&rep=rep1&type=pdf.

Ross, C., Terras, M., Warwick, C., & Welsh, A. (2011). Enabled backchannel: conference Twitter use by digital humanists. *Journal of Documentation*, *67*(2), 214–237. https://doi.org/10.1108/00220411111109449.

Seidel, M.-D. L. (2018). The role of conferences in the emergence of developmental professional culture. *Journal of Management Inquiry*, *27*(2), 149–153. https://doi.org/10.1177/1056492617726271.

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, *311*, 18–38. https://doi.org/10.1016/J.INS.2015.03.040.

Wood, L.C., Reiners, T., & Srivastava, H. S. (2015, September 25). *Exploring sentiment analysis to improve supply chain decisions*. SSRN. https://doi.org/10.2139/ssrn.2665482.

Wu, D. D., Zheng, L., & Olson, D. L. (2014). A decision support approach for online stock forum sentiment analysis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *44*(8), 1077–1087. https://doi.org/10.1109/TSMC.2013.2295353.

Notes

1	https://tags.hawksey.info/.

2	https://www.lexalytics.com/.