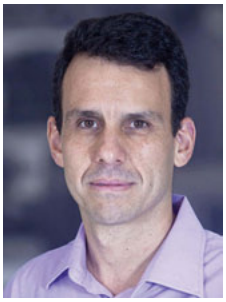
Sentiment Analysis and Knowledge Organization: An Overview of the International Literature†

Lígia Maria Arruda Café* and Renato Rocha Souza**

*Universidade Federal de Santa Catarina, Departamento de Ciência da Informação, Florianopolis, Santa Catarina Brazil, <ligiarrudacafe@gmail.com>
**Fundação Getulio Vargas—School of Applied Mathematics, Praia de Botafogo, 190, 3o andar, Rio de Janeiro Brazil 22250-900, <renato.souza@fgv.br>

Lígia Maria Arruda Café holds a PhD in linguistics from Laval University, Canada. She is Professor at the Department of Information Science at Federal University of Santa Catarina (UFSC), Brazil. Her research and teaching interests are terminology and knowledge organization.

Renato Rocha Souza holds a PhD in information science and is currently Professor at the School of Applied Mathematics, Fundação Getulio Vargas, Brazil. He is also a researcher at the Department of Organization and Theory of Information at the Federal University of Minas Gerais (UFMG). His main research interests are knowledge representation, information retrieval, natural language processing and machine learning



Café, Lígia Maria Arruda and Renato Rocha Souza. 2017. "Sentiment Analysis and Knowledge Organization: An Overview of the International Literature." *Knowledge Organizaation* 44(3): 199-214. 32 references.

Abstract: Knowledge organization (KO) as an activity is, among other meanings, a process for conceptual modeling of knowledge domains that produces a consensual abstraction model of this domain with a particular purpose. It adopts a myriad of techniques to analyze and build efficient knowledge organization systems, and one of these techniques is called sentiment analysis (SA) or opinion mining, which is emerging as promising and useful in a variety of ways. It is based in NLP and AI algorithms, and aims at identifying opinions and emotions toward any person, organization or subject; evaluating them as positive or negative, in both binary and graded fashions. This study sought to show various aspects of the implementation of SA for knowledge organization tasks as registered in the scientific literature. We began with exploratory bibliographic research and built a corpus of 91 scientific papers, written in English, selected in the LISA Database, between 2000 to 2016. We analyzed these papers and extracted title, year of publication, author(s) and institution(s), title of the journal where they were published, keywords, the LISA classification code, methods/techniques adopted and its application areas. Our main findings are that theoretical papers still prevail, which may indicate a field in the early stages. We found many institutions and authors from Asia, which points to a new shift in world expertise. We concluded that SA is still a novelty in the KO field, being slowly adopted as an aid to the main tasks, as document classification.

Received: 25 November 2016; Revised 8 March 2017; Accepted: 9 March 2017-03-11

Keywords: sentiment analysis, opinion mining, knowledge organization systems

† The authors wish to thank CNPq for the concession of a research sponsorship (PQ).

1.0 Introduction

Sentiment analysis (SA) is a technique used to identify the attitude or "sentiment" of subjects toward a company, a person, a product or any other kind of information entity. Commonly, SA applies techniques derived from natural language processing, text analysis and computational linguistics to extract subjective information and apply a grading scale or binary classification. SA has been widely used for a variety of tasks, including marketing and consumer research, political analysis and social media strategies.

We could perhaps postulate that SA constitutes a (somewhat) new domain, in the sense defined by Smiraglia (2012, 114): "A domain is best understood as a unit of analysis for the construction of a KOS. That is, a domain is a group with an ontological base that reveals an underlying teleology, a set of common hypotheses, epistemological consensus on methodological approaches, and social semantics." Based on our assumption that knowledge organization (KO) and SA are closely related, we have conducted exploratory research, aiming at investigating the potential of the set of techniques collectively named as sen-

timent analysis when applied to the KO area. In order to have a glimpse of the current knowledge on the topic, we have searched the Knowledge Organization Literature database (http://www.isko.org/lit.html) by the terms "sentiment" or "opinion," and this attempt yielded 29 papers, both articles in journals and proceedings. The topics ranged from "math KO theory" (Zhao and Yuan 2011); "automatic or semi-automatic indexing methods" (Li et al 2011; Hu and Li 2011; Na et al 2011; Sarvabhotla et al 2011; Fan 1997); "automatic online indexing" (Xia et al 2011); "automatic classification" (Na et al 2004); "psychology" (Chang 2012); "word/sentence meaning" (Hu and Li 2012); "NLP problems/methods" (Zhang et al 2009; Hu and Li 2011; Hung et al 2012; Morsy and Rafea 2012); "online access/queries" (Kang et al 2012; Na et al 2005, King et al 2009); "search engines" (Sleem-Amer et al 2012; Ku et al 2009; Belacel et al 2010; Kang et al 2013); "online systems evaluation" (Na and Thet 2009); "terminological work" (Sarvabhotla et al 2011); "multilingual IR" (Boiy et al 2009); "various NBM indexing" (Song 2014); "facet analysis" (Vechtomova 2010) and "science methodology" (Kiduk 2015). We also found two papers presenting literature reviews on the topic (Xia et al 2011; Hongwei et al 2010) that captured a spike on this publications after 2011, but the latter is more specific and both are outdated with regard to our work. No papers were published in *Knowledge Organization*. Having concluded that the topic is present in the KO domain, we then decided to query the LISA database to update the state of art and enlarge the current literature review, helping to make SA more explicit and familiar to the KO researchers and practitioners.

2.0 Materials and methods

We have searched the international literature for works that claim to adopt these SA encompassing techniques, shedding light in theoretical or practical experiences that seek to solve KO-related problems. Therefore, the study presented here adopted as a methodological approach a literature review, retrieving documents from the Library and Information Science Abstracts (LISA) database, an international abstracting and indexing tool designed for information specialists, library professionals and researchers. We decided, as our search strategy, to disregard the KO literature review previously described, as we do not know which criteria were used to establish the set of documents and also because we could compare the number of documents that would be found at the same time in both databases. We did not invest in other databases at this stage, but we plan to broaden this scope in the future.

We started over with a new search using the term "sentiment analysis," restricting the results to papers written in English and published in academic peer-reviewed journals in the period between 2000 and June of 2016. One alternative to widen the coverage would be to explore the keywords derived from the retrieved articles, as they could serve as semantic connections among the analyzed domains. For now, we consider the results as sufficient. Refusing articles in proceedings was intentional and aimed at observing the research that was more mature and got to be published in academic journals. The search retrieved 141 references and, after eliminating repeated and irrelevant items, and also those with no free access to the full text from within our institutions, we selected 91 articles that constituted the final research corpus. From these, only 4 were also found in the KO literature. Each article was thoroughly examined by observing the following items: the title of the paper, year of publication, author(s) and institution(s), title of the journal, the keywords and the LISA classification code assigned. When keywords were not present, we have taken those assigned by the LISA platform, with consequent loss of specificity. After analyzing the subject and the content of the documents, we have extracted the methods/techniques adopted in the research and its application areas. We have then built a controlled vocabulary as shown in Table 1.

These application areas were chosen to be generic in spite of losing a bit of specificity. Nevertheless, they represent the richness of the domains in which sentiment analysis is being applied. In the next sections, we present some visualizations of the data gathered along with interpretations and comments. We also made some cross-comparisons, highlighting the main characteristics of cutting-edge SA research related to KO.

3.0 Results

In this section, we are going to present the main results of the research divided in topics. The first is the diachronic analysis in which we present the evolution of the number of papers by year in the scope of the research. Then follows an analysis of institutions, countries, authors and journals to characterize the extrinsic aspects of the papers dealing with the subject of sentiment analysis. We finally analyze the LISA classification code assigned to the papers, the keywords, the application areas and the methods/techniques to illustrate the intrinsic aspects of the papers.

3.1 Diachronic analysis

The identification of the papers' publishing chronology demonstrates the evolution of the research in the KO domain. Our search was limited to papers that were published and indexed after 2000, but the first few papers we see were published from 2007 on (less than ten years

Application area	Description
Non Applied (Theoretical)	Theoretical aspects of the field: experiments with news techniques or variations of known techniques. The papers may describe a small application, but only for the sake of demonstrating the technique.
Social Media	Social media environments such as Twitter, blogs, YouTube communities, etc.
Reviews Analysis	Corpora of users' reviews targeting a service, institution or product.
Lexicon Building	Construction of dictionaries or other lexical resources for sentiment analysis.
Consumer Opinion	Analysis of the opinion/polarity targeting a service, institution, or product.
Language Specific	Language specific aspects of sentiment analysis
Political Analysis	Sentiment analysis applied to political analysis
Document Classification	Sentiment analysis applied to document classification
E-commerce	E-commerce strategies using sentiment analysis to characterize consumers
E-learning	Sentiment analysis as a strategy to evaluate users on digital learning environments
Gender Identity	Sentiment analysis as an aid to classify gender
Stock Market	Stock market and financial markets predictions based on sentiment analysis
Business Strategies	Sentiment analysis as an aid for business strategic modelling.
Citation Analysis	Application of sentiment analysis on the vicinity of citations on texts to identify the polarity towards the cited authors
Stylistic Analysis	Sentiment analysis as a component of stylistic analysis

Table 1. Description of the application areas.

ago). It may indicate the degree of novelty of the topic among other KO subjects. We can clearly see that there was an increase in publications from 2007 to 2013, and, after 2014, the trend seemed to fade a little bit. We have observed an unexpected low number of publications in 2015. This effect could have been caused by a lag in the indexing process. Nevertheless, in spite of having taken into account only half of the 2016 publications, the numbers increase again. We may assume they probably won't surpass 2014 totals if the trend continues, as we can see in Figure 1.

The numbers show that SA and KO have been sharing a great many publications as we will confirm when examining the subjects.

3.2 Institutions

The institutions where the authors work is an important factor to take into account, and it can serve as a proxy on priority research areas and whether an institution recognizes a particular field. It is believed that the more an institution publishes on a particular subject, with papers targeting high impact factor journals, the greater its prestige in the scientific community. There were 125 institutions when counting all occurrences from all the authors, but if we take only the first authors of each paper, we have 68 distinct institutions. Table 2 shows the 10 institutions that have more than two articles published, comprising less than 15% of the total. The most productive are Nanyang Technological University (8 papers), the University of Wolverhampton (6 papers) and the University of Jaen (4 papers), respectively located in Singapore, the United Kingdom and Spain.

3.3 Countries

The relationship between the number of articles and institutional affiliation of the authors is an interesting fact that shows the degree of dispersion of research on a particular topic. As can be seen in Table 2 and also in Table 3, scientific production in Singapore (11) and the United Kingdom (8) is focused in a small number of institutions among which are those mentioned in the previous section, indicating a high concentration of research in sentiment analysis in these countries. Alternatively, there is China, where albeit we can see a significant number of papers (14), these are dispersed among 13 different institutions. The panorama seen in China resembles the USA that, in turn, presents 10 papers published by 9 institutions. We can also verify relative dispersion in Spain (11

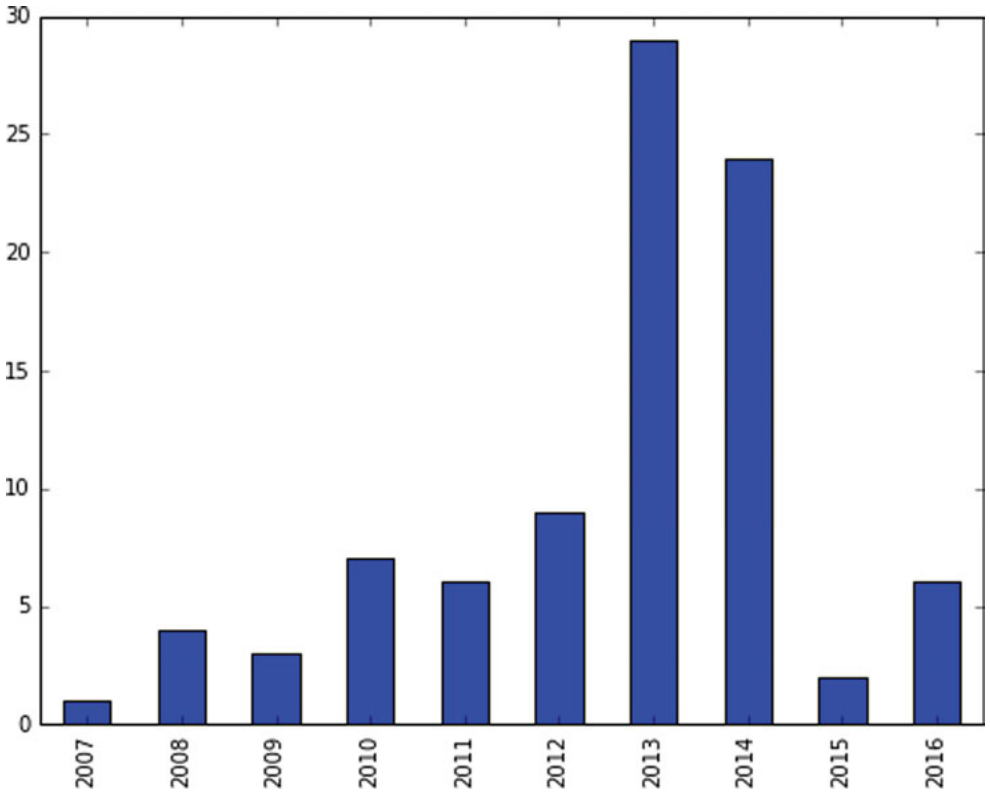


Figure 1. Years of publication.

Institution	Number of papers	Country
Nanyang Technological University	8	Singapore
University of Wolverhampton	6	United Kingdom
University of Jaén	4	Spain
National University of Singapore	3	Singapore
National Taiwan University	2	Taiwan
Isfahan University of Technology	2	Iran
University of Seville	2	Spain
University of Kentucky	2	USA
Chung Yuan Christian University	2	Taiwan
Xi'an Jiaotong University	2	China

Table 2. Number of papers per institution.

papers produced by 7 institutions); Italy (4 papers in 4 institutions); Korea (3 papers in 3 institutions); and Germany (3 papers in 3 institutions). These latter countries also show a scatter chart, but with a much lower production than the other countries. The observed dispersion may indicate that the subject of technical application of SA in knowledge organization is in process of consolidation. Table 3 also registers Taiwan with 6 papers published by 4 institutions.

Countries	Number of papers	Institutions
China	14	13
Spain	11	7
Singapore	11	2
USA	10	9
United Kingdom	8	3
Taiwan	6	4
Italy	4	4
Korea	3	3
Germany	3	3

Table 3. Number of papers and institution by countries.

3.4 Authors

The authors are responsible for the scientific communication of an area, and its reputation reveals who are the agents that act in a scientific knowledge network. Moreover, the identification of those who publish more can be a strong indicator of leadership in the area. It is exactly what Table 4 highlights when it presents the 7 most productive authors who appear as first authors in articles, ranging from 4 to 7 papers by author. It also emphasizes that the 7 authors from Table 4 represent circa 8% of the first authors in the articles, which indicates a very concentrated scientific productivity in a small number of large producers. Meadows (1974) already found this phenomenon in scientific productivity, reporting that there are few researchers who concentrate a prominent quantitative production and there are many smaller producers. The ranking leadership belongs to Erik Cambria, currently assistant professor at Nanyang Technological University (NTU) at the School of Computer Science and Engineering (Singapore), which ranks in first place with the largest number of papers (7), followed by Jin-Cheon Na (also from Singapore) and Mike Thelwall, both with 6 publications each; Tun Thura Thet with 5 papers and the other 3 authors with 4 papers each.

3.5 Journals

The scientific paper is the means by which experts disclose the results of research. In the process of scientific communication, content is validated by peers and then published in the leading scientific communication channel, the academic journals. This procedure aims at conferring legitimacy to the scientific knowledge produced within the researchers' communities. Given its importance, Table 5 shows the 10 journals with 2 or more published articles on SA. Among them, *Knowledge-Based Systems* is the one showing the highest number of articles (18). It is also observed in the number of journals that 5 of them are genuinely from the field of information science, namely: *Journal of the American Society for Information Science and Technology*, *Information Processing & Management*, *Journal of Information Science*, *Online Information Review* and the *Journal of the Association for Information Science and Technology*. Among these, 3 have significant records of production, proving the importance of SA techniques in the field of knowledge organization. They are: *Journal of the American Society for Information Science and Technology* with 12 articles; *Information Processing & Management* with 10 articles; and *Journal of Information Science* with 9 articles.

Author	Number of papers	Country
Cambria, Erik	7	Singapore
Na, Jin-Cheon	6	Singapore
Thelwall, Mike	6	United Kingdom
Thet, Tun Thura	5	Singapore
Schuller, Björn	4	Germany
Buckley, Kevan	4	United Kingdom
Khoo, Christopher Soo-Guan	4	Singapore

Table 4. Number of papers per author.

Journal	Number of papers
Knowledge-Based Systems	18
IEEE Intelligent Systems	15
Journal of the American Society for Information Science and Technology	12
Information Processing & Management	10
Journal of Information Science	9
Online Information Review	4
Internet Research	3
First Monday	3
International Journal on Semantic Web and Information Systems	2
Journal of the Association for Information Science and Technology	2

Table 5. Number of papers per journal.

3.6 LISA classification code

Broughton (2004) states that "classification is a fundamental tool in the process of organizing a collection and the complementary process of searching for and retrieving information." The LISA database has its own classification code, but those assigned to the papers in the context of SA were not much of help in identifying the specificities of the subject of papers. Being a generic classification, it fails in discriminating among the papers' subjects; we can perceive that almost 74% of the papers were classified as "automatic text analysis, automatic indexing, machine translation," which barely represents the actual topics being presented. Table 6 illustrates the high concentration among the classes.

Lisa Classification Code	Number of papers
13.13 Automatic Text Analysis, Automatic Indexing, Machine Translation	67
14.11 Communications and Information Technology—Networks	6
13.14 Information Storage and Retrieval-Searching	5
10.0 Information Communication	4
14.19 Computer Applications	3

Table 6. Number of papers by LISA classification code.

It is curious to see the apparent lack of distinguishing in class 14.11 (Communications and Information Technology—Networks) and class 10.0 (Information Communication), and also the emptiness of class 14.19 (Computer Applications) when compared with class 13.13 (Automatic Text Analysis, Automatic Indexing, Machine Translation) that, surprisingly, belongs to another main class. It is inherent in the classification schemas to lag behind the knowledge domain they represent, but, nevertheless, they may face the need to adapt to new subjects rapidly. We could not get information on the process of updating the classification codes or how often new terms are included and revised.

3.7 Keywords

The analysis of the keywords, however, is very informative and seems to capture the main trends in the domain as we can see in Table 7. Besides presenting some synonymies to the main topic in the most frequent words—"opinion mining," "sentiment classification," "attitudes," "emotions" and even "aspect detection,"—it is curious to verify that the term we have used in the search task, "sentiment analysis," was not present among the keywords appearing more than twice. The other group of frequent keywords show how connected the field is with computer science tasks and topics. Words like "machine learning," "data mining," "natural language processing," "automatic text analysis," "text mining," "automatic classification," "sentic computing" and "algorithm" are very common within the frequent keywords.

Keyword	Number of papers
opinion mining	13
data mining	13
natural language processing	13
automatic text analysis	11
social network	10
emotions	9
sentiment classification	6
machine learning	6
methods	5
twitter	5
social media	5
text mining	4
blogs	4
automatic classification	4
attitudes	3
semantics	3
context	3
sentic computing	3
feature selection	3
product reviews	3
self-training	3
algorithm	3
content analysis	3
e-commerce	3
reviews	3
aspect detection	3
classification	3

Table 7. Number of papers by keyword assigned.

Figure 2. Keywords tag cloud.

Application Area	Number of papers	Institutions
Non Applied (Theoretical)	25	18
Social Media	13	13
Reviews Analysis	11	8
Lexicon Building	8	7
Consumer Opinion	7	7
Political Analysis	6	6
Language Specific	6	5
Document Classification	4	4
E-commerce	2	2
E-learning	2	2
Stock Market	2	2
Gender Identity	2	1
Business Strategies	1	1
Citation Analysis	1	1
Stylistic Analysis	1	1

Table 8. Number of papers and institutions by application area.

We can also identify a group of frequent keywords related both to the contexts of application (i.e., "social network," "twitter," "blogs," "e-commerce," "reviews") and specific techniques such as "content analysis" and "classification." Among the frequent keywords, we also find some generic terms, like "methods" and "context," which hardly convey any meaning at all for classification purposes. Figure 2 illustrates both the frequent keywords and also some other not-so-common keywords.

3.8 Application areas

Table 8 illustrates the application areas related both to the number of papers and the institutions. We can see that

Application Area	Institution	Country
Document Classification	National Institute of Information and Communications Technology	Japan
	University of California	USA
	University of Sfax	Tunísia
	University of Waterloo	Canada

Table 9. Institutions by application area: Document Classification.

Application Area	Institution	Country
Social Media	Chonbuk National University	Korea
	Harbin Institute of Technology	China
	Hof University of Applied Sciences	Germany
	Korea University	Korea
	The University of Waikato	New Zealand
	Universidad Politécnica de Madrid	Spain
	Universidad Politécnica de Valencia	Spain
	University of Amsterdam	Netherlands
	University of Jaén	Spain
	University of Macedonia	Macedonia
	University of Regina	Canada
	University of Southern Indiana	USA
	University of Wolverhampton	United Kingdom

Table 10. Institutions by application area: Social Media.

circa 27% of the papers are purely theoretical and do not present real world applications. When they do, they are using artificial corpora just for the sake of demonstrating the algorithms and techniques presented. We also noted that circa 14% of the papers deal with experiments in social networks and social media, which is very coherent with the natural application envinronment of these techniques.

Regarding the institutions, we can see that the topics are scattered among them with no concentration of topics in a few places. We have chosen 4 specific topics by their importance to the field of knowledge organization to take a closer look.

Beginning with "Document Classification" in Table 9, we can observe 4 papers in 4 different institutions and countries, which may show the relatively low correlation of the topic with sentiment analysis. The same cannot be said with respect to "Social Media," which, as shown before, is scattered among 14 institutions and concentrates circa 14% of the papers. As can be seen in Table 10, this topic presents Spain and Korea as the most prominent countries.

The area of "Political Analysis" is also scattered through institutions and countries, showing no leadership from either point of view, as can be seen in Table 11. It is curious, though, to see Turkey and Israel as players in this field, which may denote the fact that the recent turmoil in both countries may have been fostering innovative ways of dealing with politics.

Lastly, we present the major area of application (Table 12) with countries and also number of papers in the only area which is not as scattered as the previous ones. With more than one fourth of the papers, it encompasses a surprisingly low number of distinct countries (only 9), being

Application Area	Institution	Country
Political Analysis	Hacettepe Universitesi	Turkey
	Nanyang Technological University	Singapore
	National Institute of Informatics	Japan
	San Diego State University	USA
	University of Haifa	Israel
	University of Torino	Italy

Table 11. Institutions by application area: Political Analysis.

Application Area	Institution	Country	Papers
Non Applied (Theoretical)	National University of Singapore	Singapore	3
	University of Wolverhampton	United Kingdom	3
	Nanyang Technological University	Singapore	2
	University of Kentucky	USA	2
	University of Seville	Spain	2
	Fondazione Bruno Kessler	Italy	1
	Institute of Computing Technology	China	1
	La Trobe University	Australia	1
	Nanjing University of Science and Technology	China	1
	National Cheng Kung University	Taiwan	1
	Open University	United Kingdom	1
	Shanxi University	China	1
	Sun Yat-sen University	China	1
	UNED	Spain	1
	Universiti Teknologi Malaysia	Malaysia	1
	University of Jaén	Spain	1
	Universitá degli Studi di Salerno	Italy	1
	Xi'an Jiaotong University	China	1

Table 12. Institutions by application area: Non Applied (Theoretical).

China, Spain, Italy and Singapore the leaders in the theoretical studies.

Again, this corroborates the fact the subject is being developed away from the main players of the Ivy League and European equivalents, which motivates all sort of interpretations.

3.9 Methods

We have identified a great number of techniques applied on the SA tasks. Most of them are derived from the machine learning and natural language processing domains. We have tried to classify the papers with regard to those

techniques, but the difference in the specificity of the description made the effort dubious. We have, otherwise, extracted the techniques and proposed a taxonomy, using Brownlee (2013), Azure Machine Learning Team (2015) and Scikit-learn Developers (2016) as a basis. The complete taxonomy is presented as an appendix, and the terms that were found in our literature are underlined.

4.0 Conclusion

After presenting the data we have drawn from the papers, some conclusions could be made. Perhaps the most important is that SA is still a novelty in the knowledge organization field, being slowly adopted as an aid to the tasks such as document classification. We can envision information retrieval systems in the future taking into account user's emotional perspectives as something to be valued when retrieving documents. It is still to be seen whether the trend will continue the coming years, given the drop in production in recent years.

Regarding the data per se, we can highlight the theoretical papers still being the norm, which may indicate a field in its early stages. There are many institutions and authors from Asia (36% of papers), which points to a new axis in world expertise, on which Nanyang Technological University and National University of Singapore are the most productive. However, no institution emerges as a hub in the field, given the dispersion of topics among them.

We believe there are some biases in the information presented, as our corpus was created from searches on LISA only. An idea for future work would be to extend the search to other databases and to try the search and retrieval of papers in areas beyond information science, to verify the occurrence of the topic in correlated areas, such as computer science and management sciences.

References

Azure Machine Learning Team. 2015. "Azure Machine Learning Studio Capabilities Overview." http://aka.ms/MLStudioOverview

Belacel, Nabil, Juan A. Carretero and Prabhat Mahanti. 2010. "AI and Opinion Mining, Part 2." *IEEE intelligent systems* 25, no. 4: 72-79.

Boiy, Erik and Marie-Francine Moens. 2009. "A Machine Learning Approach to Sentiment Analysis in Multilingual Web Texts." *Information Retrieval* 12: 526-58.

Broughton, Vanda. 2004. *Essential Classification*. London: Facet Publishing.

Brownlee, Jason. 2013. "A Tour of Machine Learning Algorithms." http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

Chang, Moon-Soo. 2012. "Empirical Sentiment Classification using Psychological Emotions and Social Web Data." *Korean Institute of Intelligent Systems* 22: 563-69.

Fan, David P. 1997. "Computer Content Analysis of Press Coverage and Prediction Of Public Opinion for the 1995 Soverignty Referendum in Quebec." *Social Science Computer Review* 15: 351-66.

Gang, Li, Wang Zhongyi and Kou Guangzeng. 2011. "Research on Computing Method for Sentiment's Intensity in Sentiment Classification."*Journal of the China Society for Scientific and Technical Information* 30: 292-98.

Hongwei, Wang, Pei Xie and Liu N.K. James. 2010. "Literature Review of Sentiment Classification on Web Text." *Journal of the China Society for Scientific and Technical Information* 29: 931-38.

Hu, Yi and Wenjie Li. 2011. "Document Sentiment Classification by Exploring Description Model of Topical Terms." *Computer Speech and Language* 25: 386-403.

Hung, Chili, Chih-Fong Tsai and Hsinyi Huang. 2012. "Extracting Word-of-mouth Sentiments via SentiWordNet for Document Quality Classification." *Recent Patents on Computer Science* 5: 145-52.

Huosong, Xia, Tao Min, Wang Yi and Wei Xiang. 2011. "The Influence of Stop Word Removal on the Chinese Text Sentiment Classification based on SVM Technology." *Journal of the China Society for Scientific and Technical Information* 30: 347-52.

Kang, Beomil, Min Song and Whasun Jho. 2013. "A Study on Opinion Mining of Newspaper Texts based on Topic Modeling." *Journal of the Korean Society for Library and Information Science* 47: 315-34.

Kang, Hanshoon, Seong Joon Yoo and Dongil Han. 2012. "Senti-lexicon and Improved Naive Bayes Algorithms for Sentiment Analysis of Restaurant Reviews." *Expert Systems with Applications* 39: 6000-10.

Kiduk, Yang. 2015. "Fusion Approach to Targeted Opinion Detection in Blogosphere." *Journal of Korean Library and Information Science Society* 46: 321-44.

Kim, Sea Woo, Chin-Wan Chung and DaeEun Kim. 2009. "An Opinion-based Decision Model for Recommender Systems." *Online Information Review* 3: 584-602.

Ku, Lun-Wei, Hsiu-Wei Ho and Hsin-His Chen. 2009. "Opinion Mining and Relationship Discovery using CopeOpi Opinion Analysis System." *Journal of the American Society for Information Science and Technology* 60: 1486-1503.

Library and Information Science Abstracts. Lisa Classification Codes. http://proquest.libguides.com/ld.php?content_id=3028387

Meadows, Arthur. Jack. 1974. *Communication in Science*. London: Butterworths.

Morsy, Sara A. and Ahmed Rafea. 2012. "Improving Document-level Sentiment Classification using Con-

textual Valence Shifters." *Lecture Notes in Computer Science* 7337: 253-58.

Na, Jin-Cheon and Tun Thura Thet. 2009. "Effectiveness of Web Search Results for Genre and Sentiment Classification." *Journal of Information Science* 35: 709-27.

Na, Jin-Cheon, Christopher S.G. Khoo, Syin Chan and Norraihan Bte Hanzah. 2005. "Sentiment-based Search in Digital Libraries." In *JCDL '05 Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries 7-11 June 2005 Denver, CO*. New York: ACM, 143-44.

Na, Jin-Cheon, Haiyan Sui, Christopher Khoo, Syin Chan and Yunyn Zhou. 2004. "Effectiveness of Simple Linguistic Processing in Automatic Sentiment Classification of Product Reviews." In *Knowledge Organization and the Global Information Society: Proceedings of the Eighth International ISKO Conference, London, England, July 13-16, 2004,* ed. Ia C. McIlwaine. Advances in Knowledge Organization 9. Würzburg: Ergon Verlag, 49-54.

Na, Jin-Cheon; Tun Thura Thet, Arie Hans Nasution and Fauzi Munif Hassan. 2011. "A Sentiment-Based Digital Library of Movie Review Documents Using Fedora." *The Canadian Journal of Information and Library Science* 35: 307-37.

Sarvabhotla, Kiran, Prasad Pingali and Vaseduva Varma. 2011. "Sentiment Classification: A Lexical Similarity based Approach for Extracting Subjectivity in Documents." *Information Retrieval* 14: 337-53.

Scikit-learn Developers. 2016. "Scikit-learn algorithm cheat-sheet." http://scikit-learn.org/stable/tutorial/machine_learning_map/

Sleem-Amer, Mona, Ivan Bigorgne, Stéphanie Brizard, Leeley Daio Pires Dos Santos, Yacine El Bouhairi, Bénédicte Goujon, Stéphane Lorin, Claude Martineau, Loïs Rigouste, and Lidia Varga. "Intelligent Semantic Search Engines for Opinion and Sentiment Mining." In *Next Generation Search Engines: Advanced Models for Information Retrieval*. Hershey, Pa.: IGI Global, 2012, 191-215.

Song, Min. 2014. "An Ontology-based Approach to Sentiment Classification of Mixed Opinions in Online Restaurant Review." *한국비블리아 발표논집* 2014: 73-89.

Vechtomova, Olga. 2010. "Facet-Based Opinion Retrieval from Blogs. *Information Processing & Management* 46: 71-88.

Xia, Huosong, Liuyan Peng and Menglin Yu. 2011. "A Literature Review on Automatic Sentiment Classification." *Journal of the China Society for Scientific and Technical Information* 30: 530-39.

Zhang, Changli, Daniel Zeng, Jiexun Li, Fei-Yue Wang and Wanli Zuo. 2009. "Sentiment Analysis of Chinese Documents: From Sentence to Document Level." *Journal of the American Society for Information Science and Technology* 60: 2474-87.

Zhao, Narisa and Li Yuan. 2011. "Online-review-based Fuzzy Computing and Inference of Consumer Sentiment." *Journal of the China Society for Scientific and Technical Information* 30: 417-23.

Appendix 1: Corpus of reference

Argamon, Shlomo, Casey Whitelaw, Paul Chase, Sobhan Raj Hota, Navendu Garg and Shlomo Levitan. 2007. "Stylistic Text Classification Using Functional Lexical Features." *Journal of the American Society for Information Science and Technology* 58: 802-22.

Bae, Younggue and Hongchul Lee. 2012. "Sentiment Analysis of Twitter Audiences: Measuring Positive or Negative Influence of Popular Twitterers." *Journal of the American Society for Information Science and Technology* 63: 2521-35.

Bagheri, Ayoub, Ayoub Saraee and Franciska De Jong. 2014. "ADM-LDA: An Aspect Detection Model Based on Topic Modelling Using The Structure of Review Sentences." *Journal of Information Science* 40: 621-36.

Bagheri, Ayoub, Ayoub Saraee and Franciska De Jong. 2013. "Care More About Customers: Unsupervised Domain-Independent Aspect Detection for Sentiment Analysis of Customer Reviews." *Knowledge-Based Systems* 52: 201-13.

Bertola, Frederico and Viviana Patt. 2016. "Ontology-Based Affective Models to Organize Artworks in the Social Semantic Web." *Information Processing & Management* 52: 139-62.

Bosco, Cristina, Viviana Patti and Andrea Bolioli. 2013 "Developing Corpora for Sentiment Analysis: The Case of Irony and Senti-TUT." *IEEE Intelligent Systems* 28: 55-63.

Bravo-Marquez, Felipe, Marcelo Mendoza and Barbara Poblete. 2014. "Meta-Level Sentiment Models for Big Social Data Analysis." *Knowledge-Based Systems* 69: 86-99.

Cambria, Erik, Björn Schuller, Bing Liu, Haixun Wang and Catherine Havasi. 2013. "Knowledge-Based Approaches to Concept-Level Sentiment Analysis." *IEEE Intelligent Systems* 28, no. 2: 12-4.

Cambria, Erik, Björn Schuller, Bing Liu, Haixun Wang and Catherine Havasi. 2013a. "Statistical Approaches to Concept-Level Sentiment Analysis." *IEEE Intelligent Systems* 28, no. 3: 6-9.

Cambria, Erik, Björn Schuller, Yunqing Xia and Catherine Havasi. 2013. "New Avenues in Opinion Mining and Sentiment Analysis." *IEEE Intelligent Systems* 28, no. 2: 15-21.

Cambria, Erik, Yangqiu Song, Haixun Wang and Newton Howard. 2014. "Semantic Multidimensional Scaling for Open-Domain Sentiment Analysis." *IEEE Intelligent Systems* 29, no. 2: 44-51.

Carrillo-de-Albornoz, Jorge and Laura Plaza. 2013. "An Emotion-Based Model of Negation, Intensifiers, and Modality for Polarity and Intensity Classification." *Journal of the American Society for Information Science and Technology* 64: 1618-33.

Chen, Li, Feng Wang, Luole Qi and Fengfeng Liang. 2014. "Experiment on Sentiment Embedded Comparison Interface." *Knowledge-Based Systems* 64: 44-58.

Cho, Heeryon, Songkuk Kim, Jongseo Lee and Jong-Seok Lee. 2014. "Data-Driven Integration of Multiple Sentiment Dictionaries for Lexicon-Based Sentiment Classification of Product Reviews." *Knowledge-Based Systems* 71: 61-71.

Cruz, Fermín L., Carlos G. Vallejo, Fernando Enríquez and José A. Troyano. 2012. "PolarityRank: Finding an Equilibrium between Followers and Contraries in a Network". *Information Processing & Management* 48: 271-82.

Cruz, Fermín L. 2012. "Feature-Based Opinion Extraction: A Practical, Domain-Adaptable Approach." *AI Communications* 25: 369-71.

García-Moya, Lisette, Henry Anaya-Sánchez and Rafael Berlanga-Llavori. 2013. "Retrieving Product Features and Opinions from Customer Reviews." *IEEE Intelligent Systems* 28, no. 3: 19-27.

Fortuny, Enric Junqué de, Tom De Smedt, David Martens and Walter Daelemans. 2014. "Evaluating and Understanding Text-based Stock Price Prediction Models. *Information Processing & Management* 50: 426-41.

Garcia-Silva, Andrés., Víctor. Rodriguez-Doncel and Oscar Corch. 2013. "Semantic Characterization of Tweets Using Topic Models: A Use Case in the Entertainment Domain." *International Journal on Semantic Web and Information Systems* 9, no. 3: 1-13.

Greaves, Felix, Daniel Ramirez-Cano, Christopher Millett, Ara Darzi and Liam Donaldson. 2013. "Use of Sentiment Analysis for Capturing Patient Experience From Free-Text Comments Posted Online." *Journal of Medical Internet Research* 15, no. 11: 1-9.

Habernal, Ivan, Tomáš Ptáčˇ and Josef Steinberger. 2014. "Supervised Sentiment Analysis in Czech Social Media." *Information Processing & Management* 50: 693-707.

Hajmohammadi, Mohammad Sadegh, Roliana Ibrahim and Ali Selamat. 2014. "Bi-View Semi-Supervised Active Learning for Cross-Lingual Sentiment Classification." *Information Processing & Management* 50: 718-32.

He, Yulan and Deyu Zhou. 2011. "Self-Training from Labeled Features for Sentiment Analysis." *Information Processing & Management* 47: 606-16.

Hoeber, Orland, Larena Hoeber, Maha El Meseery, Odoh Kenneth and Radhika Gopi. 2016. "Visual Twitter Analytics (Vista)." *Online Information Review* 40, no. 1: 25-41.

Hung, Chihli. 2008. "A Personalized Word of Mouth Recommender Model." *Webology* 5, no. 3.

Hung, Chihli and Hao-Kai Lin. 2013. "Using Objective Words in SentiWordNet to Improve Word-of-Mouth Sentiment Classification." *IEEE Intelligent Systems* 28, no. 2: 47-54.

Jiménez-zafra, Salud M., M. Teresa Martín-Valdivia, Eugenio Martínez-Cámara and L. Alfonso Urenã-López. 2016. "Combining Resources to Improve Unsupervised Sentiment Analysis at Aspect-Level." *Journal of Information Science* 42: 213-29.

Jurafsky, Dan, Victor Chahuneau, Bryan R. Routledge and Noah A. Smith. 2014. "Narrative Framing of Consumer Sentiment in Online Restaurant Reviews. *First Monday* 19, no. 4: 15-9.

Kalampokis, Evangelos, Efthimios Tambouris and Konstantinos Tarabanis. 2013. "Understanding the Predictive Power of Social Media." *Internet Research* 23: 544-59.

Kato, Yoshikiyo, Sadao Kurohashi and Kentaro Inui. 2008. "Classifying Information Sender of Web Documents." *Internet Research* 18: 191-203.

Kechaou, Zied, Ali Wali, Mohamed Ben Ammar, Hichem Karray and Adel M. Alimi. 2013. "A Novel System for Video News' Sentiment Analysis." *Journal of Systems and Information Technology* 15: 24-44.

Khoo, Christopher Soo-guan, Armineh Nourbakhsh and Jin-cheon Na. 2012. "Sentiment Analysis of Online News Text: A Case Study of Appraisal Theory." *Online Information Review* 36: 858-78.

Kwon, A-Rong and Kyung-Soon Lee. 2013. "Opinion Bias Detection with Social Preference Learning in Social Data." *International Journal on Semantic Web and Information Systems* 9: 57-76.

Li, Gang and Fei Liu. 2012. "Application of a Clustering Method on Sentiment Analysis." *Journal of Information Science* 38: 127-39.

Li, Sheng-Tun and Fu-Ching Tsai. 2013. "A Fuzzy Conceptualization Model for Text Mining with Application in Opinion Polarity Classification." *Knowledge-Based Systems* 39: 23-33.

Loia, Vincenzo and Sabrina Senatore. 2014. "A Fuzzy-Oriented Sentic Analysis to Capture the Human Emotion in Web-Based Content." *Knowledge-Based Systems* 58: 75-85.

Luo, Wenjuan, Fuzhen Zhuang, Weizhong Zhao, Qing He and Zhongzhi Shi. 2015. "QPLSA: Utilizing Quad-Tuples for Aspect Identification and Rating." *Information Processing & Management* 51: 25-41.

Malo, Pekka, Ankur Sinha, Pekka Korhonen, Jyrki Wallenius and Pyry Takala. 2014. "Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts." *Journal of the Association for Information Science and Technology* 65: 782-96.

Malouf, Robert and Tony Mullen. 2008. "Taking Sides: User Classification for Informal Online Political Discourse." *Internet Research* 18: 177-90.

Martínez-Cámara, Eugenio, M. Teresa Martín-Valdivia, M. Dolores Molina-González and José M. Perea-Ortega. 2014. "Integrating Spanish Lexical Resources by Meta-Classifiers for Polarity Classification." *Journal of Information Science* 40: 538-54.

Miao, Qingliang, Qiudan Li and Daniel Zeng. 2010. "Fine-Grained Opinion Mining by Integrating Multiple Review Sources." *Journal of the American Society for Information Science and Technology* 61: 2288-99.

Montejo-Ráez, A., M. C. Díaz-Galiano, F. Martínez-Santiago and L. A. Ureña-López. 2014. "Crowd Explicit Sentiment Analysis." *Knowledge-Based Systems* 69: 134-9.

Na, Jin-Cheon, Tun Thura Thet, Arie Hans Nasution and Fauzi Munif Hassan. 2011. "A Sentiment-Based Digital Library of Movie Review Documents Using Fedora." *The Canadian Journal of Information and Library Science* 35: 307-37.

Na, Jin-Cheon and Tun Thura Thet. 2009. "Effectiveness of Web Search Results for Genre and Sentiment Classification." *Journal of Information Science* 35: 709-26.

Na, Jin-Cheon, Tun Thura Thet and Christopher S. G. Khoo. 2010. "Comparing Sentiment Expression in Movie Reviews from Four Online Genres." *Online Information Review* 34: 317-38.

Peetz, Maria-Hendrike, Maartende Rijke and Rianne Kaptein. 2016. "Estimating Reputation Polarity on Microblog Posts." *Information Processing & Management* 52: 193-216.

Perez Rosas, Verónica Pérez, Rada Mihalcea and Louis-Philippe Morency. 2013. "Multimodal Sentiment Analysis of Spanish Online Videos." *IEEE Intelligent Systems* 28, no. 3: 38-45.

Pon, R. K., A. F. Cardenas, D. J. Buttler and T. J. Critchlow. 2010. "Measuring the Interestingness of Articles in a Limited User Environment." *Information Processing & Management* 47: 97-116.

Popescu, Octavian and Carlo Strapparava. 2014. "Time Corpora: Epochs, Opinions and Changes." *Knowledge-Based Systems* 69: 3-13.

Poria, Soujanya, Alexander Gelbukh, Erik Cambria, Amir Hussain and Guang-Bin Huang. 2014. "EmoSenticSpace: A Novel Framework for Affective Common-Sense Reasoning." *Knowledge-Based Systems* 69: 108-23.

Poria, Soujanya, Alexander Gelbukh, Amir Hussain, Newton Howard, Dipankar Das and Sivaji Bandyopadhyay. 2013. "Enhanced SenticNet with Affective Labels for Concept-Based Opinion Mining." *IEEE Intelligent Systems* 28, no. 2: 31-8.

Poria, Soujanya, Erik Cambria, Grégoire Winterstein and Guang-Bin Huang. 2014. "Sentic Patterns: Dependency-Based Rules for Concept-Level Sentiment Analysis." *Knowledge-Based Systems* 69: 45-63.

Rao, Yanghui. 2016. "Contextual Sentiment Topic Model for Adaptive Social Emotion Classification." *IEEE Intelligent Systems* 31: 41-7.

Reyes, Antonio, Paolo Rosso and Tony Veale. 2013. "A Multidimensional Approach for Detecting Irony in Twitter." *Language Resources and Evaluation* 47: 239-68.

Rigo, Sandro José, Isa Maria da Rosa Alves and José Luis Victória Barbosa. 2015. "A Linguistic Approach to Identify the Affective Dimension Expressed in Textual Messages." *International Journal of Information and Communication Technology Education* 11: 44-59.

Rill, Sven, Dirk Reinel, Jörg Scheidt and Roberto V. Zicari. 2014. "PoliTwi: Early Detection of Emerging Political Topics on Twitter and the Impact on Concept-Level Sentiment Analysis." *Knowledge-Based Systems* 69: 24-33.

Rushdi-Saleh, Mohammed, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López and José M. Perea-Ortega. 2011. "OCA: Opinion Corpus for Arabic." *Journal of the American Society for Information Science and Technology* 62: 2045-54.

Salas-Zárate, María del Pilar, Estanislao López-López, Rafael Valencia-García, Nathalie Aussenac-Gilles, Ángela Almela and Giner Alor-Hernández. 2014. "A Study on Liwc Categories for Opinion Mining in Spanish Reviews." *Journal of Information Science* 40: 749-60.

Sela, Shlomi, Tsvi Kuflik and Gustavo S. Mesch. 2008. "Changes in the Discourse of Online Hate Blogs: The Effect of Barack Obama's Election in 2008." *First Monday* 17, no. 11: 1-11.

Small, Henry. 2011. "Interpreting Maps of Science Using Citation Context Sentiments: A Preliminary Investigation." *Scientometrics* 87: 373-88.

Smiraglia, Richard P. 2012. "Epistemology of Domain Analysis." In *Cultural Frames of Knowledge*, ed. Richard P. Smiraglia and Hur-li Lee. Würzburg: Ergon Verlag, 111-24.

Sommer, Stefan, Andreas Schieber, Kai Heindrich and Andreas Hilbert. 2012. "What is the Conversation About? A Topic-Model-Based Approach for Analyzing Customer Sentiments in Twitter." *International Journal of Intelligent Information Technologies* 8: 10-25.

Suzuki, Takafumi. 2009. "Extracting Speaker-Specific Functional Expressions from Political Speeches Using Random Forests in Order to Investigate Speakers' Political Styles." *Journal of the American Society for Information Science and Technology* 60: 1596-606.

Syed, Afraz Z., Muhammad Aslam and Ana Maria Martinez-Enriquez. 2014. "Associating Targets with Sentiunits: A Step Forward in Sentiment Analysis of Urdu Text." *Artificial Intelligence Review* 41: 535-61.

Thelwall, Mike. 2010. "Emotion Homophily in Social Network Site Messages." *First Monday* 15, no. 4: 1-7.

Thelwall, Mike, Kevin Buckley, Georgios Paltoglou, Di Cai and Arvid Kappas. 2010. "Sentiment Strength Detection in Short Informal Text." *Journal of the American Society for Information Science and Technology* 61: 2544-58.

Thelwall, M. and K. Buckley. 2013. "Topic-Based Sentiment Analysis for the Social Web: The Role of Mood and Issue-Related Words." *Journal of the American Society for Information Science and Technology* 64: 1606-17.

Thelwall, Mike, Kevin Buckley and Georgios Paltoglou. 2011. "Sentiment in Twitter Events." *Journal of the American Society for Information Science and Technology* 62: 406-18.

Thelwall, Mike, Kevin Buckley and Georgios Paltoglou. 2012. "Sentiment Strength Detection for the Social Web." *Journal of the American Society for Information Science and Technology* 63: 163-73.

Thelwall, Mike, David Wilkinson and Sukhvinder Uppal. 2010. "Data Mining Emotion in Social Network Communication: Gender Differences in MySpace." *Journal of the American Society for Information Science and Technology* 61: 190-9.

Thet, Tun Thura, Jin-Cheon Uppal and Christopher S. G. Khoo. 2010. "Aspect-Based Sentiment Analysis of Movie Reviews on Discussion Boards." *Journal of Information Science* 36: 823-48.

Tian, Feng, Pengda Gao, Longzhuang Li, Weizhan Zhang, Huijun Liang, Yanan Qian and Ruomeng Zhao. 2014. "Recognizing and regulating e-learners' emotions based on interactive Chinese texts in e-learning systems." *Knowledge-Based Systems* 55: 148-64.

Tsai, Angela Charng-Rurng, Chi-En Wu, Richard Tzong-Han Tsai and Jane Yung-jen Hsu. 2013. "Building a Concept-Level Sentiment Dictionary Based on Commonsense Knowledge." *IEEE Intelligent Systems* 28, no. 2: 22-30.

Vechtomova, Olga. 2010. "Facet-Based Opinion Retrieval from Blogs. *Information Processing & Management* 46: 71-88.

Wang, Suge, Deyu Li, Lidong Zhao and Jiahao Zhang. 2013. "Sample Cutting Method for Imbalanced Text Sentiment Classification Based on BRC." *Knowledge-Based Systems* 37: 451-61.

Weichselbraun, Albert, Stefan Gindl and Arno Scharl. 2013. "Extracting and Grounding Contextualized Sentiment Lexicons." *IEEE Intelligent Systems* 28, no. 2: 39-46.

Wöllmer, Martin, Felix Weninger, Tobias Knaup, Björn Schuller, Congkai Sun, Kenji Sagae and Louis-Philippe Morency. 2013. "YouTube Movie Reviews: Sentiment Analysis in an Audio-Visual Context." *IEEE Intelligent Systems* 28, no. 3: 46-53.

Wu, Chi-En and Richard Tzong-Han Tsai. 2014. "Using Relation Selection to Improve Value Propagation in a Conceptnet-Based Sentiment Dictionary." *Knowledge-Based Systems* 69: 100-7.

Xia, Rui, Chengqing Zong, Xuelei Zong and Erik Cambria. 2013. "Feature Ensemble Plus Sample Selection: Domain Adaptation for Sentiment Classification." *IEEE Intelligent Systems* 28, no. 3: 10-8.

Xianghua, Fu, Liu Guo, Guo Yanyan and Wang Zhiqiang. 2013. "Multi-Aspect Sentiment Analysis for Chinese Online Social Reviews Based on Topic Modeling and Hownet Lexicon." *Knowledge-Based Systems* 37: 186-95.

Xu, Tao, Qinke Peng and Yinzhao Cheng. 2012. "Identifying the Semantic Orientation of Terms Using S-HAL for Sentiment Analysis." *Knowledge-Based Systems* 35: 279-89.

Yan, Gongjun, Wu He, Jiancheng Shen and Chuanyi Tang. 2014. "A Bilingual Approach for Conducting Chinese and English Social Media Sentiment Analysis". *Computer Networks* 75: 491-503.

Yang, Dong-Hui and Guang Yu. 2013. "A Method of Feature Selection and Sentiment Similarity for Chinese Micro-Blogs." *Journal of Information Science* 39: 429-41.

Yildiz, Mete and Ayşegül Saylam. 2013. "E-Government Discourses: An Inductive Analysis." *Government Information Quarterly* 30: 141-53.

Yu, Liang-Chih, Jheng-Long Wu, Pei-Chann Chang and Hsuan-Shou Chu. 2013. "Using a Contextual Entropy Model to Expand Emotion Words and their Intensity for the Sentiment Classification of Stock Market News." *Knowledge-Based Systems* 41: 89-97.

Yu, Ning. 2013. "Domain Adaptation for Opinion Classification: A Self-Training Approach." *Journal of Information Science Theory and Practice* 1: 10-26.

Yu, Ning. 2014. "Exploring Co-Training Strategies for Opinion Detection." *Journal of the Association for Information Science and Technology* 65: 2098-110.

Zhang, Changli, Daniel Zeng, Jiexun Li, Fei-Yue Wang and Wanli Zuo. 2009. "Sentiment Analysis of Chinese Documents: From Sentence to Document Level." *Journal of the American Society for Information Science and Technology* 60: 2474-87.

Zhang, Pu and Zhongshi He. 2013. "A Weakly Supervised Approach to Chinese Sentiment Classification Using Partitioned Self-Training." *Journal of Information Science* 39: 815-31.

Zhang, Xianfeng, Yang Yu, Hongxiu Li and Zhangxi Lin. 2016. "Sentimental Interplay Between Structured and Unstructured User-Generated Contents." *Online Information Review* 40: 119-45.

Zhang, Zhu. 2008. "Weighing Stars: Aggregating Online Product Reviews for Intelligent E-Commerce Applications." *IEEE Intelligent Systems* 23, no. 5: 42-9.

Zheng, Xiaolin, Zhen Lin, Xiaowei Wang, Kwei-Jay Lin and Meina Song. 2014. "Incorporating Appraisal Expression Patterns into Topic Modeling for Aspect and Sentiment Word Identification." *Knowledge-Based Systems* 61: 29-47.

Appendix 2: Machine Learning Tentative Taxonomy

Learning Styles
- Supervised learning
  - Classification
  - Regression
    - Iterative regression
- Unsupervised learning
- Semi-supervised learning
- Reinforcement learning.
- Transduction
- Learning to learn

Algorithm Types
- Linear Classifiers
  - MultiClass Classifier
- Regression
  - Logistic Regression
    - Perceptron
- Support Vector Machine (SVM)
  - Boundary region cutting (BRC)
- Quadratic Classifiers
- Ordinary Least Squares Regression (OLSR)
  - Linear Regression
  - Logistic Regression
  - Stepwise Regression
  - Multivariate Adaptive Regression Splines (MARS)
  - Locally Estimated Scatterplot Smoothing (LOESS)
- Instance-based
  - k-Nearest Neighbour (kNN)
  - Learning Vector Quantization (LVQ)
  - Self-Organizing Map (SOM)
  - Locally Weighted Learning (LWL)
- Regularization
  - Ridge Regression
  - Least Absolute Shrinkage and Selection Operator (LASSO)
  - Elastic Net
  - Least-Angle Regression (LARS)
- Decision Tree
  - Classification and Regression Tree (CART)
  - Iterative Dichotomiser 3 (ID3)
  - C4.5 and C5.0
    - J48
  - Chi-squared Automatic Interaction Detection (CHAID)
  - Decision Stump
  - M5
  - Conditional Decision Trees
- Bayesian
  - Naive Bayes
    - Gaussian Naive Bayes
    - Multinomial Naive Bayes
  - Averaged One-Dependence Estimators (AODE)
  - Bayesian Belief Network (BBN)
  - Bayesian Network (BN)
- Clustering
  - k-Means
  - k-Medians
  - Expectation Maximisation (EM)
  - Fuzzy clustering
  - Hierarchical Clustering
  - Topic Modeling
    - Contextual sentiment topic model (CSTM)
  - Latent Dirichlet Allocation
    - AEP-based Latent Dirichlet Allocation (AEP-LDA) model
- Association Rule Learning
  - Apriori algorithm
  - Eclat algorithm
- Artificial Neural Network
  - Perceptron
  - Back-Propagation
  - Hopfield Network
  - Radial Basis Function Network (RBFN)
  - Deep Learning
    - Deep Boltzmann Machine (DBM)
    - Deep Belief Networks (DBN)
    - Convolutional Neural Network (CNN)
    - Stacked Auto-Encoders
- Dimensionality Reduction
  - Principal Component Analysis (PCA)
  - Principal Component Regression (PCR)
  - Partial Least Squares Regression (PLSR)
  - Sammon Mapping
  - Multidimensional Scaling (MDS)
  - Projection Pursuit
  - Linear Discriminant Analysis (LDA)
  - Mixture Discriminant Analysis (MDA)
  - Quadratic Discriminant Analysis (QDA)
  - Flexible Discriminant Analysis (FDA)
  - t-distributed stochastic neighbor embedding (t-SNE)
- Ensemble Algorithms
  - Boosting
  - Bootstrapped Aggregation (Bagging)
  - AdaBoost

  - LogitBoost
  - Stacked Generalization (Blending)
  - Gradient Boosting Machines (GBM)
  - Gradient Boosted Regression Trees (GBRT)
  - Random Forest
- Preparation and Evaluation
  - Feature extraction
    - Bag-of-Words (BoW)
      - Term Frequency (TF)
      - Term Frequency/Inverse Document Frequency (TF/IDF)
    - Bag-of-N-Gram (BoNG)
  - Feature selection
  - Cross-validation
  - Accuracy evaluation
  - Performance measures
- Related Algorithms
  - Computational intelligence (evolutionary algorithms, etc.)
  - Computer Vision (CV)
  - Graphical Models
  - Natural Language Processing (NLP)
    - Text Mining
    - Porter stemming

Product Types
- Recommender Systems
- Social Media Optimization (SMO)
- Sentiment Analysis
  - Crowd Explicit Sentiment Analysis (CESA)
  - Multimodal Sentiment Analysis
  - Concept-based opinion mining
  - Polarity Classification
  - Intensity Classification
  - Speech-based emotion recognition
  - Sentiment Analysis Software and Algorithms
    - PolarityRank
    - SenticNet
    - SentiSense
    - Sentic computing
    - SentiUnits
    - SentiStrength
    - SentiWordNet
    - Electronic word of mouth (eWOM)
    - EmoSenticSpace
    - Sentiment Hyperspace Analogue to Language (S-HAL)