A TEXT MINING CLASSIFICATION OF ARTICLES ON SOCIAL SUSTAINABILITY IN SUPPLY CHAINS

Tayeb Basta

Faculty, College of Engineering and Technology, University of Science and Technology of Fujairah, Fujairah, P.O. Box: 2202, United Arab Emirates tayebasta@gmail.com

Mohamed Basta

École de technologie supérieure, 1100, rue Notre-Dame Ouest, Montréal H3C 1K3 Québec, Canada mohamed.basta.1@etsmtl.net

Abstract

The purpose of this paper is to categorize articles on social sustainability in supply chains based on the methodology used to address the topic. By reading all articles one by one, the paper considers the use of text mining as an alternative to manual classification. This work reduces the time it takes to find information in a body of literature from months to minutes.

Practitioners should ensure that the data language used to describe the topic is correct and complete. The findings of the paper assist researchers and postgraduate students in completing time-consuming tasks such as literature reviews.

The current work could be improved by incorporating contextual information classification rather than frequency classification.

Keywords: Social risk; social sustainability; supply chain; systems thinking; industry approach; text mining.

1. Introduction

Other issues related to supply chain management include social risk and social sustainability in supply chains. They are, however, frequently addressed in terms of labor rights and working conditions [1, 2, 3]. This could be due to their inability to be managed through measurement, monitoring, and prediction [4, 5]. When considering social risk and social sustainability in supply chains, one of the most important questions to ask is what approaches have been taken to address such a problem. The authors of [6] conducted an analysis on a collection of research articles dealing with social risk and social sustain-ability in supply chains and concluded that industry standards, such as corporate social responsibility and social life cycle assessment, represent the overwhelming solutions used to address the problem, whereas the use of systems thinking is very limited in this context. Regardless of the quality of the conclusions they reached, their work was extremely difficult; while they collected the data set of articles using a search engine, most of their research work was done manually by reading at least the title, abstract, introduction, and conclusion of each article to reach their conclusions.

In the current study, we employ text mining techniques to perform the same analysis on the methodologies used to address the issues of social risk and social sustainability in supply chains. We developed a method, converted the articles to text files, and ran the method to obtain similar results to [6], but in a matter of seconds rather than working days, if not months.

The remainder of the paper is structured as follows: Section 2 explains the purpose of the current project. Section 3 describes the method used to mine the supply chain articles. Section 3 delves into the text mining method in depth. Section 4 de-scribes the data used in the current experimental study. Section 5 discusses and analyzes the experimental results. Finally, the paper concludes in section 6.

DOI: 10.21817/indjcse/2022/v13i5/221305201 Vol. 13 No. 5 Sep-Oct 2022 1378

Tayeb Basta et al. / Indian Journal of Computer Science and Engineering (IJCSE)

e-ISSN: 0976-5166 p-ISSN: 2231-3850

2. Objective of The Paper

The current project aims to automate the content analysis of articles on social sustainability in supply chain management using machine learning techniques. The goal of the analysis is to classify articles into four different categories based on the methodology used to solve the problem. These methodologies are systems thinking, industry approach, a combination of systems thinking and industry approach, and undefined whenever neither of the two methodologies is used.

The authors of proposed twelve questions to guide their analysis in [6]. They use these questions to help uncover trends and patterns in what supply chains are doing to address social risk and social sustainability issues. In this paper, we choose six of those questions that we believe are more practical to include in a machine learning process:

- (1) How prevalent is the application of the industry approach in the literature?
- (2) What is the most used terminology in the literature for industry approaches?
- (3) How prevalent is explicit use of systems thinking in the literature?
- (4) What is the most explicitly used terminology in the literature for systems thinking?
- (5) Is the terminology of systems thinking and the industry approach used in tandem?
- (6) How common is the use of approaches in the literature that are not associated with explicit systems thinking or industry approaches?

3. The Method

Data mining is the process of extracting valuable patterns and relationships hidden in large amounts of data. Whereas data mining methods anticipate a highly structured data format, text mining, a subset of data mining, seeks to uncover hidden patterns and trends in semi-structured and unstructured text documents [7, 8]. Text mining refers to a set of technologies that "convert text into numbers" so that powerful algorithms can be applied to large document databases [9].

Email and spam filtering [10], sentiment analysis [11], marketing [12], social network analysis [13], and financial statement fraud detection [14] are examples of text mining applications. Text mining tools have also been used to help biomedical researchers gain access to biological knowledge encoded in massive expert-curated biological databases [8].

3.1. The text mining process

The current work's designed method is a keyword-based text mining process comprised of three sub-processes. PDF articles are converted to text files during the preprocessing step.

- (1) The first sub-process loops through all the files, reads the articles, and replaces certain industry approach terms that contain numbers with letter words. It also removes preferences, as we discovered through simple tests that the references in articles have a significant impact on the mining results.
- (2) The second sub-process, which consists of the standard text mining tasks. It includes text mining operators for tokenizing file content, removing stop words, converting content to lower case, removing tokens with lengths less than three, and re-removing all tokens except those that are related to approaches used to solve the problem of social sustainability in supply chains. These are the methodologies of industry approach and systems thinking. The following operator will generate n-grams. N-grams are contiguous sequences of n items from a token sequence [15]. We chose to generate 4-grams in this work because certain terms in the used data language consist of four-word phrases: "social life cycle assessment," "social environment and governance criteria," and "quality function deployment for the environment." The next operator will filter out the newly generated n-grams attributes, keeping only those related to social sustainability in supply chains and those related to industry approach and systems thinking. The result of this sub-process is a dataset of rows (examples) with values for the aforementioned attributes.
- (3) The third sub-process includes operators for filtering examples (i.e., dataset rows) based on the presence of specific attributes.
- The first operator of the third sub-process indicates the presence of supply chain terminology; it sums up the occurrence frequencies of the set of attributes logistics, value_chain, value_chains, supply_chain, supply_chain, supply_chain, supply_chain, supply_chains under the common name supplychain. The second operator indicates the presence of social sustainability; it sums up the occurrence frequencies of the attributes social_risk, social_risks, social_sustainability under the common name sustainability.
- The next two operators are to sum up the frequency of terminology in both the industry approach and systems thinking.
- We provide a second version of the method that generates frequencies of terms from both the industry approach and systems thinking for benchmarking the current method's results against the findings of [6].

DOI: 10.21817/indjcse/2022/v13i5/221305201 Vol. 13 No. 5 Sep-Oct 2022 1379

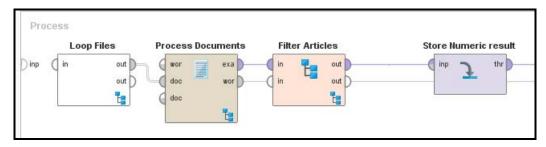


Fig. 1. Main text mining process.

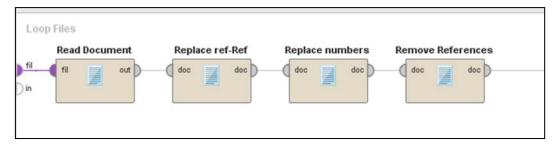


Fig. 2. Loop Files sub-process.

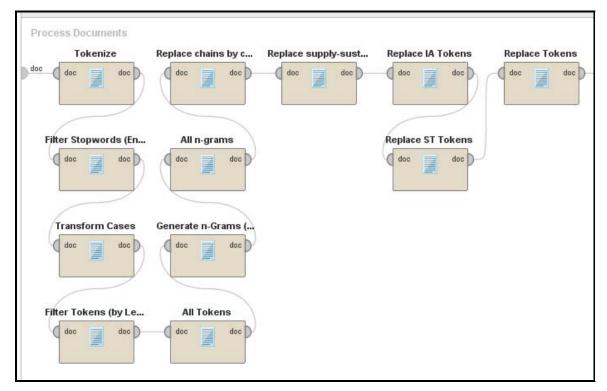


Fig. 3. Processing Documents sub-process.

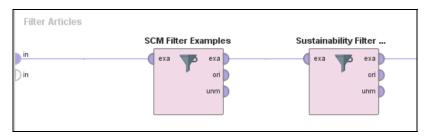


Fig. 4. Filter Articles sub-process.

Tayeb Basta et al. / Indian Journal of Computer Science and Engineering (IJCSE)

e-ISSN: 0976-5166 p-ISSN: 2231-3850

3.2. The data language

The data language used in this analysis is explicitly described as follows:

- (1) Systems thinking
- Definition: includes all sampling units that use one or more systems thinking terms explicitly.
- Terminology: operations research, operational research, system dynamics, complexity theory, organizational cybernetics, chaordic systems thinking, strategic assumption surfacing and testing, interactive planning, soft systems thinking, critical systems heuristics, team syntegrity, post-modern systems thinking, socio-technical systems, system dynamics.
- Examples: we investigate using critical systems heuristics to include all those affected by the supply chain's procedures and operations in the decision-making process.
- (2) Industry Approach
- Definition: includes all sampling units that use any of the supply chain social risk and social sustainability industry approach.
- Terminology: CSR: CSR, corporate social responsibility, corporate citizenship, corporate conscience, corporate responsibility, corporate sustainability and social responsibility, CSSR, responsible business, corporate global responsibility, CGR, social responsibility, corporate social responsiveness
 - SLCA: SLCA, S-LCA, social life cycle assessment, life cycle assessment, life cycle analysis, SIA, social impact assessment
 - ESG: ESG, environmental social and governance, responsible investing, socially sustainable measurement, social environment and governance criteria
 - Sustainable Design: quality function deployment for the environment, QFDE, sustainable design, responsible design
 - Industry Standard: industry standard, iso26000, agenda21, millennium development goals, the accountability assurance standard, aa1000as, social accountability 8000, sa8000, isae3000
 - Sustainable Reporting: GRI, global reporting initiative, sustainable reporting, stakeholder communication
- Example: using QFDE could prove effective in ameliorating the social sustainability performance of the supply chain, it allows social concerns to be incorporated during the design phase of a product or service
- (3) Undefined
- Definition: the recording unit does not state or use any of the Terminology of the variables from the first two variables explicitly and proposes an unknown or new technique(s).
- Terminology: any technique that is not enumerated in the first two variables
- Example: In paper we propose a novel framework using KPI and Scorecards to assess the supply chain's social footprint

We defined thirty-one attributes relevant to systems thinking and fifty-two attributes to industry approach based on this language.

- (1) Systems thinking attributes:
- chaordic, adaptive, complex, complexity, theory, critical, heuristics, interactive, planning, operational, operations, research, post, modern, socio, sociotechnical, social, technical, cybernetics, soft, structuralist, system, dynamics, systems, dynamic, team, syntegrity, strategic, assumption, surfacing, testing.
- (2) Industry approach attributes:
- csr, citizenship, conscience, sustainability, csr, business, global, cgr, responsibility, corporate, responsiveness, slca, life, cycle, analysis, sia, impact, assessment, esg, environmental, investing, socially, measurement, governance, criteria, quality, function, deployment, environment, qfde, responsible, design, industry, approach, isotwentysix, agendatwentyone, millennium, development, goals, assurance, standard, aathousandas, social, accountability, saeightthousand, isaethree, sustainable, gri, reporting, initiative, stakeholder, communication.

4. The Data

The current work's data is made up of articles culled from ten different databases. They were sifted through to find as many articles as possible about social risk and social sustainability in the supply chain. The articles are chosen from these databases using the same inclusion and exclusion criteria as in [6]. The criteria state that the article must be about supply chain, social risk, and social sustainability; it must not be about philanthropy, charity, or social innovation; it must have been published between 2004 and 2015; it must be written in English; and it must be available in PDF format.

The databases are American Society of Civil Engineering, Compendex, Emerald, Inspec, Proquest, Worldcat, Scopus, Web of Science, Wiley Online Library, and Science Direct. Because of its size, the Science Direct database is considered a dataset, while the others are grouped into another. The data mining model is then run on each dataset separately.

DOI: 10.21817/indjcse/2022/v13i5/221305201 Vol. 13 No. 5 Sep-Oct 2022 1381

5. RESULTS

To begin, the method excludes articles that do not address social sustainability in supply chains. The articles are then classified into one of four categories based on the approach used to address the subject in the article: industry approach, systems thinking, both approaches, and neither approach.

- (1) Industry Approach: the article employs the industry approach.
- (2) Systems Thinking: the article employs systems thinking.
- (3) Both Approaches: the article employs both the industry approach and systems thinking.
- (4) Neither Approach: the article considers social sustainability in supply chains outside the scope of both the industry approach and systems thinking.

Finally, we ran the method's second version to explicitly generate the frequencies of terms used in industry approach and systems thinking methodologies.

The data has been divided into two datasets. The first is the Science Direct database, and the second is a compilation of all the other databases into a single dataset. Each of these datasets has approximately 500 articles. To evaluate the method's performance, the method is run on each dataset separately. The method's goal is to find rational values for certain parameters. These parameters are the number of occurrences of search strings that qualify an article to belong to a specific class, as shown in the table below.

Attributes	The attribute represents	Occurrence of an attribute indicates the article is
supplychain	Logistics, supply chain, supply chains, value	addressing the supply chain management problem
	chain, value chains	
sustainability	social risk, social risks, social sustainability	addressing the social sustainability
industryapproach	Industry approach terminology	using the industry approach
systemsthinking	Systems thinking terminology	using the systems thinking approach

Table 1. Articles classification terminology.

At this point, the text mining method converts the problem from an unstructured collection of text documents to a structured problem in the form of a two-dimensional table. As illustrated in Figure 5, each row of the two-dimensional table describes a data article in terms of the frequency with which the attributes of Table 1 occur in that article.

supplychain	sustainability	industryapproach	systemsthinking	metadata_file
142.0	2.0	39.0	2.0	\Aboelmaged-2015-Adoption-of-Supply-Chain-Sus
3.0	1.0	.0	.0	\Ackrill-2012-Sweetness and Power – Public Poli.tx
12.0	1.0	101.0	1.0	\Acutt-2004-Perspectives on corporate social re.txt
17.0	1.0	23.0	.0	\Adrien-2009-International sourcing, social respon
1.0	1.0	.0	.0	\Agwu-2013-From Reluctance to Resistance- Study
84.0	24.0	3.0	1.0	\Ahmad-2015-Enterprise systems are we ready for
25.0	1.0	2.0	.0	\Ahmed-2011-Elusive challenges of e-change mana

Fig. 5. Sample of text mining model output.

What remains is to classify these examples into one of the four categories mentioned above. Figure 5 must be expressed in the considered classification context before selecting any classification mining algorithm. A zero in a cell indicates that there is no occurrence of an attribute in the example, indicating that the article does not belong to the attribute class. A number greater than zero indicates that the article belongs to that class. As a result, the table in Figure 5 corresponds to the table in Figure 6 below.

supplychain	sustainability	industryapproach	systemsthinking	metadata_file
1	1	1	1	\Aboelmaged-2015-Adoption-of-Supply-Chain-Sus
1	1	0	0	\Ackrill-2012-Sweetness and Power – Public Poli.tx
1	1	1	1	\Acutt-2004-Perspectives on corporate social re.txt
1	1	1	0	\Adrien-2009-International sourcing, social respon-
1	1	0	0	\Agwu-2013-From Reluctance to Resistance- Study
1	1	1	1	\Ahmad-2015-Enterprise systems are we ready for

Fig. 6. Structured dataset in bi-nominal data types.

We choose to run the k-means clustering algorithm with the number of clusters parameter set to four. Clustering is concerned with grouping objects that are similar to one another in one cluster but different from the objects in other clusters. The k-means clustering algorithm is an exclusive clustering algorithm, which means that each object is assigned to exactly one of a set of clusters [16].

5.1. Classification of articles

Table 2 shows the results of running our method on a single database that includes the American Society of Civil Engineers, Compendex, Emerald, Inspec, Proquest, Worldcat, Scopus, Web of Science, and Wiley Online Library.

Supply chain	Social sustainability	Industry Approach (IA)	Systems Thinking (ST)	articles	IA	ST	both	none
1	1	1	1	430	51%	8%	23%	17%
1	1	2	2	430	55%	4%	11%	29%
1	1	2	1	430	44%	22%	22%	25%
2	2	2	2	226	59%	4%	15%	22%
Previous	work	436	81%	9%	1.4%	10%		

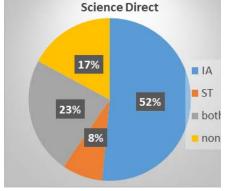
Table 2. American Society of Civil Engineering, Compendex, Emerald, Inspec, Proquest, Scopus, Web of Science, Wiley Online Library, and Worldcat databases. The first two columns indicate the filtering criteria used to consider an article is addressing the social.

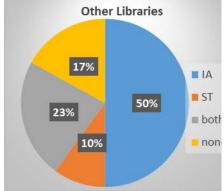
The goal of this step is to fine-tune the parameters of our data mining model so that the results are more similar to the previous work. Table 2 first two columns contain filtering criteria parameters. When supply chain = 1 and social sustainability = 1, it means that an article addresses social sustainability in supply chains if the occurrence of supply chain and social sustainability terminology is at least 1. The next two columns indicate whether the article is using the industry approach, systems thinking, both, or neither. For example, IA = 1 indicates that an article is using the industry approach if the number of occurrences of IA terms is at least one; IA = 2 indicates that an article is only using the industry approach if the number of occurrences of IA terms is at least two. The column "articles" shows how many articles in the dataset met the criteria in the first four columns. The last four columns show the percentage of articles that were addressed using the industry approach, systems thinking, both approaches, or none at all. The final row is a summary of previous work results.

We choose the values for the following parameters: supply chain=1, social sustainability=1, IA=1, ST=1. These values produce results that are similar to the previous work.

Library	Number of articles	IA	ST	both	none
Science Direct	430	51%	8%	23%	17%
Other libraries	417	50%	10%	23%	17%
Previous work	436	81%	9%	1.4%	10%

Table 3. Results of using Industry Approach and Systems Thinking in Science Direct and Other libraries databases.





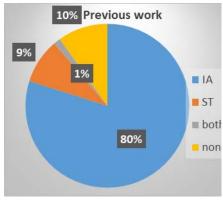


Fig. 7. Comparison of the current and previous work in terms of IA and ST usage. It appeared that the percentage of ST is the same in both works. IA usage is the highest in both; however, it looks higher in the previous work.

Table 3 values, as shown in Figure 7, indicate the following facts:

- (1) The method produces consistent results on both datasets: Science Direct and the other libraries.
- (2) Articles are filtered to see if they address social sustainability in supply chains or not by testing supply chain and social sustainability parameters. Table 2 shows that supply chain >= 1 and social sustainability >= 1 produce results that are similar to those reported in [6]. When these values are increased to 2, the number of filtered articles falls to 226; this is a significant decrease from the 436 reported in the previous work
- (3) When comparing the current solution to [6,] while there are some differences in terms of values, the order of usage of the two approaches, systems thinking and industry approach, generated by their work and the current solution is the same. In the previous work, the use of the industry approach was nine times that of the use of systems thinking, and it is nearly seven times in the current study.
- (4) The current method is frequency-based; the absence of any terminology from either methodology indicates that neither is used in the relevant article. Though some articles in previous work were classified as using the industry approach or systems thinking, the current method asserts that no terminology of the two methodologies can be found in those articles.
- (5) There is a difference in the number of articles that did not use any of the methodologies in [6] and the current study; nearly 43 articles (i.e., 10% of 436) in the former and 75 (i.e., 17% of 430) in the latter. This fact prompted us to read the 75 articles in order to discover that two of them use operations research in the form of optimization problems without mentioning the terms operation research or operational research. 15 articles did not use any of the methodologies, but they did not address social sustainability in supply chains. These have gotten past the first round of filtering because they have at least one supply chain or value chain occurrence and one occurrence of social risk or social sustainability. In conclusion, the current method is more accurate for classifying articles as not using either of the two methodologies if the first filtering phase is so accurate.
- (6) There is a noticeable divergence between the two methods in terms of the number of articles addressing social sustainability in supply chains using both methodologies together, and this is where the current method falls short.

5.2. Most explicitly used terminology

To be consistent in running the second version of the method, we opt for the occurrence of a data language term at least once in an example (article) to be explicitly considered using the respective methodology.

Tables 5 and 6, as well as Figures 8 and 9, show the percentage of industry approach and systems thinking terminology usage in all databases. These tables and figures show the results of our method, which was run on Science Direct and other libraries, as well as the results of previous work.

The percentage of terminology equal to the number of articles that contain the terminology divided by the number of articles that use a specific methodology (IA or ST). For example, if the number of articles is 10, values (article1) are CSR and SLCA, values (article2) are CSR and System Dynamics, and values (article3) are SLCA. We have two CSRs, two SCLAs, and one System Dynamics.

Figure 8 depicts the percentage of people who use Industry Approach terminology. It confirms that our method yields consistent results on both the Science Direct and other libraries databases. It also demonstrates the consistency of the current method's and previous work's results.

Terminology	CSR	ESG	Industry standard	SLCA	Sustainable design	Sustainable reporting
Abbreviation	CSR	ESG	IS	SLCA	SD	SR
Science direct	35%	3%	21%	17%	11%	13%
Other libraries	42%	5%	20%	9%	9%	15%
Previous work	40%	1%	20%	17%	2%	18%

Table 4. Most explicitly used Industry Approach terminology in all databases.

Terminology	Chaordic systems thinking	Complex adaptive systems	Complexity theory	Critical systems heuristics	Cybernetic systems	Operations research	Post modern systems	Socio-technical systems	Soft systems thinking	Structuralist systems	Systems dynamics	Gen. Systems thinking	Team syntegrity
Abbreviation	CST	CAS	CT	CSH	CS	OR	PS	STS	SST	SS	SD	GST	TS
Science direct	0%	21%	5%	1%	0%	40%	1%	7%	2%	0%	17%	8%	0%
Other libraries	0%	22%	6%	0%	0%	41%	3%	5%	3%	5%	7%	12%	0%
Previous work	3%	8%	8%	0%	3%	4%	0%	28%	3%	0%	22%	14%	3%

Table 5. Most explicitly used Systems Thinking terminology in all databases.

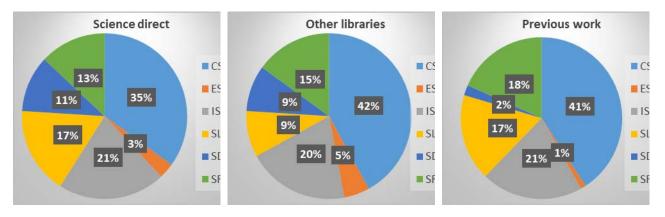


Fig. 8. Comparison of Industry Approach terminology usage between current work and the work of Error! Reference source not found..

Figure 8 depicts the consistency of our findings with previous research, while Figure 9 depicts the difference in the percentage of System Thinking usage in the current and previous works' terminology. Particularly in terms of Operations Research, Socio-technical Systems, and Complex Adaptive Systems. Further investigation is required to determine the source of such divergence.

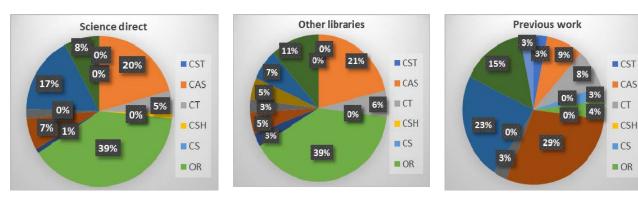


Fig. 9 Comparison of Systems Thinking methodology usage between current work and the work of [6].

6. Conclusion

Data mining is a process that extracts valuable patterns and relationships hidden in large amounts of data if the data is presented in a highly structured format. Text mining, a subset of data mining, converts unstructured data such as text documents into structured format [16].

We presented a text mining method in this paper to mimic the work of [6]. In that work, the classification of articles in terms of the methodology used for addressing social sustainability in supply chains was done manually on a large number of articles; it is a difficult and error-prone task.

The current method is a frequency-based process that employs industry terminology and systems thinking. The method employs text mining techniques to generate structured datasets from documents, with the values in the datasets consisting of terminology frequency. The presence of a data language term in an article indicates that the article addresses supply chain sustainability using the methodology of that term. This characteristic can be

p-ISSN: 2231-3850 Tayeb Basta et al. / Indian Journal of Computer Science and Engineering (IJCSE)

viewed as both a weakness and a strength of the current method. The former viewpoint is reflected in statements such as "operations research is not widely used in the field of..." operations research is a terminology used in systems thinking.

While the sentence confirms that operations research is not used in the article to address supply chain sustainability, the current method considers the inverse. To some extent, the latter viewpoint is manifested by ensuring that an article has not addressed supply chain sustainability using a methodology if its terminology is completely absent.

The method produces consistent results across two datasets: The Science Direct database and the other databases. The current method's results and those of [6] are mostly very close. However, there is some divergence where the percentage of certain systems thinking terms usage differs in both works.

The findings of this paper validate the findings of previous work, so academics and practitioners alike can confidently use the findings to enrich existing knowledge clusters or close the highlighted knowledge gaps, all while being aware of the various biases and trends that were discovered regarding what industry approach and systems thinking methodologies were used throughout the supply chain phase.

More work is needed to improve the current method. This could be accomplished in two ways. The first is to revise the data language, and the second is to use an information-extraction approach that inputs semantic information rather than keywords.

References

- Seuring, S.; Müller, M. (2008): From a literature review to a conceptual framework for sustainable supply chain management, Journal of Cleaner Production, 16(15), pp. 1699-1710
- Kleindorfer, P. R.; Saad, G. H. (2005): Managing disruption risk in supply chain Production and Operation Management, 14(1): pp.
- Linton, J. D.; Klassen, R.; Jayaraman, V. (2007): Sustainable supply chains: An introduction Journal of Operations Management, 25, pp. 1075-1082.
- [4] Kates, R. W.; Parris, T. M.; Leiserowitz, A. (2005): What is sustainable development? Goals, indicators, values, and practice Environment, 47(3): 9-21.
- Agbonkhese, S. E. (2010): Measuring environmental and social sustainability in the apparel supply chain, Master Thesis, Massachusetts Institute of Technology,
- Basta, M.; Lapalme, J.; Paquet, M. (2016): The Top Approaches for Managing Supply Chain Social Risk and Social Sustainability, 11th International Conference on Modeling, Optimization and Simulation - MOSIM16, Montréal, Québec, Canada, pp. 22-24.
- [7] Gupta, V.; Lehal, G. S. (2009): A Survey of Text Mining Techniques and Applications, Journal of Emerging Technologies in Web Intelligence, 1(1), pp. 60-76.
- Wei, C. H.; Kao, H. Y.; Lu, Z. (2013): PubTator: a web-based text mining tool for assisting biocuration, Nucleic Acids Res, 41Web Server issue, pp. W518-W522
- [9] Miner, G.; Elder, J. (2012). Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications, Elsevier.
- [10] Esichaikul, V.; Guha, S.; Juntapoln, C. (2011): Monitoring email transaction logs by text-mining email contents, 3rd International Conference on Data Mining and Intelligent Information Technology Applications ICMiA.
- [11] Li, J. (2014): Hierarchical Classification in Text Mining for Sentiment Analysis, International Conference on Soft Computing and Machine Intelligence ISCMI.
- [12] Netzer, O.; Feldman, R.; Goldenberg, J.; Fresko, M. (2012): Your Own Business: Market-Structure Surveillance Through Text Mining, Marketing Science, 31(3): 521–543
- [13] Inkpen, D. (2015): Text Mining in social media for Security Threats, Recent Advances in Computational Intelligence in Defense and Security, 621, pp. 491-517
- [14] Gupta R.; Gill, N. S. (2012): Financial Statement Fraud Detection using Text Mining, International Journal of Advanced Computer Science and Applications, 3(12), pp. 189-191.
- [15] Guthrie, D. (2006): A Closer Look at Skipgram Modelling, Proceedings of the 5th international Conference on Language Resources and Evaluation, pp. 1-4.
- [16] Han, J.; Kamber, M.; Pei, J. (2012). Data Mining: Concepts and Techniques, 3rd edition, Morgan Kaufmann.

Conflicts of Interest

The authors have no conflicts of interest to declare.

Authors Profile



Tayeb Basta, received his PhD from the computation department of the university of Manchester in UK. Currently, Tayeb teaches at the University of Science and Technology of Fujairah in the United Arab Emirates.



Mohamed Basta, Experienced IT Lead with a demonstrated history of working in the investment banking trade surveillance industry. Strong professional with a PhD. of Engineering focused on Supply Chain Management, Systems Thinking and Social Responsibility from École de Technologie Supérieure in Canada.