

2020

A Text-mining Approach to Evaluate the Importance of Information Systems Research Themes

Navid Aghakhani

University of Tennessee at Chattanooga, navid-aghakhani@utc.edu

Arben Asllani

University of Tennessee at Chattanooga, beni-asllani@utc.edu

Follow this and additional works at: <https://scholarworks.lib.csusb.edu/ciima>



Part of the [Management Information Systems Commons](#)

Recommended Citation

Aghakhani, Navid and Asllani, Arben (2020) "A Text-mining Approach to Evaluate the Importance of Information Systems Research Themes," *Communications of the IIMA*: Vol. 18 : Iss. 1 , Article 3.

Available at: <https://scholarworks.lib.csusb.edu/ciima/vol18/iss1/3>

This Article is brought to you for free and open access by CSUSB ScholarWorks. It has been accepted for inclusion in Communications of the IIMA by an authorized editor of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.

A Text-mining Approach to Evaluate the Importance of Information Systems Research Themes

Cover Page Footnote

We wish to submit the attached manuscript for the consideration for possible publication in the Communications of the IIMA. We attest that no portions of the manuscript have been copyrighted, published, accepted for publication, or are under review elsewhere. Furthermore, the manuscript has never been submitted to Communications of the IIMA, and no working papers, prior drafts, or other version of this manuscript are posted online.

A Text-Mining Approach to Evaluate the Importance of Information Systems Research Themes

***Abstract:** This study combines text-mining techniques and exploratory factor analysis to identify significant research themes in the information systems (IS) field during the last decade. We analyzed 483 papers that have been published in the Management Information Systems Quarterly (MISQ) journal since 2010. We found that the most published themes in the MISQ are: (1) the impact of information systems and digital innovation on organizational performance, (2) the influence of computer-mediated communication and social media within individuals and groups, (3) information systems development and project management, (4) design science, and (5) decision making and information Systems. We also perform econometric analysis to study the effect of these research themes on the number of downloads, citations and reads in Google Scholar (SC), ResearchGate (RG), and the Association for Information Systems (AIS) library. This study shows that “the impact of information systems and digital innovation on organizational performance” and “the influence of computer-mediated communication and social media within individuals and groups” are positively associated with the number of reads, downloads, and citations. However, “information systems development and project management” is negatively associated with the number of reads, downloads, and citations. The methodology used here can be implemented in other areas of research by journal editors and research institutions to identify the most sought-after areas of future research. Also, understanding the most relevant themes of IS research can help universities revise their IS doctorate curriculum to better meet the needs of today's academic workforce.*

Keywords: Text Mining, IS Research Themes, Social Media, IS Curriculum

1. Introduction

The study of research themes in any academic discipline is essential to investigating changes in these themes and the plethora of emerging challenges (Palvia, Kakhki, Ghoshal, Uppala, & Wang, 2015). Leading academic journals play an essential role in addressing this need. In response, premium business journals have published similar studies to inform scholars about how the prominent research themes in their discipline have evolved (Donthu, Gremler, Kumar, & Pattnaik, 2020; Lamberton & Stephen, 2016; Linnenluecke, Marrone, & Singh, 2020). Similar studies are also required in the Information Systems (IS) discipline due to the rapid changes in Information Technology and its corresponding societal impacts (Goyal, Ahuja, & Guan, 2018).

This study, particularly, is motivated by two important gaps in the prior research. First, previous studies that have investigated IS themes are somewhat limited. They either analyze a specific research area in IS or use traditional methodologies (e.g., content analysis) to examine IS themes (Hughes et al., 2019; Palvia et al., 2015; Senyo, Addae, & Boateng, 2018). Using content analysis as a research tool is a labor-intensive process, which makes it challenging to analyze a vast number of journal articles. As such, there is a need to use data-driven approaches to evaluate IS research themes (Goyal et al., 2018; Okoli, 2015). Second, the emergence of social media and social networking platforms has transformed the way individuals interact with one another (Aghakhani, Karimi, & Salehan, 2018). This transformation is also evident in the online community of practice, particularly scholar communities (Yu, Wu, Alhalabi, Kao, & Wu, 2016). Researchers usually rely on sharing their research output to gain peer recognition or funding opportunities (Mas-Bleda, Thelwall, Kousha, & Aguillo, 2014). The emergence of social platforms has enabled researchers to conveniently connect with their peers and collaborate and to promote their research. For instance, someone can promote his/her research by sharing a link to his/her

recently published article(s) on LinkedIn. ResearchGate is an example of a professional social networking service for scholars that facilitates peer recognition by allowing researchers to follow each other, share their projects, and receive feedback on their research.

Understanding the drivers of peer recognition and attention in online communities has been an important research topic in IS (Bhattacharyya, Banerjee, Bose, & Kankanhalli, 2020). For instance, in the context of online consumer reviews, studies use text mining and natural language processing techniques to examine how the content of online reviews can predict review readership or review helpfulness (Lah & Dahlan, 2017; Siering, Muntermann, & Rajagopalan, 2018). The same analogy can be used in the scholars' online social community, where citation counts or readership counts can signal the members' reputation and peer recognition. This paper aims to fill these gaps by addressing the following research questions:

- RQ1: How can IS research themes be evaluated using big data analytics?
- RQ2: What are the major IS research themes based on big data analytics?
- RQ3: What is the association between IS research themes and the number of reads, citations, and downloads?

To address these questions, we propose a data-driven method that combines text mining and econometrics models to identify the major IS research themes and examine their effect on citation counts, download counts, and readership counts in various platforms. Our sample includes 483 papers published in MISQ since 2010. For each article in our sample, we collected the citation counts from Google Scholar and ResearchGate. We also obtained download counts from the AIS library and readership counts from ResearchGate. This study contributes to the field by extending the notion of IS research themes to the context of attention and peer recognition in the online social platforms of scholars. This study also informs scholars about the major research themes in the

Information Systems (IS) discipline and their association with the number of citations, number of reads, and number of downloads in the Google Scholar, ResearchGate, and the AIS Library.

2.Literature

Any science's business is to develop new understandings of past, persisting, or newly identified natural phenomena (Jackson, 1996). In this quest, a series of meta-scientific methods are developed to provide insights into the main scientific themes and identify features of theories that predict their long-term survival (Faust & Meehl, 2002). Previous research suggests that theme analysis plays an essential role in forming perceptions (Linell, 2010) and influencing how decisions are made (Putnam & Cooren, 2004). There is evidence of using theme analysis outside the realm of scientific research. For example, organizational scholars have demonstrated that theme analysis can help them understand drivers of corporate sustainability (Sodhi & Tang, 2018). It is also used to describe mergers and acquisitions (Christofi, Leonidou, & Vrontis, 2017), corporate restructuring (Sonenshein, 2010), organizational innovations (Bartel & Garud, 2009), and strategy formulations (Barry & Elmes, 1997). Several other studies use annual reports (Subramanian, Insley, & Blackwell, 1993), shareholder letters (Jameson, 2000), press releases (Henry, 2008), and corporate websites (Pollach, 2003) as the object of a theme analysis.

Research theme analysis has also been an essential topic in IS research and education, mainly due to the rapid technological advances in Information Technology (e.g., the emergence of social media or the shift toward AI and big data analytics) and its corresponding impact on individuals and organizations. Broadly, theme analysis in IS can be classified into two major streams in terms of adopted methodologies. First, early studies have used literature review and meta-analysis methods to examine the research themes in a specific domain of IS or to discover primary IS research from papers published in major IS journals. For instance, Nevo and

Kotlarsky (2020) have analyzed the major research published on crowdsourcing in information systems literature. Similar studies also have been done in other IS fields, such as analyzing the effect of Information Technology on productivity (Appiahene, Ussiph, & Missah, 2018) and evaluating health information systems research (Haried, Claybaugh, & Dai, 2019). Other studies in this line of work have examined the foci of IS research and have discovered the core research themes. For instance, Palvia et al. (2015) have conducted a meta-analysis on seven major IS journals and found that electronic commerce is a significant research theme and survey analysis is the dominant research methodology. In a similar vein, Mazaheri, Lagzian, and Hemmat (2020) have reported that electronic commerce/business, information systems research, and IS usage/adoption are the dominant research topics in IS.

The second stream of research has adopted objective methodologies such as quantitative content analysis and text mining to address the lack of accuracy and labor-intensiveness of the literature review method (Goyal et al., 2018; Okoli, 2015; Palvia et al., 2015). Weigel, Rainer Jr, Hazen, Cegielski, and Ford (2013) have used a quantitative content analysis method to explore 2,188 papers published over a ten-year period in medical informatics, information systems, as well as medical journals and discovered significant themes and sub-themes in this domain. Sidorova, Evangelopoulos, Valacich, and Ramakrishnan (2008) have used Latent Semantic Analysis (LSA) on the abstracts of research papers published in MISQ, ISR, and JMIS over a ten-year period and identified five core research themes. Similarly, Goyal et al. (2018) have used the LSA method on the abstracts of papers published in MISQ, ISR, JMIS, and JAIS to identify the emergent research themes in IS and the specific research theme that a particular journal desires most.

Our research intends to contribute to these prior studies in two important ways. First, the literature suggests that text mining has become a prominent method to objectively identify IS research themes. While previous text mining studies have mainly used the abstracts of the papers as the main text corpus in their analysis, this study uses the entire journal text. Also, our study proposes a novel analytical framework by integrating a modified version of MapReduce with Exploratory Factor Analysis to identify essential keywords that will be used later to build research theme dictionaries using the Linguistic Inquiry and Word Count (LIWC) program. Second, our research will extend the notion of theme analysis to peer recognition in scholars' online social communities. We will notably examine how the research themes are associated with citation counts, download counts, and readership on Google Scholar, the AIS Library, and ResearchGate platforms.

3. Methodology

Text mining is a variety of techniques used to discover patterns in a collection of text documents (Feldman & Sanger, 2007). Text mining has been used successfully in theme analysis to identify patterns and quantify emerging keywords, providing insight into the structure, context, and trends of themes (Galati & Bigliardi, 2019; Nie & Sun, 2017; Shin et al., 2018). When combined with Big Data platforms and techniques, text mining becomes a powerful approach in theme analysis. Specifically, the methodology proposed in this study encompasses the following steps.

Step 1: Data Collection

The analysis source is the full text of the 482 articles published in MISQ since 2010. The articles are downloaded from the AIS Library. The full text and keywords for articles are organized into folders. For longitudinal studies, the information is classified into subfolders

according to the publication year. Also, for every article in our corpus, we have collected the citation counts from Google Scholar, download counts from AIS library, and citation counts and readership counts from ResearchGate.

Step 2: Apply MapReduce to count keyword frequencies

At this step, a modified version of the MapReduce program is used to count the frequencies for each keyword. The traditional MapReduce algorithm (Dean & Ghemawat, 2008) is modified to measure the frequency of words found in a full-text corpus, using the authors of the published articles' keyword list.

Step 3: Use the Pareto principle to reduce the number of keywords

The exploratory factor analysis that follows requires a small number of measurable variables or keywords. Following the Pareto principle (Arnold, 2015), only the most frequently used keywords (e.g., the top 20 percent) are used in the analysis. The 80/20 principle is only a rule of thumb and may not apply in all circumstances, although an approximate application of the rule is recommended.

Step 4: Perform exploratory factor analysis

Each file represents the data points for the factor analysis (e.g., journal article). Once the most frequent keywords are identified, the frequency index f_{ij} of each keyword i in article j is calculated as follows:

$$f_{ij} = \frac{F_{ij}}{T_j}$$

where F_{ij} is the frequency of the keyword i in article j , and T_j is the total number of words in article j . When applying factor analysis in this methodology, it is essential to make sure that the sample size is adequate, i.e., the number of text files is significantly large. One must also ensure that at least three variables are associated with each of the expected factor categories or themes.

A theme is determined by the similarity between the underlying variables (keywords). However, it is suggested that the total number of variables used to identify themes should be no more than one-third of the sample size. Like other cases in factor analysis, Bartlett's test must be significant, where a small value of significance ($p < 0.05$) indicates that factor analysis produces useful results from the data (Snedecor & Cochran, 1989).

Table 1 shows the output of the factor analysis. It shows the loading coefficients for each keyword into major factors or themes. We omitted loading values less than 0.35 for clarity of representation. As shown, the factor analysis yields five factors based on the top frequent keywords from Step 4. These factors and their corresponding keywords represent the major themes in our data. The authors have labeled these factors based on their similar research themes in IS as follows: Factor (1) The Impact of Information Systems and Digital Innovation on Organizational Performance; Factor (2) The Influence of Computer-Mediated Communication and Social Media within Individuals and Groups; Factor (3) Information Systems Development and Project Management; Factor (4) Design Science; and Factor (5) Decision Making and Information Systems.

Table 1: Loading of Keywords into Factors

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
business	0.68				
strategic	0.607				
management	0.572				
strategy	0.558				
organization	0.488				
innovation	0.411				
digital	0.395				
individual		0.568			
social		0.504			
models		0.447			
computer		0.425			
communication		0.411			
technology		0.406			
group		0.368			
influence		0.364			
development			0.695		
software			0.666		
project			0.656		
team			0.556		
offshore			0.509		
design				0.636	
science				0.591	
problem				0.569	
field				0.472	
theories				0.453	
decision					0.597
making					0.528
task					0.398
performance					0.351

Step 5: Measure theme presence for a specific period

For this part of the methodology, it is recommended that the researcher use the LIWC tool. LIWC is a text analysis program that counts words in psychologically meaningful categories. LIWC has the “ability to detect meaning in a wide variety of experimental settings” (Tausczik & Pennebaker, 2010). The LIWC software program allows a researcher to create a unique dictionary to investigate a specific category or a set of categories in a text file or series of

text documents. The themes identified in Step 4 and the keywords associated with these themes are used as the basis for a custom dictionary file. This dictionary is uploaded in the LIWC program, and the full text of articles of any given year are processed. Table 2 shows the first few records of the output of the LIWC program. As shown, LIWC calculates the percentage of the words that belong to a given theme for each paper. For example, in the first paper, "A Trigger ...", 0.52 percent of the words in the article belong to the "innovation performance" theme and 2.33 percent belong to the "communication and social media" theme.

Table2: Output Sample of LIWC

1		Theme 1	Theme 2	Theme 3	Theme 4	Theme 5
2	Year and Paper Title	Innovation performance	Communication and Social Media	Project Management	Design Science	Decision Making
3	2010 A_Trigger_Model_for_Technology_Adaptation_Interventions.txt	0.52	2.33	2.07	0.66	0.55
4	2010 Circuits_of_Power.txt	0.91	1.07	0.40	0.50	0.40
5	2010 User_Participation_in_IS_Security_Risk_Management.txt	1.83	0.76	0.64	0.36	0.67
6	2010 Investigating_Contradictory_Views_of_Formative_Measurement.txt	0.38	2.69	0.40	0.43	0.62
7	2010 Studying_Technology_Adoption_by_Groups.txt	0.30	5.28	0.39	0.25	1.05
8	2010 Information_Systems_Strategy.txt	5.79	0.64	0.38	0.27	0.38
9	2010 Effects_of_IT_Fashion_on_Organizations.txt	1.19	0.79	0.21	0.44	1.02
10	2010 Brand_Positioning_Strategy.txt	0.38	0.70	0.10	0.26	0.43
11	2010 Impact_of_IT_and_Transactive_Memory_Systems.txt	1.28	1.63	4.41	0.63	0.97
12	2010 IS_and_Environmentally_Sustainable_Development.txt	0.56	1.07	0.33	0.64	0.66
13	2010 Neural_Gender_Differences_in_Online_Trust.txt	0.16	1.18	0.13	0.21	0.62
14	2010 Market_Value_of_Voluntary_Disclosures.txt	0.26	0.75	0.04	0.14	0.07
15	2010 Consumer_Reviews_on_Amazon.com.txt	0.38	0.75	0.23	0.19	0.52
16	2010 Trust_and_Distrust.txt	0.12	0.83	0.09	0.26	0.26
17	2010 Impact_of_Information_Capabilities_Design_on_BPO_Performance	1.05	1.03	0.96	1.05	1.00
18	2010 Malicious_Agents_and_Enterprise_Software.txt	0.23	0.52	0.85	0.21	0.21
19	2010 Employee_IS_Security_Policy_Violations.txt	0.23	1.14	0.12	0.39	0.17
20	2010 Information_About_Information.txt	0.27	1.18	0.54	0.60	0.78
21	2010 Formation_and_Value_of_IT_Enabled_Resources.txt	1.55	0.83	0.30	0.61	0.86
22	2010 Organization_Enterprise_System_Fit.txt	0.67	1.24	0.56	0.27	0.74
23	2010 The_Willoughby_Monograph_Series.txt	1.31	0.75	0.56	0.75	0.19
24	2010 Direct_&_Indirect_Effects_of_Emotions_on_IT_Use.txt	0.17	1.58	0.14	0.21	0.94
25	2010 Web_2.0_and_Politics.txt	0.45	2.37	0.28	0.36	0.26

Step 6: Econometric analysis

In this step, we use the output of LIWC to examine the association between the research themes and the corresponding citation counts, download counts, and readership counts on Google Scholar, the AIS library, and ResearchGate. We use negative binomial regression, one of the Poisson model variations (Greene, 1994; Schindler & Bickart, 2012), to examine each theme's effect on the metrics mentioned above. To ensure that negative binomial regression was

suitable over the Poisson model, we tested whether the overdispersion parameter α was significantly different from zero. Our results gave a p-value < 0.001 , which confirms the existence of overdispersion in our dataset, validating that it is appropriate to use negative binomial regression over the Poisson model. Table 3 shows the results of the negative binomial regression.

Table 3: Negative Binomial Regression

	Google Scholar Citation	ResearchGate Read	ResearchGate Citation	AIS Download
Theme 1	.23*** (.0608161)	.41*** (.0845174)	.27*** (.0617112)	.18*** (.0565704)
Theme 2	.26*** (.0580593)	.24** (.0859941)	.26*** (.059043)	.12*** (.0547912)
Theme 3	-.12** (.0520033)	-.21*** (.0760049)	-.13** (.0544821)	-.10** (.0500155)
Theme 4	.33*** (.1013019)	.32** (.1520257)	.41*** (.1029127)	.04 (.0874751)
Theme 5	-.32*** (.1001028)	.11 (.1631319)	-.32*** (.1015707)	-.06 (.101219)
Year	.35 *** (.017909)	.13 *** (.0248495)	.36*** (.017818)	.13*** (.0176224)
<p>Note 1: Theme 1: The Impact of Information Systems and Digital Innovation on Organizational Performance Theme 2: The Influence of Computer-Mediated Communication and Social Media within Individuals and Groups Theme 3: Information Systems Development and Project Management Theme 4: Design Science Theme 5: Decision Making and Information Systems</p> <p>Note 2: *p<0.05; **p<0.01; ***p<0.001</p> <p>Note 3: Numbers in parenthesis are standard errors</p> <p>Note 4: Year represents the increasing time since publication</p>				

4. Discussion

As shown in Table 3, our results show a consistent relationship between three research themes and the download counts, citation counts, and readership counts on the three platforms.

In our analysis, we control for the effect of years elapsed from the publication date to the date the

data was collected. Therefore, as our results show, the greater the time since publication, the greater the likelihood of receiving more citations, downloads, and reads. ResearchGate represents a multidisciplinary forum for scholars to share their research output with their peers. Thus, the number of reads in ResearchGate represents the interdisciplinary attention an article receives. AIS, however, is an electronic library for Information Systems professionals. Therefore, the number of reads represents the attention an article gets among Information Systems professionals.

Our analysis shows that research theme 1 (the impact of information systems and digital innovation on organizational performance) and research theme 2 (the influence of computer-mediated communication and social media within individuals and groups) have a positive association with the citation counts in both Google Scholar and ResearchGate, as well as scholarly attention in both AIS and ResearchGate. This finding is probably related to the notion that the individual and organization-level impacts of information systems have been the focal research themes in the IS discipline since its inception. Simultaneously, as information systems have evolved and business processes and business models have become digitized, the role of information systems in other disciplines such as organization science or marketing has become more prevalent.

Research theme 3 (information systems development and project management), has a negative association with all the dependent variables (citation counts, readership counts, and download counts). The negative correlation might be related to how this research theme and its related theories have matured over time. Most of the prior studies have focused on developing and managing enterprise systems (e.g., ERP systems). With the emergence of AI-based systems, however, more research needs to be done on the management and design of such systems,

shifting the focus away from the majority of articles published on enterprise systems that fall under theme 3 and leading to a negative correlation overall (Berente, Gu, Recker, & Santhanam, 2019).

Our results also indicate that while theme 4, design science, has a positive association with citation counts in Google Scholar as well as citation counts and readership counts in ResearchGate, it has no significant association with download counts in the AIS library. We submit one explanation for this result. It is reasonable to argue that the Information Systems discipline had adopted the design science method from computer science as one of its reference disciplines. However, over the past decade, behavioral research and the corresponding methodologies for data analysis (e.g., qualitative and survey data analysis) have been the major focus in IS journals, as well as designing Ph.D. courses and seminars. It is plausible to argue that with the emergence of big data analytics and artificial intelligence, the role of design science in IS research will become more prominent, which implies the importance of the change in IS curriculum and training scholars in response to this change.

Finally, the negative or not significant association between theme 5 and citation counts as well as readership counts might be an indicator of the maturity of research on the traditional decision support systems. The majority of research on decision making has focused on the adoption of conventional decision support systems by individuals, groups, or organizations. The emergence of data-driven and AI-enabled decision support systems, however, requires developing new theories to understand challenges in adopting and managing these systems (Berente et al., 2019). Therefore, we expect that conducting studies on theoretical aspects of adoption and managing AI-enabled decision support systems will result in positive trends in readership and citation counts of such studies.

5. Conclusion

This study contributes to the literature by proposing a data-driven method to examine IS research themes. Specifically, we go beyond prior studies by using the full text of articles in our corpus and combining classic statistics with a wordcount algorithm and linguistic inquiry tools. Also, we contribute to the attention and peer recognition literature by introducing the notion of theme analysis and examining the effect of research themes on various important peer recognition metrics in the online social media of academics. This study also informs scholars, publishers, academics, and social media platforms about the specific research themes that would receive more citations or more recognition on particular platforms. Finally, universities may use these findings in designing or revising their IS doctorate curriculum in response to the IS academic market demand and emerging research.

Like other research, this study comes with limitations. First, we used cross-sectional data to examine the effect of research themes on the dependent variables. Therefore, our results are not immune from unobserved heterogeneity associated with the cross-sectional data. Second, our data is limited to MISQ papers, which limits the generalizability of our findings. Thus, we suggest that future research use panel data and incorporate multiple journal outlets to investigate the causal effects of research themes and increase the findings' generalizability.

References and Citations

- Aghakhani, N., Karimi, J., & Salehan, M. (2018). A unified model for the adoption of electronic word of mouth on social network sites: Facebook as the exemplar. *International Journal of Electronic Commerce*, 22(2), 202-231.
- Appiahene, P., Ussiph, N., & Missah, Y. M. (2018). Information technology impact on productivity: a systematic review and meta-analysis of the literature. *International Journal*

- of Information Communication Technologies and Human Development (IJICTHD), 10(3), 39-61.
- Barry, D., & Elmes, M. (1997). Strategy retold: Toward a narrative view of strategic discourse. *Academy of Management Review*, 22(2), 429-452.
- Bartel, C. A., & Garud, R. (2009). The role of narratives in sustaining organizational innovation. *Organization Science*, 20(1), 107-117.
- Berente, N., Gu, B., Recker, J., & Santhanam, R. (2019). Managing AI. Call for papers MISQ special issue on managing AI. *MIS Quarterly*, 1-5.
- Bhattacharyya, S., Banerjee, S., Bose, I., & Kankanhalli, A. (2020). Temporal effects of repeated recognition and lack of recognition on online community contributions. *Journal of Management Information Systems*, 37(2), 536-562.
- Christofi, M., Leonidou, E., & Vrontis, D. (2017). Marketing research on mergers and acquisitions: a systematic review and future directions. *International Marketing Review*, 34(5), 629-651.
- Dean, J., & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107-113.
- Donthu, N., Gremler, D. D., Kumar, S., & Pattnaik, D. (2020). Mapping of Journal of Service Research Themes: A 22-Year Review. *Journal of Service Research*, December 2020. doi:10.1177/1094670520977672.
- Faust, D., & Meehl, P. E. (2002). Using meta-scientific studies to clarify or resolve questions in the philosophy and history of science. *Philosophy of Science*, 69, S185-S196.
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*: Cambridge University Press.
- Galati, F., & Bigliardi, B. (2019). Industry 4.0: Emerging themes and future research avenues using a text mining approach. *Computers in Industry*, 109, 100-113.
- Goyal, S., Ahuja, M., & Guan, J. (2018). Information Systems Research Themes: A Seventeen-year Data-driven Temporal Analysis. *Communications of the Association for Information Systems*, 43(1), 23.
- Greene, W. H. (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. Leonard N. Stern School of Business, New York University. Retrieved January 2nd, 2021 from <https://archive.nyu.edu/bitstream/2451/26263/2/94-10.pdf>.
- Haried, P., Claybaugh, C., & Dai, H. (2019). Evaluation of health information systems research in information systems research: A meta-analysis. *Health Informatics Journal*, 25(1), 186-202.
- Henry, E. (2008). Are investors influenced by how earnings press releases are written? *The Journal of Business Communication* (1973), 45(4), 363-407.
- Hughes, L., Dwivedi, Y. K., Misra, S. K., Rana, N. P., Raghavan, V., & Akella, V. (2019). Blockchain research, practice and policy: Applications, benefits, limitations, emerging research themes and research agenda. *International Journal of Information Management*, 49, 114-129.
- Jackson, E. A. (1996). The second metamorphosis of science: A second view. Santa Fe Institute Foundational Papers. Retrieved January 2nd, 2021 from <https://bit.ly/38ZnMG5>
- Lah, N. S. B. C., & Dahlan, H. M. (2017). A concept-level approach in analyzing review readership for E-Commerce persuasive recommendation. Paper presented at the 2017 International Conference on Research and Innovation in Information Systems (ICRIIS).

- Lamberton, C., & Stephen, A. T. (2016). A thematic exploration of digital, social media, and mobile marketing: Research evolution from 2000 to 2015 and an agenda for future inquiry. *Journal of Marketing*, 80(6), 146-172.
- Linell, P. (2010). Communicative activity types as organisations in discourses and discourses in organisations. *Discourses in Interaction* (203), 33.
- Linnenluecke, M. K., Marrone, M., & Singh, A. K. (2020). Sixty years of Accounting & Finance: a bibliometric analysis of major research themes and contributions. *Accounting & Finance*, 60(4), 3217-3261.
- Mas-Bleda, A., Thelwall, M., Kousha, K., & Aguillo, I. F. (2014). Do highly cited researchers successfully use the social web? *Scientometrics*, 101(1), 337-356.
- Mazaheri, E., Lagzian, M., & Hemmat, Z. (2020). Research directions in information systems field, current status and future trends: A literature analysis of AIS basket of top journals. *Australasian Journal of Information Systems*, 24, <https://doi.org/10.3127/ajis.v24i0.2045>.
- Nevo, D., & Kotlarsky, J. (2020). Crowdsourcing as a strategic is sourcing phenomenon: Critical review and insights for future research. *Journal of Strategic Information Systems*, 29(4), 101593.
- Nie, B., & Sun, S. (2017). Using text mining techniques to identify research trends: A case study of design research. *Applied Sciences*, 7(4), 401.
- Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, 37, 879-910.
- Palvia, P., Kakhki, M. D., Ghoshal, T., Uppala, V., & Wang, W. (2015). Methodological and topic trends in information systems research: A meta-analysis of IS journals. *CAIS*, 37, 30.
- Pollach, I. (2003). Communicating corporate ethics on the world wide web: a discourse analysis of selected company web sites. *Business & Society*, 42(2), 277-287.
- Putnam, L. L., & Cooren, F. (2004). Alternative perspectives on the role of text and agency in constituting organizations. *Organization*, 11(3), 323-333.
- Schindler, R. M., & Bickart, B. (2012). Perceived helpfulness of online consumer reviews: The role of message content and style. *Journal of Consumer Behaviour*, 11(3), 234-243.
- Senyo, P. K., Addae, E., & Boateng, R. (2018). Cloud computing research: A review of research themes, frameworks, methods and future research directions. *International Journal of Information Management*, 38(1), 128-139.
- Shin, S.-H., Kwon, O. K., Ruan, X., Chhetri, P., Lee, P. T.-W., & Shahparvari, S. (2018). Analyzing sustainability literature in maritime studies with text mining. *Sustainability*, 10(10), 3522.
- Sidorova, A., Evangelopoulos, N., Valacich, J. S., & Ramakrishnan, T. (2008). Uncovering the intellectual core of the information systems discipline. *MIS Quarterly*, 467-482.
- Siering, M., Muntermann, J., & Rajagopalan, B. (2018). Explaining and predicting online review helpfulness: The role of content and reviewer-related signals. *Decision Support Systems*, 108, 1-12.
- Snedecor, G. W., & Cochran, W. G. (1989). *Statistical Methods*, eight edition. Iowa State University Press, Ames, Iowa.
- Sodhi, M. S., & Tang, C. S. (2018). Corporate social sustainability in supply chains: a thematic analysis of the literature. *International Journal of Production Research*, 56(1-2), 882-901.
- Sonenshein, S. (2010). We're changing—Or are we? Untangling the role of progressive, regressive, and stability narratives during strategic change implementation. *Academy of Management Journal*, 53(3), 477-512.

- Subramanian, R., Insley, R. G., & Blackwell, R. D. (1993). Performance and readability: A comparison of annual reports of profitable and unprofitable corporations. *The Journal of Business Communication* (1973), 30(1), 49-61.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24-54.
- Weigel, F. K., Rainer Jr, R. K., Hazen, B. T., Cegielski, C. G., & Ford, F. N. (2013). Uncovering research opportunities in the medical informatics field: A quantitative content analysis. *CAIS*, 33, 2.
- Yu, M.-C., Wu, Y.-C. J., Alhalabi, W., Kao, H.-Y., & Wu, W.-H. (2016). ResearchGate: An effective altmetric indicator for active researchers? *Computers in Human Behavior*, 55, 1001-1006.