The Role of Information Organization and Knowledge Structuring in Combatting Misinformation: A Literary Analysis

Kevin Matthe Caramancion(✉)

University at Albany, State University of New York, Albany, NY 12222, USA
kcaramancion@albany.edu

Abstract. This paper seeks to explore how the three hallmark dimensions of Information Organization—(1) Access, (2) Discovery, and (3) Retrieval—each as a construct paves the way for the rise of misinformation and, consequently, be the essential areas of interest for its control and regulations. Furthermore, the role of social networking platforms, grounded on *folksonomy-designed* environments, is examined on how they function as a seminal contributing factor for the creation and persistence of misinformation. A *taxonomic* approach as an addendum for the remedy of misinformation is assessed, including suggestions for its more robust implementation. This paper concludes with a summative precis of the presented ideas and literature on this subject and stipulates the general limitations of Information Organization and Knowledge Structuring in their applications in the domain of misinformation.

Keywords: Taxonomy · Folksonomy · Misinformation

1 Introduction

Misinformation or *fake news*, albeit its deceptive nature, is still a type of information. Thus, it is still subject to the characteristics and properties of classical correct and truthful information. These include the metadata functioning as its descriptors, ontological clues dictating its domain, and even the semantical provenance that can be used to trace its origins—among other qualities.

Knowledge structure, the interrelated collection of facts or knowledge about a particular topic, is grounded on labels [1] and relations [2]. The Data-Information-Knowledge-Wisdom (DIKW) pyramid of Information Management suggests the processing and transformations that transpire in many models, including information systems and the cognitive behavior of humans. An erroneous value in any of the stages of the DIKW can be a highly probable cause of misinformation. Incorrect data, when not corrected, can lead to inaccurate analysis. Correct data, coupled with incorrect analysis, may yield inaccurate information. Correct data and analysis that leads to verified information may still be rejected by users when their cognitive capabilities can't comprehend or outright reject it due to biases or incompatibility in information-seeking motivation (i.e., impairment of knowledge and wisdom).

D. Mohaisen and R. Jin (Eds.): CSoNet 2021, LNCS 13116, pp. 319–329, 2021.
https://doi.org/10.1007/978-3-030-91434-9_28

The figure [3] below illustrates the connected and interacting elements that, in their trivial ways, when compromised, may contribute to misinformation which is represented by the blue oval as the societal challenge in this paper's context (Fig. 1).

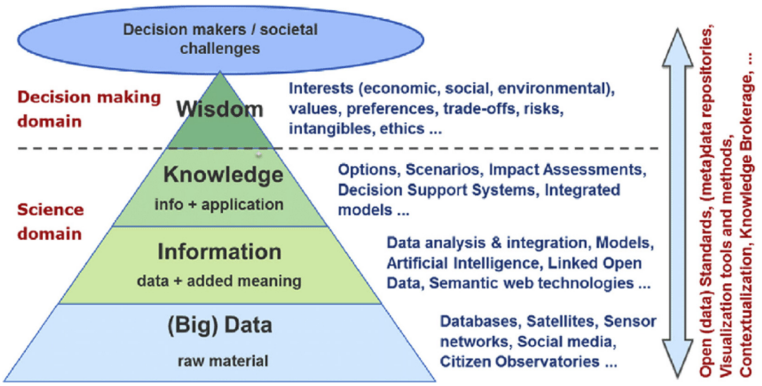

Fig. 1. The DIKW ecosystem & its elements

Given these groundings, falsehoods and misinformation is effectively controlled when prevented in the first place rather than the use of corrective mechanisms. However, the use of the latter should always be available as an essential recourse for its governance in information environments since the infallibility of the integrity in all the content remains elusive.

2 Information Access as a Factor

2.1 Access as a Definitional Construct

In the context of knowledge inquiry, *information access* is the series of actions performed by users to achieve the goals of seeking, organizing, and understanding the phenomena of any particular information [4]. Moreover, information access is the freedom or ability to identify, obtain and make use of any database or information effectively since users typically interact with information technologies for this undertaking. Information access, as an abstraction, takes in the concrete form of search queries, text summarization, and text clustering, among other representations [5]. Information access, as a dimension, covers several key issues prevailing in the current times, including but not limited to copyright issues, open-source, privacy, and even security.

One primal importance of information access is the objective to which it seeks to simplify and make it more effective for human users to access and further process large and unwieldy amounts of data and information. Information consumers engaged in the information-seeking process has one or more goals in their minds and use the search systems as tools to help achieve those goals. The fields of user-centered design and human-computer interaction (HCI) are the prevailing subfields of computing that support bridging the gap between the information-seeking motivation of users and the

interface of information technologies. To fulfill the information needs of users, accurate knowledge translations (e.g., from tacitly abstract to more explicit codified forms) and representations between these two entities is imperative.

The figure [6] below shows how this consistent interaction and their respective triggers take place (Fig. 2).

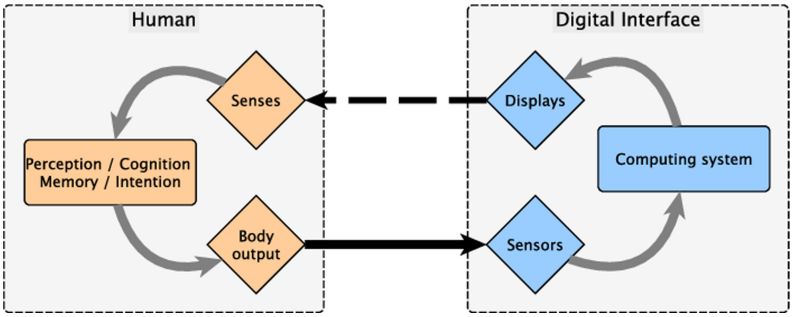

Fig. 2. The human-computer interaction

2.2 Applications on Misinformation

Social Networking Platforms, which have been historically used for entertainment purposes by users [7], have now superseded that affordance. These platforms are now the most used medium for information access, including news consumption which offers more timely, real-time updates from public figures and news channels. These platforms, functioning as containers, now allows webpages to be contained in their respective environments allowing both structured and unstructured data. The very backbone of knowledge inquiry and structuring has been maintained arduously—from queries, keyword searches, even up to reverse audio or image lookup. The flexibility of even the most abstract form of queries to be looked up has been made incrementally possible.

The rise of these platforms mainly altered two crucial features of the knowledge inquiry process. First is the transfer of inceptions in the information-seeking goals of the users to the news feed. Unlike traditional queries where users have an objective prior to their information access, users now typically access the medium first and, from there, decide which information to consume. Next, unlike the traditional information-seeking mediums where resources and artifacts generally are scrutinized and substantiated, social networking websites such as Facebook and Twitter allow almost any user to create and share contents that can be freely accessed, information-wise, by other users [8]. These are the two predominant characteristics of the current platforms that pave the way for the creation and persistence of misinformation. The developments in access technologies, in ways, made it easier for misinformation to reach an even bigger and wider audience.

3 Information Discovery as an Amplifier

3.1 Discovery as a Definitional Construct

In the context of the information search process, *information discovery* is the complex series of tasks that involves locating a particular digital object on the network through iterative research activities, which usually involve the specification of a set of criteria relevant to the resources needs of users, the organization, and ranking resources in the candidate in this candidate search, and the repeated expansion or restriction based on the characteristics of the identified resources and exploration of specific resources [9]. Information and knowledge discovery is grounded on the principles of structures and semantic relationships where related objects can be grouped based on metadata values for classificatory purposes [10]. Information discovery, as a dimension, entails controlled vocabularies, thesauri, and other related forms of taxonomic structures.

In outright retrospect, the paramount importance of information discovery is the uncovering of relevant information based on a user's information needs and seeking objectives. Information systems that seek to predict the relevance or preference to a particular resource are called recommender or recommendation systems.

The figure [11] below outlines the explicit and implicit data points used by recommender systems to attain this objective (Fig. 3).

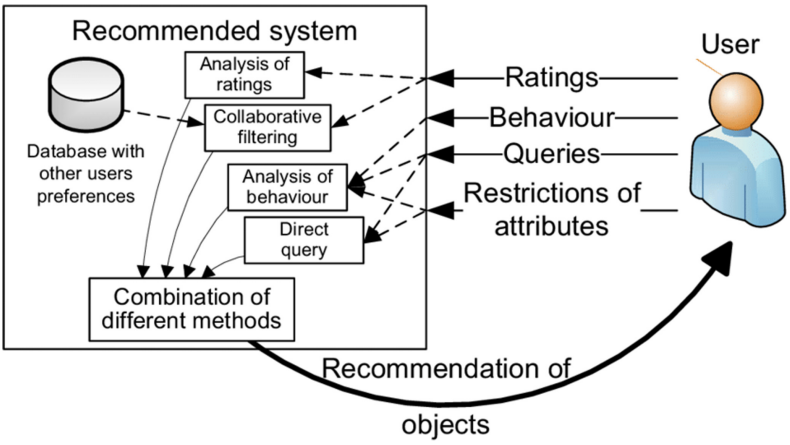

Fig. 3. A summative diagram of modern recommender systems

Information discovery is further attributed as the dimension responsible for filtering irrelevant information in the search process due to information overload [12]. Although it may seem uncomplicated, this task is especially arduous due to the vast number of resources, artifacts, and information that may be accessible to the users, especially in the digital age. Another crucial role of information discovery in the search process is the presentation of related resources that the users may find valuable and relevant for their information-seeking needs. This is due to the fact that the information search process is not static and, as a matter of truth, interactive, where learning and discovery are both unequivocal components of the process.

3.2 Applications on Misinformation

There has never been a failure in the transference of the principles of information discovery from its traditional mediums and formats to the contemporary social networking platforms. As a matter of truth, the components and enforcements of knowledge discovery did not just transfer, but rather, is even more amplified to reach more than its intended potential. This is particularly visible in the emergence of recommender systems in the queries of users, which have been so advanced that issues of privacy have been raised in the data collection methods used in these current technologies. The upgrades in mining and analytic techniques have been ubiquitous [13]. Unlike before, where usually structured data, metadata, and textual information are processed and mined, sophisticated advances now allow unstructured data not just processed but even manipulated. Computer vision as a field, for instance, offers complete analysis and control of images and videos alike [14].

The methods of information discovery transgressed the traditional forms of structured and unstructured data and included an analysis and mining of the behavior of the users. Included herein but is not limited to are mouse clicks, eye movement tracks, and even webpage refresh rates for the purposes of recommending more relevant information and content to the users in the form of extreme personalization, which are originally grounded on the fundamentals of information discovery. Fake news literature predominantly suggests that this model is a chief enabler of misinformation [15–17]. When users tend to acquire content to feed off their confirmation biases, even conspiracies, it creates an environment, typically referred to as a filter bubble, that is dividedly partisan and is usually a potent breeding ground of misinformation.

4 Information Retrieval as the Filter of Fake News

4.1 Retrieval as a Definitional Construct

In the information search process, *retrieval* is the process, methods, and procedures of acquiring particular information from resources relevant to users' information needs expressed through search queries. The technological underpinning that made this possible are two-fold: (a) the indexing system and (b) the query system, functioning as the interface for users. Results displayed are from the indexing system and can either be based on full-text or other specific-text extraction techniques [18]. The widely accepted variables functioning as metrics that measure a retrieval system's effectiveness are *precision and recall*. The former is the fraction of retrieved documents that are relevant to a particular query, while the latter is the fraction of the relevant documents that are successfully retrieved.

For its valuation in knowledge structuring, information retrieval compels search systems (e.g., engines, directories, etc.) to facilitate a rapid and accurate search-indexed structure based on the input(s) of users, usually in the form of keywords, to fulfill their interests and informational needs. Information retrieval, as a science, is in a continuous iterative process to keep improving this indexing design to provide an even more effective and rich search that includes texts from documents, metadata that describe data, and even unstructured data objects such as images, videos, and audio. Fields of computing are born

out of information retrieval such as natural language processing (NLP), decision trees (DT), and social search, among others. Due to these emergences, the queries, previously in textual format, can now be in the form of voice [19], reverse image, and even audio footprint—all for recognition purposes.

The figure [20] below displays the differences in indexing both structured and unstructured data objects (Fig. 4).

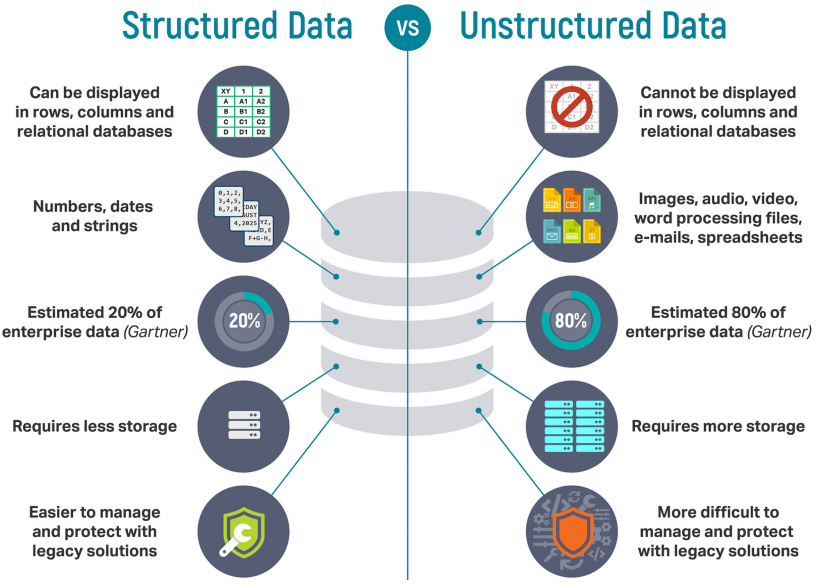

Fig. 4. Indexing style of data objects dependent on type

4.2 Applications on Misinformation

Non-factual information should cease to exist by the very doctrines of information retrieval. One of the hopes of *precision and recall* is to filter out irrelevant content that will not fulfill the information needs of the user performing the query. In the current information environments, however, the actors of misinformation have played the card of relevance all too well. Deceptive contents are indexed as appropriate since the descriptors are designed to truly mislead the users in their quest for information. Even unstructured data objects are not left untouched; manipulated photos and videos that support the claims of fake news are used to mislead the information user further. The paradox is that the quest for speed in retrieving content based on the queries of users will now often result in the inclusion of content that may be unverified.

Although retrieval systems may have been exploited to disseminate misinformation, the remedy still lies in this domain. Ranking strategies in indexing systems should promote verified, truthful content and demote the misleading ones. A strategy already enforced in retrieval systems where the computation of a numeric score on how well

each object in the database matches the query and ranks the objects according to this value. The top-ranking objects are then shown to the users. This process may then be iterated if the user wishes to refine the query since the outright removal of contents (i.e., censorship), as of this time, is a highly controversial topic since it intersects with the laws in freedom of speech and press [21]. The ranking strategy, however, is not without its controversies. Issues of discrimination, biases from paid content, privacy intrusion, and even racism are sometimes associated with this area of strategy.

5 The Contribution of Folksonomy in Misinformation Explosion

5.1 Influence in Social Networking Platforms

At the very backbone of social networking platforms lies the democratic model of user-generated content (UGC). These are any forms of content that are contributed by any user in an information environment, be it of any type, including videos and images. The question of whether these environments are anchored from folksonomic or taxonomic structures is visibly evident and has long been settled. The knowledge acquisition obtained through social tagging in almost any content has been long the classical strength of social mediums over the traditional top-down approach to information sharing [22]. The community-based UGCs and tagging lacked the traditional hierarchies of taxonomies, and for a long time, this has been a feature of strength rather than a vulnerability. This bypassed the delay due to the processing time in the indexing of contents, thereby offering anyone of almost instantaneous access to any information, unfortunately including the untruthful ones.

5.2 The Rise of Misinformation

The long flagship feature of these platforms, the democratic folksonomic model, unequivocally paved the way for misinformation. The bottom-up structure offered direction of information flows both vertical and horizontal alike almost without restrictions. One of the hopes of folksonomy is to improve collaboration [23] through the contribution of tags by a broad spectrum of users [24]. But then, the current misinformation environment reveals how impaired conspiratorial and hyper-partisan extremists used this structure to pollute these platforms. Unlike the traditional taxonomic models where the gateways (i.e., information professionals such as subject matter experts, fact-checkers) verify the integrity of information prior to its distribution, the right of the users to free speech overpowered the right to access the truth.

5.3 Crowd-Driven Tags as Enforcers of Misinformation

Tags, as an essential component of this bottom-up structure, have been weaponized to shield misinformation. The characteristics of tags to reflect personal association, categories, and concepts for content representation have been exploited. Debunked beliefs, including conspiracy theories, have been labeled as opinions and personal views by some users [25]. The dangerous contradiction is when an impaired and compromised community cannot accept a scientific truth that goes against their belief. Tags of fake news and

lies are contradictorily associated with such truths. For instance, people who refuse to vaccinate spread misinformation and claim how vaccines are the means of government to install a tracker in their bodies or even how vaccines are even more dangerous than the actual diseases. Medical and scientific campaigns that call for vaccine awareness are unfortunately tagged as lies by crowds of misinformation actors. Misinformation draws its strength from the density of users, and this model provides that possibility through the virality of content enhanced by tags.

5.4 Online Communities of Misinformation

A more concrete effect of ultra-personalization in information discovery and retrieval in these social networking platforms is the creation of online communities grounded on false beliefs. Social tags allow users to connect with other users that share common beliefs and interests, including misinformation. Unlike taxonomic classification, the vocabulary in folksonomy directly reflects the user's vocabulary, and people with the same demographic, socioeconomic, culture, and perceptions [26] tend to group themselves in these networks socially. This grouping is made easier by recommender systems. This is a factor of political polarization, a divide of people based on opinions or beliefs. The knowledge structured from communities of misinformation is based on *blind faith.* Scientific and factual truths become relative positions based on the community one belongs.

6 The Classical Functions of Taxonomic Authority as Misinformation Safeguards

6.1 Natural Characteristics of Taxonomy Against Misinformation

Taxonomies, in their own intrinsic authoritative structure, might give the impression of being dictatorial, but it creates an information environment that allows for efficient searching, sorting, and reporting. An information architecture that is grounded on a hierarchical system of classifying information [27] with clear guidelines and categorizations may put misinformation in its rightful place. For instance, the current content integrity descriptors in social networking platforms identifying posts as misleading and verified when enforced efficiently may offer a sense of clarity to users. The authoritative fact-checkers, functioning as third-party subject matter experts, dictate the integrity of the content as per their current scientific state.

This taxonomic system, when communicated regularly, may limit the growth of misinformation through a simple system of categorization. First-person accounts and opinions should be clearly labeled as personal views rather than binding claims or "claimed" personal truths. This is because there are simple truths and facts that are not subject to personal interpretations, such as on the topics of public health and national security.

6.2 The Bias Neutralizer in News and Media

Whereas the truth may be absolute, the way in which it is expressed through linguistics or lexicons may be relatively subjective. For instance, the very keyword misinformation has

many associated synonyms, including the colloquialism *fake news* or even deception. In an attempt by most media organizations to keep their subscribers and viewers maximally engaged, the need to cater to the latter's dispositions may be a priority to the former. This is the reason why news headlines may have different framing of their words and lexical structures.

A taxonomic approach to the wordings of news headlines may provide an even more objective approach to their presentation. Current proposals to implement linguistic technologies against misinformation are introduced to detect biases and to grade the political spectrums of vocabularies. These technologies are grounded on semantic relations, including synsets with the more generic, objective terms positioned on top of the hierarchy. Although the synonyms, including hyponyms, hypernyms, meronyms, may have various interpretations that change over time—the authoritative term as the reference index remains the same. This can limit the semantic inconsistencies explicitly used by some news organization that implicitly amplify the social polarization in a nation's citizens and as a natural consequence, promotes the culture of misinformation.

6.3 Achieving Conjunctive Balance with Folksonomy

While the previous sections highlight the shortcomings of the folksonomy and strength of taxonomic classifications in social networks, this paper does not recommend the termination of the former and the complete enforcement of the latter. As always, a harmonious balance between the two philosophies is the ideal model of a functioning platform grounded on efficiency and integrity. An absolute taxonomic approach to control misinformation will result in the content of the environments losing personalized and relatable topics to the users. On the other hand, the current structure in the platforms is on the opposite spectrum, where an almost absolute folksonomic culture of tagging is exactly what amplified (if not directly caused) the societal problems of polarization and misinformation. A social network knowledge model that considers the strength of both approaches where one remedies the weakness of the other creates an information environment where misinformation may be difficult to thrive.

7 Conclusion

Misinformation, in its very existence, is the antithesis of Information Organization. Through the lenses of Knowledge Structures, Representations, and Models, this paper examined how the phenomenon of misinformation emerges and lingers in an information environment. This paper, with regards to the science of misinformation, shifted the analytical focus from the typical domains of such as threat actors, victims, motivations to the system that enables it on the background, the *misinformation environment*. From an ecosystem view, the interacting elements that make it possible to happen are also the key factors for its control and regulation. At the surface, although it may appear that the most trivial of processes may not seem to affect its growth, such as index rankings, labels in content, miswording—however, the current state of the social networking platforms suggests otherwise.

As this paper had revealed, a substantial contributor to the culture of misinformation is the blueprint of the classificatory system in social media. The combination of taxonomy and folksonomy, drawing from the strength of one another, remains the optimal model. As a literary subject, misinformation is highly interdisciplinary [28]—and the place of information science in the core domains investigating its nature will remain unchallenged and imperative.

8 Limitations

As a complex phenomenon, misinformation has many contributory factors outside the scope of knowledge structures and representations. For instance, the determinative effects of sociological design in different countries and societies. Misinformation can be a symptom of an even deeper societal problem, such as the divide in a country's citizens rooted in culture, economic disparity, and politics. In cases like these, misinformation will persist outside online communities and social networking sites. The misinformation flows in these online mediums may be temporarily controlled, but people will simply find new ways of creating and disseminating them.

An even more dangerous subset of misinformation, *dezinformatsiya* or disinformation, is information created and carefully engineered to deceive. Unlike misinformation, disinformation actors smartly play the rules of knowledge structures and representations to completely mislead, usually for the purpose of cyber warfare, computed propaganda, and political manipulations. Their strategies include but are not limited to mass astroturfing and *deepfakes*, among other usually state-sponsored acts. Unlike a misinformation actor that can be corrected through training and awareness programs, disinformation actors act on blind faith. In cases such as these, more aggressive and stringent information policies grounded on legal remedies are necessary for their control. This subset of misinformation is outside the scope of this paper and falls under the domains of cybersecurity and digital forensics.

References

1. Fox, M.J., Wilkerson, P.L.: Introduction to Archival Organization and Description. Getty Publications (1999)
2. Hedden, H.: Chapter 4 creating relationships. In: The Accidental Taxonomist, pp.97–127. Information Today, New Jersey (2010)
3. Lokers, R., Knapen, R., Janssen, S., van Randen, Y., Jansen, J.: Analysis of big data technologies for use in agro-environmental science. Environ. Model. Softw. **84**, 494–504 (2016)
4. Dezsö, Z., Almaas, E., Lukács, A., Rácz, B., Szakadát, I., Barabási, A.L.: Dynamics of information access on the web. Phys. Rev. E **73**(6), 066132 (2006)
5. Tombros, A., Jose, J.M., Ruthven, I.: Clustering top-ranking sentences for information access. In: Koch, T., Sølvberg, I.T. (eds.) ECDL 2003. LNCS, vol. 2769, pp. 523–528. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45175-4_47
6. Papetti, S.: Design and perceptual investigations of audio-tactile interactions. In: Proceedings of AIA DAGA (2013)

7. Mastley, C.P.: Social media and information behavior: a citation analysis of current research from 2008–2015. Ser. Libr. **73**(3–4), 339–351 (2017)
8. Intezari, A., Taskin, N., Pauleen, D.J.: Looking beyond knowledge sharing: an integrative approach to knowledge management culture. J. Knowl. Manage. (2017)
9. Proper, H.A., Bruza, P.D.: What is information discovery about? J. Am. Soc. Inf. Sci. **50**(9), 737–750 (1999)
10. Riley, J.: Understanding Metadata: What is Metadata, and What is it For? NISO Press, Washington (2017)
11. Eckhardt, A.: Various aspects of user preference learning and recommender systems. In: DATESO, pp. 56–67 (2009)
12. Koltay, T.: Information overload, information architecture and digital literacy. Bull. Am. Soc. Inf. Sci. Technol. **38**(1), 33–35 (2011)
13. Ristoski, P., Paulheim, H.: Semantic Web in data mining and knowledge discovery: a comprehensive survey. J. Web Semant. **36**, 1–22 (2016)
14. Smiraglia, R.P.: Domain analysis of domain analysis for knowledge organization: observations on an emergent methodological cluster. KO Knowl. Organ. **42**(8), 602–614 (2015)
15. Caramancion, K.M.: An exploration of disinformation as a cybersecurity threat. In: 2020 3rd International Conference on Information and Computer Technologies (ICICT), pp. 440–444. IEEE, March 2020
16. Caramancion, K.M.: Understanding the impact of contextual clues in misinformation detection. In: 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), pp. 1–6. IEEE, September 2020
17. Caramancion, K.M.: The demographic profile most at risk of being disinformed. In: 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), pp. 1–7. IEEE, April 2021
18. Denny, M.J., Spirling, A.: Text preprocessing for unsupervised learning: why it matters, when it misleads, and what to do about it. Polit. Anal. **26**(2), 168–189 (2018)
19. Sa, N., Yuan, X.: Examining users' partial query modification patterns in voice search. J. Am. Soc. Inf. Sci. **71**(3), 251–263 (2020)
20. Lawtomated: Structured Data vs. Unstructured Data: what are they and why care? (2019). lawtomated.com
21. Caramancion, K.M.: Understanding the association of personal outlook in free speech regulation and the risk of being mis/disinformed. In: 2021 IEEE World AI IoT Congress (AIIoT), pp. 0092–0097. IEEE, May 2021
22. Deng, S., Lin, Y., Liu, Y., Chen, X., Li, H.: How do personality traits shape information-sharing behaviour in social media? Exploring the mediating effect of generalized trust. Inf. Res. Int. Electron. J. **22**(3), n3 (2017)
23. Karunakaran, A., Reddy, M.C., Spence, P.R.: Toward a model of collaborative information behavior in organizations. J. Am. Soc. Inform. Sci. Technol. **64**(12), 2437–2451 (2013)
24. Mai, J.E.: Folksonomies and the new order: authority in the digital disorder. KO Knowl. Organ. **38**(2), 114–122 (2011)
25. Cooke, N.A.: Posttruth, truthiness, and alternative facts: information behavior and critical information consumption for a new age. Libr. Q. **87**(3), 211–221 (2017)
26. Begany, G.M., Sa, N., Yuan, X.: Factors affecting user perception of a spoken language vs. textual search interface: a content analysis. Interact. Comput. **28**(2), 170–180 (2016)
27. Rosenfeld, L., Morville, P.: Information Architecture for the World Wide Web. O'Reilly Media, Inc. (2002)
28. López-Huertas, M.J.: Domain analysis for interdisciplinary knowledge domains. KO Knowl. Organ. **42**(8), 570–580 (2015)