ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 9, Issue 2, March 2020

Exploring multidisciplinary research themes in approachable to negative impacts of emerging technologies

Heeyeul Kwon[a], Haejin Jo[b], Yongtae Park[c*]

[a]Industrial Engineering Department, Seoul National University, Republic of Korea

[b]Industrial Engineering Department, Seoul National University, Republic of Korea

[c]Industrial Engineering Department, Seoul National University, Republic of Korea

Abstract: ***Ex-ante exploration as well as proactive prevention of those unprecedented adverse consequences of emerging technologies is vital but difficult task for governments and businesses wishing to outperform global technology competition. Whilst previous attempts have been made in participatory and qualitative manner, we herein propose a quantitative way of identifying various research themes regarding those negative impacts and suggesting which disciplinary fields required to resolve each problematic situation. In details, we have comprehensively observed diverse views and opinions regarding future technologies on the web by applying text mining techniques of latent semantic analysis (LSA) and IdeaGraph.The proposed methodology could first provide a number of term clusters that illustrate specific instances where social, environmental, economic, political consequences arise due to the diffusion of emerging technologies. Moreover, it could suggest different disciplinary fields, which must collaborate with the developers of the emerging technologies to resolve potential problems prior to the widespread use. This is expected to be served as an essential tool for various technology corporations and government agencies taking advantage of new technologies in their business and social development.***

Keywords: **emerging technology, idea graph, latent semantic analysis, multidisciplinary, text mining.**

I. BACKGROUND/ OBJECTIVES AND GOALS

Many technologies, especially newly emerging ones, are drastically changing our social, political, economic, and business environments. Despite its wide-ranging benefits, their adverse consequences have become more troublesome. For instance, the emergence of 3d printer has caused various safety issues when used in medicine or food and environmental issues when disposing hazardous waste. Ex-ante exploration and proactive prevention are vital tasks for technology corporate and governments to perform prior to technologies' wide-spread use. However, many are currently faced with some critical challenges in doing so: foreseeing hazardous impacts of new technologies that have not yet occurred and deciding which sort of experts to involve for preventing those situations from happening.

Numerous attempts have been made both in practice and academics to solve such issues. Technology corporates normally gather a group of experts from different backgrounds and discuss through the potential problems and solutions. In academics, the field of technology assessment (TA) has dealt approaches, including impact analysis, Delphi analysis, and risk assessment (Tran & Daim 2008). As seen, the conventional approaches incorporated

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 9, Issue 2, March 2020
participatory methodologies and performed in a qualitative manner. However, the future society affected by new technologies is just too uncertain and complex. Emerging technologies are not just equipped with unprecedented functions but also involved with diverse players and industries. We therefore attempt to approach in a quite different manner. We take advantage of the ample knowledge source provided from the web. By analyzing such a wide variety of opinions with text mining techniques like Latent Semantic Analysis (LSA) and IdeaGraph, we identify numerous research themes regarding negative impacts of future technologies and suggest which disciplinary fields are required to resolve each problematic situation.

II. METHOD

LSA is a widely accepted information retrieval algorithm for extracting and representing textual materials into a semantic structure (Dumais et al. 1988). The methodology holds the assumption that there exist some underlying and latent structures in word usage, which is obscured from variability of word choice (Berry et al. 1995). A truncated singular value decomposition (SVD) is thus utilized in the term-document matrix to estimate the structure in word usage across a number of documents (Golub & Van Loan, 2012). Such a procedure is wholly dependent on the words co-occurrence paths of terms constituting the documents. This method has been applied in various tasks, including cross-language information retrieval (CL-LSA) (Littman et al. 1998), source-code clustering (Maletic & Marcus 2000), detection of plagiarism (Cosma & Joy 2012), and most importantly, topic detection (Sidorova et al. 2008). We, as well, applied LSA for the purpose of summarizing the massive number of texts into multiple groups of general and high frequency terms. It is considered most suitable based on two primary reasons. First, LSA is capable of grouping general terms that each represents certain topic. Second, it allows us to capture latent terms that are normally undetectable due to the variability of word choice. Extended from KeyGraph text mining technique, IdeaGraph is capable of modeling text data as a network to improve human cognition and thus discovering facts, relationships and implications of hidden textual information (Wang et al. 2013). It has been recognized in the fields of knowledge discovery in texts (KDT) and chance discovery (CD). Total four steps are involved in IdeaGraph technique. First, general clusters are generated based on the calculation between the terms presented in the documents. The indicator of calculating the relation is denoted in (1).

$$\mathrm{R}\left(I_i, I_j\right) = P\left(I_i | I_j\right) + P\left(I_j | I_i\right) \quad (1)$$

Where I_i and I_j indicate two separate terms.

Second, cognitive clusters are obtained by measuring information and information density of general clusters. Only those above the pre-set threshold values are confirmed as the cognitive clusters. Third, valuable links of the terms outside the clusters are calculated for each cluster based on the indicator (2). As shown in the indicator, the terms that are low frequency but highly related to the overall constituent terms within the cluster are newly linked as supplementary terms. Finally, key items are derived using the indicator (3).

$$\mathrm{PR}(I_i, CC) = \sum_{c_k \in CC} R(I_i, c_k) \quad (2)$$

Where CC denotes cognitive clusters and c indicate a cluster.

$$\mathrm{Key(I)} = \sum_{i=0}^{N_c} PR(I, CC_i) + \sum_{\notin CC, I_k \in G} R(\mathrm{I}, I_k) \quad (3)$$

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 9, Issue 2, March 2020

Where N_c indicates the number of terms in the cluster c and G indicates the whole graph.

III. PROPOSED PROCESS

A. Step 1. Emerging technology selection

The overall methodology is illustrated in Figure 1. The first step is selecting target emerging technologies for the analysis. Due to their wide ranging types, there is in need of going through numerous futures reports and news articles and investigating the ones that are highly prospective but possible to significantly alter our society.

B. Step 2. Data collection

The second step is the construction of future-oriented database. Apparently, diverse views and opinions regarding newly emerging technologies and their impacts to our society are well-expressed on the web. We thus investigate the websites that hold information specifically about our target technologies and denote them as *future-oriented websites*. The documents are collected using web data crawling program developed based on Java.

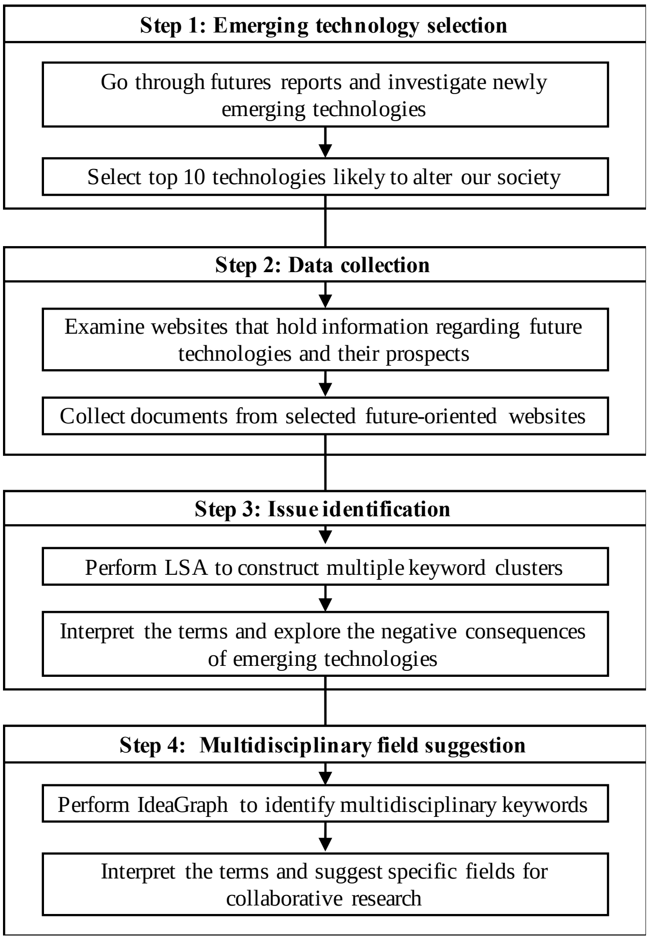

Fig 1. Overall framework of our methodology

C. Step 3. Issue identification

The third step is identification of various details of issues that depict the ways of negatively affecting our society, environment, economy, etc. Once we collect as much information as possible form the web, there will

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 9, Issue 2, March 2020

be hundreds and thousands of web documents presented in the database. Here, LSA text mining technique is applied for the purpose of comprehensively and accurately observing the voices expressed in the documents by forming multiple term clusters. We first construct term-document matrix, which represent the relations between terms and documents using the occurrence frequency values of terms in the documents, as shown in Figure 2. When LSA is performed, the term-document matrix is both decomposed and truncated into three different matrices. The multiplication of the term eigenvectors and diagonal matrix of singular values will denote the term clusters, and they can be further interpreted as the depictions of issues.

$$
\begin{array}{c c}
 & \begin{array}{ccc} \textit{Document 1} & \ldots & \textit{Document k} \end{array} \\
\begin{array}{c} \textit{Keyword 1} \\ \textit{Keyword 2} \\ \vdots \\ \textit{Keyword n} \end{array} &
\begin{pmatrix} f_{11} & f_{12} & \ldots & f_{1k} \\ \vdots & & & \vdots \\ f_{n1} & f_{n2} & \ldots & f_{nk} \end{pmatrix}
\end{array}
$$

Fig 2. Illustration of a term-document matrix

D. Step 4. Multidisciplinary field suggestion

The fourth step is suggesting multidisciplinary fields for each and every issue depiction. We set the results of term clusters as inputs for IdeaGraph text mining technique and further identify very distinct but highly related multidisciplinary terms to the issues using step 3 of IdeaGraph. The word *distinct* is used to indicate the low frequency occurrence or hidden features of the terms. The scores for distinctness is calculated based on the indicator (2). From the list of top 30 terms, we select the multidisciplinary terms that are capable of the further interpretation. When successfully processed, we observe the terms that are not initially included in the cluster due to its low frequency value but highly related when considering the overall relationships with the constituting terms. Here, the selected multidisciplinary terms can be varied depending on the focus of the issues. Once IdeaGraph captures multidisciplinary terms, we incorporate multiple futures and target technology experts and suggest the lists of specific fields, which are necessary to solve each issue.

IV. RESULTS

We are currently in the process of deriving full descriptions of issues and interpreting the multidisciplinary terms for the suggestion of disciplinary collaboration. Even though we have not derived full result of our proposed methodology, we expect some novel and useful results, which could significantly support both private and public sectors in various technology development-related decision makings. First, the methodology provides a number of term clusters that illustrate specific instances where social, environmental, economic, political consequences arise due to the diffusion of emerging technologies, as depicted in Figure 3. Within the resulting clusters, there exist the terms with fruitful details of specific places, involved actors, affected people, human values, etc. Those terms are used as the main keywords to understand the issue-related instances. For example, if the terms of *toxic, lung, lethal, safety,* and *danger* are grouped as a cluster in *fuel cell* technology, we could interpret the issue as, “the diffusion of fuel cell technology may cause some safety problems due to its toxic

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 9, Issue 2, March 2020

chemical lethally affecting our lungs”. Based on this procedure, we construct as many term clusters as possible to comprehensively overlook the issues regarding the emerging technologies. This method is considered effective because it is capable of providing short summaries from a number of negative opinions of both the experts and the public.

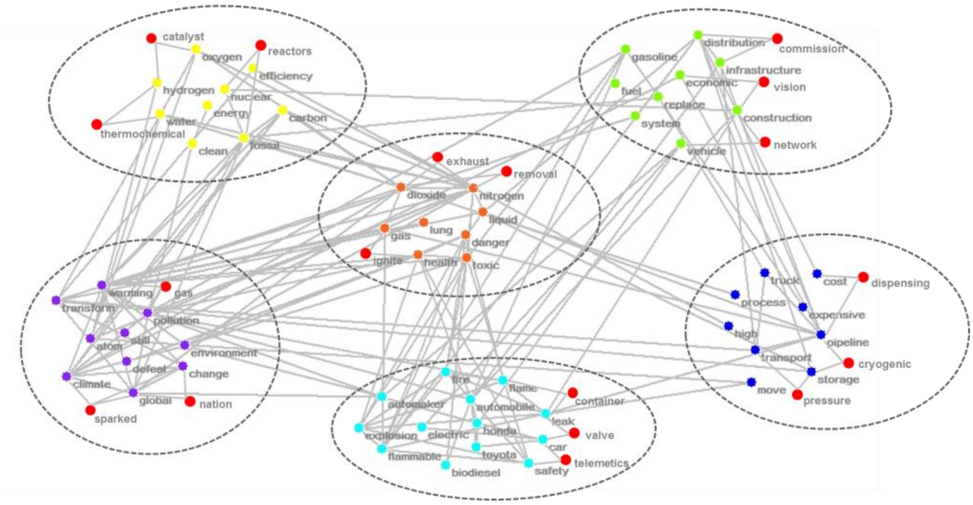

Fig 3. IdeaGraph result of fuel cell technology

Second, it suggests different disciplinary fields, which must collaborate with the developers of the emerging technologies to resolve potential problems prior to the widespread use. With the help of IdeaGraph, we observe hidden or latent multidisciplinary terms that are highly related to the issue clusters. For example, we have analyzed the issues of *virtual reality* technology and found out a privacy issue of virtual behavioral patterns. Then, the result of IdeaGraph has extracted the hidden term of *vere*, or virtual embodiment and robotic embodiment, and it was concluded that the research theme of *vere* is highly recommended with the cooperation of neuroscientists. Based on this short illustrative results, we have shown the study’s prospective potential of becoming an essential tool for various technology corporations and government agencies wishing to take advantage of new technologies in their business and social development. However, there is a major methodological limitation that must be thoroughly discussed. Although we propose a quantitative way of exploring negative issues and their potential solutions, the interpretation of issues and multidisciplinary suggestions still requires qualitative efforts from the experts. It would be useful for a future study to investigate a more systematic and sophisticated approach.

REFERENCES

[1] Berry, M. W., Dumais, S. T., & O'Brien, G. W.(1995). Using linear algebra for intelligent information retrieval. SIAM review, 37(4), 573-595.

[2] Cosma, G., & Joy, M.(2012). An approach to source-code plagiarism detection and investigation using latent semantic analysis. IEEE transactions on computers, 61(3), 379-394.

ISSN: 2319-5967
ISO 9001:2008 Certified
International Journal of Engineering Science and Innovative Technology (IJESIT)
Volume 9, Issue 2, March 2020

[3] Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R.(1988). Using latent semantic analysis to improve access to textual information. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, USA: ACM Press.

[4] Golub, G. H., & Van Loan, C. F.(2012). Matrix computations (Vol. 3). Baltimore and London: Johns Hopkins University Press.

[5] Littman, M. L., Dumais, S. T., & Landauer, T. K.(1998). Automatic cross-language information retrieval using latent semantic indexing. In G. Grefenstette (Ed.), Cross-language information retrieval (pp. 51-62). Boston, MA: Springer.

[6] Maletic, J., & Marcus, A.(2000). Using latent semantic analysis to identify similarities in source code to support program understanding. Proceedings of 12th IEEE International Conference on the Tools with Artificial Intelligence. Vancouver, BC: IEEE.

[7] Sidorova, A., Evangelopoulos, N., Valacich, J. S., & Ramakrishnan, T.(2008). Uncovering the intellectual core of the information systems discipline. MIS Quarterly, 32(3), 467-482.

[8] Tran, T. A., & Daim, T. (2008). A taxonomic review of methods and tools applied in technology assessment. Technological Forecasting and Social Change, 75(9), 1396-1405.

[9] Wang, H., Xu, F., Hu, X., & Ohsawa, Y. (2013). IdeaGraph: a graph-based algorithm of mining latent information for human cognition. In 2013 IEEE International Conference on Systems, Man, and Cybernetics. IEEE.