INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

A Survey on Forms of Visualization and Tools Used in Topic Modelling

Ruhaila Maskat [a], Shazlyn Milleana Shaharudin [b,*], Deden Witarsyah [c], Hairulnizam Mahdin [d]

[a] *Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Shah Alam, Selangor Malaysia*
[b] *Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, Tanjong Malim 35900, Malaysia*
[c] *Information System Telkom University Bandung, Indonesia*
[d] *Center of Intelligent and Autonomous System (CIAS), Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400, Batu Pahat, Johor, Malaysia*
Corresponding author: *shazlyn@fsmt.upsi.edu.my*

Abstract— In this paper, we surveyed recent publications on topic modeling and analyzed the forms of visualizations and tools used. Expectedly, this information will help Natural Language Processing (NLP) researchers to make better decisions about which types of visualization are appropriate for them and which tools can help them. This could also spark further development of existing visualizations or the emergence of new visualizations if a gap is present. Topic modeling is an NLP technique used to identify topics hidden in a collection of documents. Visualizing these topics permits a faster understanding of the underlying subject matter in terms of its domain. This survey covered publications from 2017 to early 2022. The PRISMA methodology was used to review the publications. One hundred articles were collected, and 42 were found eligible for this study after filtration. Two research questions were formulated. The first question asks, "What are the different forms of visualizations used to display the result of topic modeling?" and the second question is "What visualization software or API is used? From our results, we discovered that different forms of visualizations meet different purposes of their display. We categorized them as maps, networks, evolution-based charts, and others. We also discovered that LDAvis is the most frequently used software/API, followed by the R language packages and D3.js. The primary limitation of this survey is it is not exhaustive. Hence, some eligible publications may not be included.

Keywords—Topic visualization; topic modeling; visualization tools; review; survey.

I. INTRODUCTION

Topic modeling represents abstract models of hidden topics represented by a distribution of words highly correlated to a particular topic in a document collection [1]. According to Kherwa and Bansal [1], visualization is one of the challenges of topic modeling apart from its interpretability, computation complexity, and output stability. Visualization allows an easy understanding of the identified topics. Kherwa and Bansal also reported that listing the *top terms* and *word clouds* are the two most common visualizations in topic modeling. In this paper, we review recent literature on topic modeling, which presented its result in visualization and sought to answer the following research questions.

- What are the different visualizations used to display the result of topic modeling?
- What visualization software or API is used?

By answering these questions, NLP researchers can decide which forms of visualization they may find suitable and the tools to assist them. This could also initiate further progress to existing visualizations or their lack thereof, where we can expect to see the birth of new novel visualizations.

In this study, we adopted the PRISMA [2] systematic literature review methodology. Published papers from 2017 to 2022 were extracted, filtered, and analyzed. The search was done via the Web of Science search engine due to its high standards of publication indexing. Thus, expectedly, the papers extracted are of superior quality that have undergone a rigorous peer-review process.

Topic modeling approaches convert a text collection into a low-dimensional topic subspace, a group of words called a topic [3]. Topic models are statistical models that reveal the text data's hidden structure. It allows users to search through a large text collection, a digital library, and web material, among other things [3]. It may be used to describe the corpus

in several different ways, such as the percentage of topics in a document, the number of documents that particular topics covered, and the percentage of the collection that fits into various categories of themes [3]. The remaining part of this paper is organized as follows: Section 2 explains the PRISMA methodology. Section 3 describes the result of the survey. Section 4 discusses the limitations of this survey, and section 5 provides the study's conclusion.

II. MATERIALS AND METHODS

This section explains the steps taken to conduct this systematic review of literature based on the PRISMA methodology. The PRISMA method includes the identification, screening, eligibility, and inclusion phases. In order to perform the identification phase, research questions are formulated to reflect the review's focus and motivation. Next, one or more search queries are constructed, and the time range is determined. Now searching can take place using the queries, and extraction of the publications would yield a collection of papers. In the case there are duplications of papers, they must be removed. The eligibility of each paper is carefully determined by filtering them using predetermined inclusion and exclusion criteria. Papers that meet these criteria are then filtered again for eligibility based on their full text. Only the final listed papers are included in the review to answer the formulated research questions. Figure 1 shows the PRISMA method.
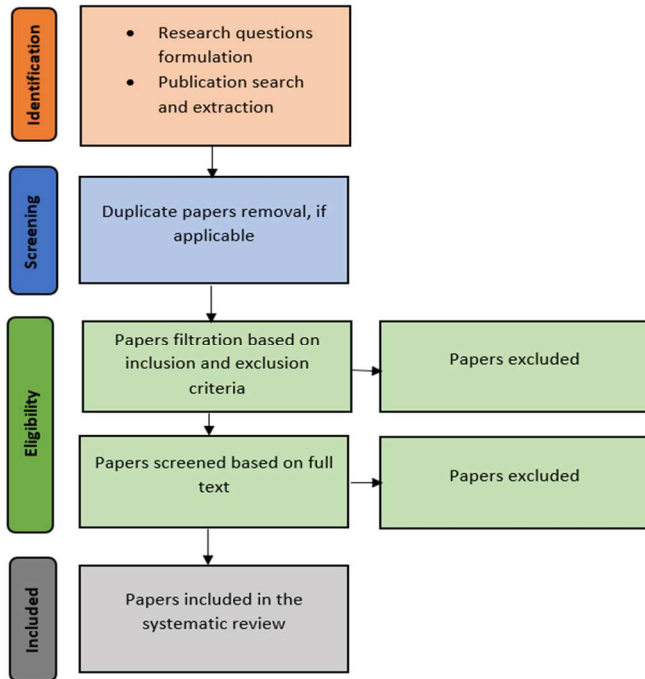

Fig. 1 PRISMA methodology used

Each following subsection describes the detailed activities performed at each phase of the PRISMA methodology.

A. Research Questions Formulation

In this phase, research questions were formulated to assist researchers in deciding the most suitable type of visualization they could use by knowing what visualization types have already been adopted by earlier works on topic modeling. Additionally, researchers would be aware of existing software

tools or API libraries. Table I list the formulated research questions and their motivations.

TABLE I
RESEARCH QUESTIONS AND MOTIVATIONS

Research Questions	Motivations
RQ1: What are the different forms of visualizations used to display the result of topic modelling?	The answer to this question allows researchers to evaluate what forms of visualizations are available and which can be suitable for their usage.
RQ2: What visualization software or API is used?	The answer to this question allows researchers to know what software tools or APIs are available and where they can access them.

B. Publication Search and Extraction

For publication searching, we resorted to the Web of Science search engine as, at present, it holds the highest level of standard in academic publication indexing, adopting the Journal Citation Reports (JCR) ranking. To conduct the search, we used the following search query, combining the logical Boolean operators of AND and OR. We aimed to find all papers on topic modeling that also presented a visualization of its result. Therefore, we specified in the Advanced Search to search for "topic modeling" in the publication title, while for "visualization" in the abstract. The logic behind this is that a topic modeling paper that displayed a visualization as a secondary contribution would often highlight it in the abstract if the work is a secondary contribution and worth mentioning.

In these papers, we expect some useful explanation, justification, or description of the visualization. For anything not highlighted in the abstract, we considered the visualization efforts to be too minimal to be included in the review. The following is the query we used. (TI=(topic modelling)) AND ((AB=(visualize)) OR AB=(visualization)) The search was conducted from 23rd to 29th May 2022. The years are scoped from 2017 to 2022. Finally, the publications are extracted and uploaded to the Mendeley reference manager for filtration.

C. Publication Filtration

The next phase of the PRISMA method consists of a two-level filtration of the collected publications to determine their eligibility (See Fig. 2). The first level was conducted by applying preconstructed criteria of inclusion and exclusion.

1) Inclusion:
- *I1: The paper should perform topic modeling as its main contribution.*
- *I2: Paper should visualize the topic modeling result as its secondary contribution or have a good justification/explanation.*

2) Exclusion:
- *E1: Papers not directly related to the topic modeling task.*
- *E2: Papers that focus is not topic modeling.*
- *E3: Papers not written in English.*
- *E4: Papers not accessible via our university's subscription.*

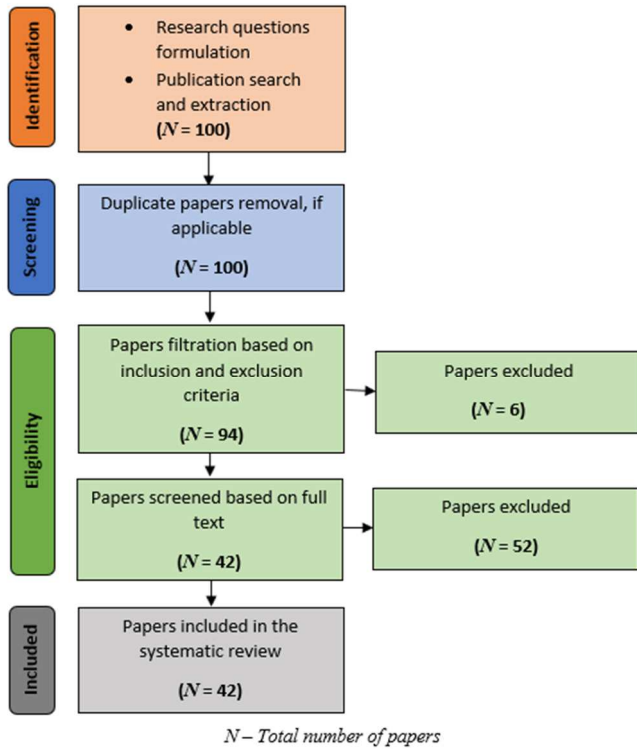- *E5: Papers that do not display any topic visualizations used/proposed.*

Fig. 2 Filtration steps using PRISMA methodology.

The second level of filtration involves careful perusal of each paper's full text to determine its true relevance to the focus of this review. They must clearly show the visualization, justify its selection, and explain it.

III. RESULTS AND DISCUSSIONS

This section discusses the result of this review, which is the list of collected publications found relevant to this review's focus (Table II) and the analysis conducted to answer the formulated research questions. Before that, we show in Figure 2 the number of papers we acquired and found eligible throughout the PRISMA phases. Our initial collection identified 100 candidate papers from the Web of Science. No duplicate documents were found. Filtration was conducted to ascertain the articles' eligibility. During the first level of filtration, based on the inclusion and exclusion criteria determined earlier, six articles were excluded, leaving only 94. After a more detailed perusal was performed at the second level of filtration, 52 papers were further excluded. Therefore, the number of published articles found eligible is 42.

A. RQ1: What are the different forms of visualizations used to display the result of topic modelling?

From our review, different visualizations were used to effectively portray topics based on their suitability with the underlying domain.

1) Maps: Studies with geoinformation can take advantage of *a* geographical map to show a clearer picture of the result. With maps, different shades of color represent different data points. This includes (Fig. 3) basic map (i), heat map (ii) time-lapse map (iii), and glyph map (iv).
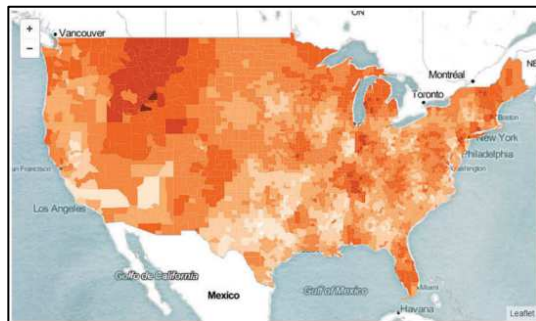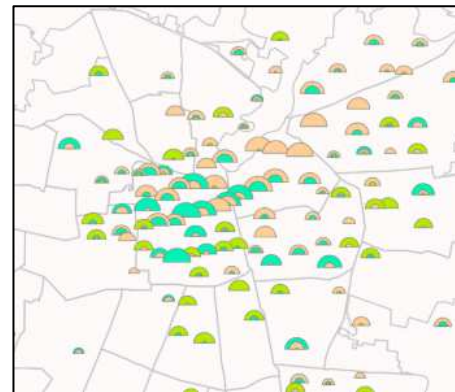

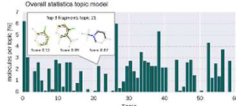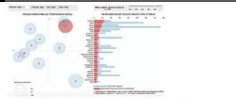(i). Basic map
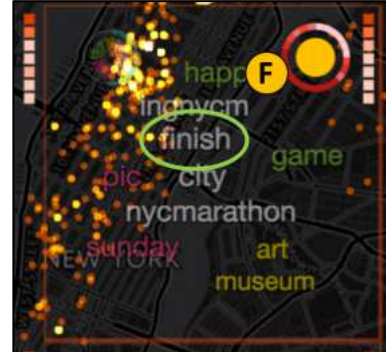
(ii). Heat map
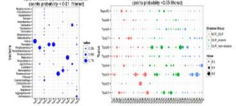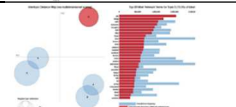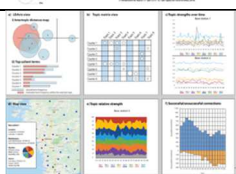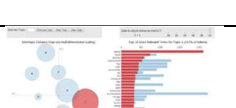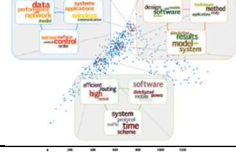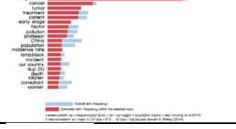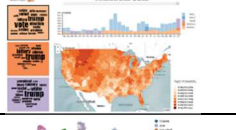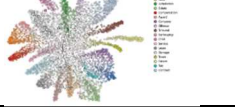
(iii). Time-lapse map
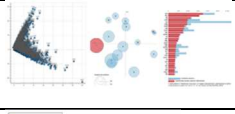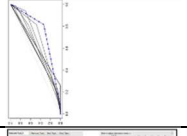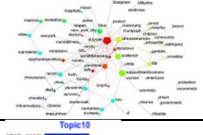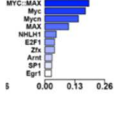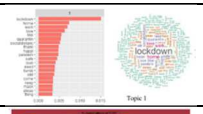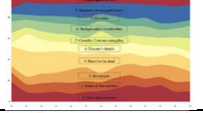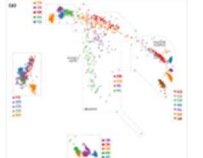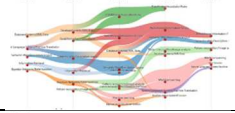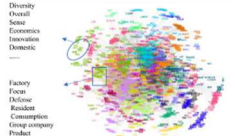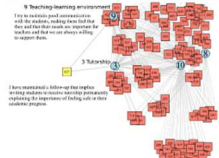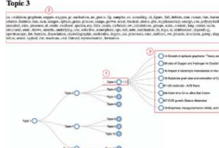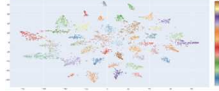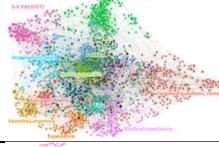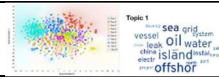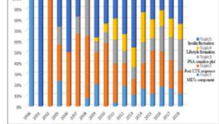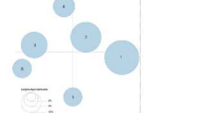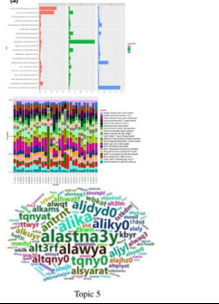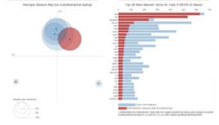
(iv). Glyph map

Fig. 3 Map Type of Visualization

TABLE II
RELEVANT PAPERS

Author	Year	Domain	Data description	Data source	Visualization form	Visualization software	Visual
Martin and Schuurman [9]	2017	Geospatial	Spatial locations of tweets in the neighbourhood.	Twitter	Map	http://www.github.com/ mikedotonline/VisualTopicModels	
Schneider et al. [10]	2017	Chemical molecular	Molecular fragment data.	ChEMBL database	Bar chart	http://www.github.com/rdkit/CheTo	
Liu and Jansson [11]	2017	Urban development	City and regional events.	Instagram, Twitter	Stacked bar, Bubble chart	LDAvis	
Yan et al. [12]	2017	Metagenomics	Oral metagenomics data	Self-collected	Relation matrix	https://github. com/bm2-lab/MetaTopics	
Manna and Phongpanangam [13]	2018	Crisis event	Tweets on crisis	CrisisLexT6 dataset	Stacked bar, Bubble chart	LDAvis	
Helldin et al. [14]	2018	Telecommunication network	Mirrored coherent processes present in the network data	Base station counter readings	Stacked bar, Bubble chart, Map, Matrix, Line chart, Stacked area chart, Double bar chart,	LDAvis	
Cheng et al. [15]	2018	Ecology, environment and poverty	Published literature records with title and abstract	Academic online databases	Stacked bar, Bubble chart	LDAvis	
Choi et al. [16]	2018	Real-time natural disaster detection	Tweets of real-time disaster	Twitter	Glyphs, Heatmap, Word cloud	Self-developed	
Jin et al. [17]	2019	Academic publications	Published literature records	DBLP and hep-th datasets	Word cloud	Not stated	
Liu et al. [18]	2019	Medical	News articles on thirdhand smoking	Wiser and Factiva databases	Stacked bar, Bubble chart	LDAvis	
Prasad et al. [19]	2019	Social healthcare	Health tweets, health key phrases benchmarks	Twitter, TREC2015, TREC2014	Word cloud, Visual clusters	Not stated	
Koylu [20]	2019	Political	Interpersonal communication	Twitter	Word cloud, Map, Bar chart	http://www.geo-social.com/topicviz/mentiontopics.html	
Carter and Rahmani [21]	2019	Law and legislation	Court decisions	Australian High Court	2D scatterplot	Uniform Manifold Approximation and Projection ('UMAP')	

Goswami et al. [22]	2019	e-commerce	Transaction data	Walmart	Word cloud	Self-developed	
Ray et al. [23]	2019	News	Hindi news articles	Amar Ujala and Navbharat Times newspapers	2D scatterplot, Stacked bar, Bubble chart	LDAvis, Bokeh	
Zafari and Ekin [24]	2019	Medical	Medical claims	Centres for Medicare and Medicaid	Concentration function	R's stm package	
Negara et al. [25]	2019	Open domain	Tweets related to Sports, Economy, Military & Technology	Twitter	Stacked bar, Bubble chart	LDAvis	
Asghari et al. [26]	2020	Medical	US new on the cost of insulin for diabetics	Twitter	Time-lapse map	Not stated	
Odlum et al. [27]	2020	Covid-19	Covid-19 tweets related to African American	Twitter	Network	Not stated	
Yang et al. [28]	2020	Genome	Combinatorial binding events of multiple transcription factors (TFs)	ChIP-Seq datasets	Bar chart	https://github.com/lichen-lab/tfLDA	
Tao et al. [29]	2020	Covid-19	Covid-19 related tweets	Twitter	Bar chart, Word cloud	Not stated	
Deng et al. [30]	2020	Crisis communication	Tianjin explosion-related messages	Weibo	Streamgraph	R's streamgraph package	
Juan et al. [31]	2020	Bioinformatics	Genome similarity	1000 Genomes Project FTP site	Heatmap	R's smacof package	
Miyata et al. [32]	2020	Library and information science (LIS)	Academic articles relating to LIS	Information Processing & Management (IPM) Journal of Documentation (JDOC) Journal of the Association for Information Science and Technology (JASIST) Library & Information Science Research (LISR) Library Quarterly (LQ)	2D scatterplot	LDAvis	
Liu et al. [33]	2020	Evolutionary path	AI-related scientific publication	Citeseer, Cora, Aminer	Evolutionary path map	D3.js	
Zhang et al. [34]	2020	Mining scientific literature	Scientific literature	Natural Language Processing group of Fudan University, China	Semantic graph	Not stated	

Reference	Year	Domain	Data	Data source	Visualization	Tool	Example
Buenano-Fernandez et al. [35]	2020	Higher learning institution	Teacher self-assessment survey containing open-ended questions	Ecuadorian university	Network	Not stated	
Cheng et al. [36]	2020	Bioengineering	Documents relating to bioengineering popular terms	National Centre for Biotechnology Information (NCBI), Google Scholar, CiteSeerX	Hierarchy tree	https://sites.google.com/unomaha.edu/topexplorer	
Bulatov et al. [37]	2020	Open domain	Accepted documents	Neural Information Processing Systems (NIPS) machine learning conference	2D scatterplot	github.com/ machine-intelligence-laboratory/TopicNet	
Opitz et al. [38]	2020	Geographic visualization	Situation-related photo images in Santiago, Chile	Yahoo Flickr Creative Commons 100 Million data set	Glyph map	Self-developed	
Kim and Cho [39]	2020	Online communities' user behaviour	Articles in subReddit Politics and AskReddit	Reddit, Slashdot, PGR21	Network diagram	Gephi	
Han et al. [40]	2020	Online news	English and Chinese blockchain-related online news articles	English and Chinese news websites	Word cloud, Map, Bubble chart, Stacked bar chart	D3.js, LDAvis	
Ertek and Kailas [41]	2021	Wind energy	News report relating to wind turbines	CWIF dataset, Google search	2D scatterplot, Word cloud	KNIME	
Zoya et al. [42]	2021	Open domain	Urdu language tweets	Twitter	Word cloud	Not stated	
Kim et al. [43]	2021	Medical	Publications relating to cancer survivors	PubMed, EMBASE, Cochrane, CINAHL, SCOPUS, Web of Science	Stacked bar	Not stated	
Chen et al. [44]	2021	Medical	Cancer-related messages	Weibo	Bubble charts	LDAvis	
Alharbi et al. [45]	2021	News	Arabic newspapers' archives related to security challenges	https://aitnews.com https://www.secure-enough.com http://security4arabs.com https://www.unlimit-tech.com	Word cloud, Bar chart, Stacked bar	R's machine learning and statistical software toolset	
Zhong and Liu [46]	2021	Medical	Web posts by adolescents with IBS and the parents' posts concerning the IBS symptoms experienced by their children	Teens and Children's Issues on the IBS Group forum	Stacked bar, Bubble chart	LDAvis	

Kumari et al. [47]	2021	Robot technology	Publication and patent data related to robot technology	Web of Science, Global Patent Analysis Service system (G-Pass, KISTI) https://gpass.kisti.re.kr	Word cloud, Network	VOSviewer	
Liu et al. [48]	2021	Covid-19	Ongoing advancements in early scientific research on Covid-19	Covid-19 Open Research Dataset (CORD-19), PubMed Central (PMC), BioRXiv preprint servers, MedRxiv preprint servers and WHO Covid-19 database.	Word cloud	Not stated	
Won et al. [49]	2021	Public health	Articles related to infection nursing	Journal of the Association of Nurses in AIDS Care, Journal of Advanced Nursing, Heart and Lung, Journal of Clinical Nursing, Nursing Clinics of North America	Sociogram	NetMiner	
Noble et al. [50]	2021	Medical	Clinical narratives for dogs	Small Animal Veterinary Surveillance Network (SAVSNET)	Word cloud, Stacked bar, Bubble chart	LDAvis	

2) Networks: Studies having connections or associations used networks to display them. Networks implemented in the review (Fig. 4) are semantic graph (i), network diagram (ii), and sociogram (iii). Semantic graphs store data in a rich, contextually, and conceptually formed format – complex data that reflects the real world. A network diagram is a type of data visualization that helps users comprehend the relationship between the data. A sociogram is a graph database that illustrates the connections between individuals in a group to map that group's social network.

(i). Semantic Graph

(ii). Network Diagram

(iii). sociogram

Fig. 4 Network Type of Visualization

3) Evolution-based: Streamgraph, evolutionary path map, and line chart have been applied in the collected literature. All are evolution based, which ties strongly to the element of time. The aim is to show the progression of changes that occurred throughout a specific period. Refer Fig 5. Streamgraph (i) displays an area graph stacked together and placed over the X-axis. This axis represents the change in time, be it in months, years, or decades. The peaks and valleys show the total values over time. Unlike streamgraph, evolutionary path map (ii) does not emphasize the total values over time but instead focuses on the evolutionary aspect of the object of interest over time. Line chart (iii) is a classical visualization often used to show changes in trend as time passes.

(i). Streamgraph

(ii). Evolutionary Path Map

(iii). Line Chart

Fig. 5 Evolution-based Type of Visualization

4) Charts: We simply categorized all visualizations that do not involve geoinformation, connections, or time progression as charts. They include bars, stacked bars, bubbles, scatterplots, etc. A minimum of two axes can be found in these charts. As they are quite common, no further elaborations are necessary.

5) Others: Falling within this category are visualization forms created specifically to satisfy the underlying domain's purpose. They are relation matrix, visual clusters, concentration function, and hierarchy tree. As they are specially purposed, interested readers are invited to refer to the original publication for further perusal.

B. *RQ2: What visualization software or API is used?*

The result of our review showed that LDAvis [4] is the most frequently used software to visualize topics among NLP researchers. It was built using a combination of R language and D3.js [5]. LDAvis is web-based and interactive, offering not only an overview of discovered topics but also permitting drilling into the details of the terms related to each topic. This gives users the flexibility to investigate topic-term relationships based on their relevance.

The second most frequently used software is the R package. The packages used are *stm* [6], which focuses on structural topic modelling, *streamgraph* [7] offering the plotting and construction of a stream graph, *smacof* [8] implements tools relating to multidimensional scaling, machine learning packages offering a variety of algorithms such as Caret (Classification and Regression Training), and statistical packages for statistical analysis.

Another popular API is the D3.js built from the JavaScript library for interactive data visualizations in web browsers. D3 stands for Data-Driven Documents where it is able to utilize Scalable Vector Graphics (SVG), HTML5 and Cascading Style Sheets (CSS). D3 is known to be lightweight and suitable across most web standards, making it suitable for web-based visualization of topic models. Other less popular software/APIs used are Bokeh, UMAP, Gephi, KNIME, VOSviewer, and NetMiner. The rest of the articles involve self-development or simply were not stated.

C. *Limitations to Review Assessment*

This review is not exhaustive; hence, there is room for improvement; future researchers can apply a different perspective to convey another point of view. This section discusses limitations of this study.

The search query coverage may appear limited. However, it can be argued that the terms chosen for the query were popularly used and were able to tap into a large pool of publications, implementing various forms and tools to visualize topic modelling. We do not dismiss there may be publications using a less popular term that were not included.

The search was executed only on the *Web of Science (WoS)* platform. We acknowledge this good objection and the restriction that entails. Nevertheless, this review was intended to inform researchers of the choices they have in visualizing and implementing topic modeling and this review is not a comprehensive list of every possible variety that existed. WoS is a prestigious platform where one could expect the latest state-of-the-art studies are published.

D. *Future Directions*

This exercise allowed us to look closely into the plethora of selections used in visualizing topic modeling results. Apparently, their creations were motivated by the goal to highlight essential aspects discovered. Maps, networks, and charts, each has specific unshareable features. Research and technological advancement are the underlying forces driving the design of these visualizations. To guess the future direction is to guess future research interest and technological opportunities. Hence, with the ever-increasing generation of data, we expect to see a faster improvement in sophistication for the network-type visualizations. This is to support the expanding complex relationships and links of open data.

Tools implementable through APIs to visualize results in topic modelling are, to a degree, chosen based on the algorithm a researcher utilized, for example LDAvis. Pre-existing generic libraries, e.g. R, which are the building blocks behind these user-friendly APIs, require considerable programming knowledge to build visualizations from scratch. Data science is an area with a high diversity of entries; technically or functionally adept. With the progressive reach of data science to more domains, the latter group of data scientists' growth will supercede the former, hence, a more user-friendly tool will be demanded.

IV. Conclusion

In conclusion, this paper presented a recent survey on the different forms of visualization and tools used in topic modeling. The period covered was from 2017 to early 2022. After conducting the PRISMA method of systematic reviewing, 42 papers were found eligible for this study. Based on our findings, we discovered that different types of visualizations serve different purposes in terms of display. They were classified as maps, networks, evolution-based, charts, and others. We also discovered that the most commonly used software/API is LDAvis, followed by R language packages and D3.js. The primary limitation of this survey is that it is not exhaustive; thus, some eligible publications may be missed. Our review is not without limitations. The first limitation involves the finalized list of articles reviewed. This list is not exhaustive. It depended substantially on the subscription to online databases by our university. Additionally, this paper's writing time may plausibly exclude newer eligible publications.

References

[1] P. Kherwa and P. Bansal, "Topic Modeling: A Comprehensive Review EAI Endorsed Transactions on Scalable Information Systems," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 7, no. 24, pp. 1–16, 2019.

[2] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *The BMJ*, vol. 372, 2021, doi: 10.1136/bmj.n71.

[3] U. Chauhan and A. Shah, "Topic Modeling Using Latent Dirichlet allocation: A Survey," *ACM Computing Surveys*, vol. 54, no. 7, 2022, doi: 10.1145/3462478.

[4] C. Sievert and K. Shirley, "LDAvis: A Method for Visualizing and Interpreting Topics," in *Workshop on Interactive Language Learning, Visualization and Interfaces*, 2015, pp. 63–70, doi: 10.3115/v1/w14-3110.

[5] P. N. Castillo, *Mastering D3. js*. Packt Publishing Ltd, 2014.

[6] M. E. Roberts, B. M. Stewart, and D. Tingley, "stm: R Package for Structural Topic Models," *Journal of Statistical Software*, vol. 91, no. 1, pp. 1–40, 2019, doi: 10.18637/jss.v000.i00.

[7] A. A. Haidar, B. Yang, and J. G. Ganascia, "Visualizing the first world war using StreamGraphs and information extraction," *Proceedings of the International Conference on Information Visualisation*, vol. 2016-Augus, pp. 290–293, 2016, doi: 10.1109/IV.2016.81.

[8] J. de Leeuw and P. Mair, "Multidimensional scaling using majorization: SMACOF in R," *Journal of Statistical Software*, vol. 31, no. 3, pp. 1–30, 2009, doi: 10.18637/jss.v031.i03.

[9] M. E. Martin and N. Schuurman, "Area-Based Topic Modeling and Visualization of Social Media for Qualitative GIS," *Annals of the American Association of Geographers*, vol. 107, no. 5, pp. 1028–1039, 2017, doi: 10.1080/24694452.2017.1293499.

[10] N. Schneider, N. Fechner, G. A. Landrum, and N. Stiefl, "Chemical Topic Modeling: Exploring Molecular Data Sets Using a Common Text-Mining Approach," *Journal of Chemical Information and Modeling*, vol. 57, no. 8, pp. 1816–1831, 2017, doi: 10.1021/acs.jcim.7b00249.

[11] S. Liu and P. Jansson, "City event detection from social media with neural embeddings and topic model visualization," *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, vol. 2018-Janua, no. 2012, pp. 4111–4116, 2017, doi: 10.1109/BigData.2017.8258430.

[12] J. Yan *et al.*, "MetaTopics: An integration tool to analyze microbial community profile by topic model," *BMC Genomics*, vol. 18, no. Suppl 1, pp. 1–5, 2017, doi: 10.1186/s12864-016-3257-2.

[13] S. Manna and O. Phongpanangam, "Exploring Topic Models on Short Texts: A Case Study with Crisis Data," *Proceedings - 2nd IEEE International Conference on Robotic Computing, IRC 2018*, vol. 2018-Janua, pp. 377–382, 2018, doi: 10.1109/IRC.2018.00078.

[14] T. Helldin, H. J. Steinhauer, A. Karlsson, and G. Mathiason, "Situation Awareness in Telecommunication Networks Using Topic Modeling," *2018 21st International Conference on Information Fusion, FUSION 2018*, pp. 549–556, 2018, doi: 10.23919/ICIF.2018.8455529.

[15] X. Cheng *et al.*, "Topic modelling of ecology, environment and poverty nexus: An integrated framework," *Agriculture, Ecosystems and Environment*, vol. 267, no. July, pp. 1–14, 2018, doi: 10.1016/j.agee.2018.07.022.

[16] M. Choi *et al.*, "TopicOnTiles: Tile-based spatio-temporal event analytics via exclusive topic modeling on social media," *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2018-April, pp. 1–11, 2018, doi: 10.1145/3173574.3174157.

[17] D. Jin *et al.*, "A novel generative topic embedding model by introducing network communities," *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, pp. 2886–2892, 2019, doi: 10.1145/3308558.3313623.

[18] Q. Liu, Q. Chen, J. Shen, H. Wu, Y. Sun, and W. K. Ming, "Data analysis and visualization of newspaper articles on thirdhand smoke: A topic modeling approach," *JMIR Medical Informatics*, vol. 7, no. 1, pp. 1–9, 2019, doi: 10.2196/12414.

[19] K. R. Prasad, M. Mohammed, and R. M. Noorullah, "Hybrid topic cluster models for social healthcare data," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 11, pp. 490–506, 2019, doi: 10.14569/IJACSA.2019.0101168.

[20] C. Koylu, "Modeling and visualizing semantic and spatio-temporal evolution of topics in interpersonal communication on Twitter," *International Journal of Geographical Information Science*, vol. 33, no. 4, pp. 805–832, 2019, doi: 10.1080/13658816.2018.1458987.

[21] D. J. Carter and A. Rahmani, "Proximity and Neighbourhood: Using Topic Modelling to Read The Development of Law in the High Court of Australia," *Monash University Law Review*, vol. 45, no. 3, pp. 785–824, 2019.

[22] A. Goswami, P. Mohapatra, and C. Zhai, "Quantifying and visualizing the demand and supply gap from e-commerce search data using topic models," *The Web Conference 2019 - Companion of the World Wide Web Conference, WWW 2019*, pp. 348–353, 2019, doi: 10.1145/3308560.3316605.

[23] S. K. Ray, A. Ahmad, and C. A. Kumar, "Review and Implementation of Topic Modeling in Hindi," *Applied Artificial Intelligence*, vol. 33, no. 11, pp. 979–1007, 2019, doi: 10.1080/08839514.2019.1661576.

[24] B. Zafari and T. Ekin, "Topic modelling for medical prescription fraud and abuse detection," *Journal of the Royal Statistical Society. Series C: Applied Statistics*, vol. 68, no. 3, pp. 751–769, 2019, doi: 10.1111/rssc.12332.

[25] E. S. Negara, D. Triadi, and R. Andryani, "Topic Modelling Twitter Data with Latent Dirichlet Allocation Method," *ICECOS 2019 - 3rd International Conference on Electrical Engineering and Computer Science, Proceeding*, no. October 2019, pp. 386–390, 2019, doi: 10.1109/ICECOS47637.2019.8984523.

[26] M. Asghari, D. Sierra-Sosa, and A. S. Elmaghraby, "A topic modeling framework for spatio-temporal information management," *Information Processing and Management*, vol. 57, no. 6, p. 102340, 2020, doi: 10.1016/j.ipm.2020.102340.

[27] M. Odlum *et al.*, "Application of topic modeling to tweets as the foundation for health disparity research for COVID-19," *Studies in Health Technology and Informatics*, vol. 272, pp. 24–27, 2020, doi: 10.3233/SHTI200484.

[28] G. Yang, A. Ma, Z. S. Qin, and L. Chen, "Application of topic models to a compendium of ChIP-Seq datasets uncovers recurrent transcriptional regulatory modules," *Bioinformatics*, vol. 36, no. 8, pp. 2352–2358, 2020, doi: 10.1093/bioinformatics/btz975.

[29] G. Tao, Y. Miao, and S. Ng, "COVID-19 Topic Modeling and Visualization," *Proceedings of the International Conference on Information Visualisation*, vol. 2020-Septe, no. Iv, pp. 734–739, 2020, doi: 10.1109/IV51561.2020.00129.

[30] Q. Deng, Y. Gao, C. Wang, and H. Zhang, "Detecting information requirements for crisis communication from social media data: An interactive topic modeling approach," *International Journal of Disaster Risk Reduction*, vol. 50, no. January, p. 101692, 2020, doi: 10.1016/j.ijdrr.2020.101692.

[31] L. Juan, Y. Wang, J. Jiang, Q. Yang, G. Wang, and Y. Wang, "Evaluating individual genome similarity with a topic model," *Bioinformatics*, vol. 36, no. 18, pp. 4757–4764, 2020, doi: 10.1093/bioinformatics/btaa583.

[32] Y. Miyata, E. Ishita, F. Yang, M. Yamamoto, A. Iwase, and K. Kurata, "Knowledge structure transition in library and information science: topic modeling and visualization," *Scientometrics*, vol. 125, no. 1, pp. 665–687, 2020, doi: 10.1007/s11192-020-03657-5.

[33] H. Liu, Z. Chen, J. Tang, Y. Zhou, and S. Liu, *Mapping the technology evolution path: a novel model for dynamic topic detection and tracking*, vol. 125, no. 3. Springer International Publishing, 2020.

[34] T. Zhang, B. Lee, Q. Zhu, X. Han, and E. M. Ye, "Multi-Dimension Topic Mining Based on Hierarchical Semantic Graph Model," *IEEE Access*, vol. 8, pp. 64820–64835, 2020, doi: 10.1109/ACCESS.2020.2984352.

[35] D. Buenano-Fernandez, M. Gonzalez, D. Gil, and S. Lujan-Mora, "Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach," *IEEE Access*, vol. 8, pp. 35318–35330, 2020, doi: 10.1109/ACCESS.2020.2974983.

[36] K. S. Cheng, Z. Wang, P. C. Huang, P. Chundi, and M. Song, "TopExplorer: Tool Support for Extracting and Visualizing Topic Models in Bioengineering Text Corpora," *IEEE International Conference on Electro Information Technology*, vol. 2020-July, pp. 334–343, 2020, doi: 10.1109/EIT48999.2020.9208294.

[37] V. Bulatov *et al.*, "TopicNet: Making additive regularisation for topic modelling accessible," *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, no. May, pp. 6745–6752, 2020.

[38] D. Opitz, E. Graells-Garrido, and I. Pérez-Messina, "Toward Characterizing Cities with Social Media Images Using Activity Recognition, Topic Modeling and Visualization," *The Web Conference 2020 - Companion of the World Wide Web Conference, WWW 2020*, pp. 688–693, 2020, doi: 10.1145/3366424.3384361.

[39] S. H. Kim and H. G. Cho, "User-topic modeling for online community analysis," *Applied Sciences (Switzerland)*, vol. 10, no. 10, 2020, doi: 10.3390/APP10103388.

[40] S. Han, S. Ye, and H. Zhang, "Visual exploration of Internet news via sentiment score and topic models," *Computational Visual Media*, vol. 6, no. 3, pp. 333–347, 2020, doi: 10.1007/s41095-020-0178-4.

[41] G. Ertek and L. Kailas, "Analyzing a decade of wind turbine accident news with topic modeling," *Sustainability (Switzerland)*, vol. 13, no. 22, 2021, doi: 10.3390/su132212757.

[42] Zoya Seemab Latif, F. Shafait, and R. Latif, "Analyzing LDA and NMF Topic Models for Urdu Tweets via Automatic Labeling," *IEEE Access*, vol. 9, pp. 127531–127547, 2021, doi: 10.1109/ACCESS.2021.3112620.

[43] J. S. Kim, H. Kim, E. Lee, and Y. Seo, "Analysis of research on metabolic syndrome in cancer survivors using topic modeling and social network analysis," *Science Progress*, vol. 104, no. 4, pp. 1–15, 2021, doi: 10.1177/00368504211061974.

[44] L. Chen, P. Wang, X. Ma, and X. Wang, "Cancer communication and user engagement on Chinese social media: Content analysis and topic modeling study," *Journal of Medical Internet Research*, vol. 23, no. 11, pp. 1–9, 2021, doi: 10.2196/26310.

[45] A. R. Alharbi, M. Hijji, and A. Aljaedi, "Enhancing topic clustering for Arabic security news based on k-means and topic modelling," *IET Networks*, vol. 10, no. 6, pp. 278–294, 2021, doi: 10.1049/ntw2.12017.

[46] B. Zhong and Q. Liu, "Medical insights from posts about irritable bowel syndrome by adolescent patients and their parents: Topic modeling and social network analysis," *Journal of Medical Internet Research*, vol. 23, no. 6, pp. 1–13, 2021, doi: 10.2196/26867.

[47] R. Kumari, J. Y. Jeong, B. H. Lee, K. N. Choi, and K. Choi, "Topic modelling and social network analysis of publications and patents in humanoid robot technology," *Journal of Information Science*, vol. 47, no. 5, pp. 658–676, 2021, doi: 10.1177/0165551519887878.

[48] J. Liu *et al.*, "Tracing the Pace of COVID-19 Research: Topic Modeling and Evolution," *Big Data Research*, vol. 25, p. 100236, 2021, doi: 10.1016/j.bdr.2021.100236.

[49] J. Won *et al.*, "Trends in nursing research on infections: Semantic network analysis and topic modeling," *International Journal of Environmental Research and Public Health*, vol. 18, no. 13, 2021, doi: 10.3390/ijerph18136915.

[50] P. J. M. Noble, C. Appleton, A. D. Radford, and G. Nenadic, "Using topic modelling for unsupervised annotation of electronic health records to identify an outbreak of disease in UK dogs," *PLoS ONE*, vol. 16, no. 12 December, pp. 1–18, 2021, doi: 10.1371/journal.pone.0260402.