COMMUNICATION

Supporting Information Visualization Research in an Academic Library

Lessons Learned from an Analysis of the Literature

Michael Groenendyk and Tomasz Neugebauer

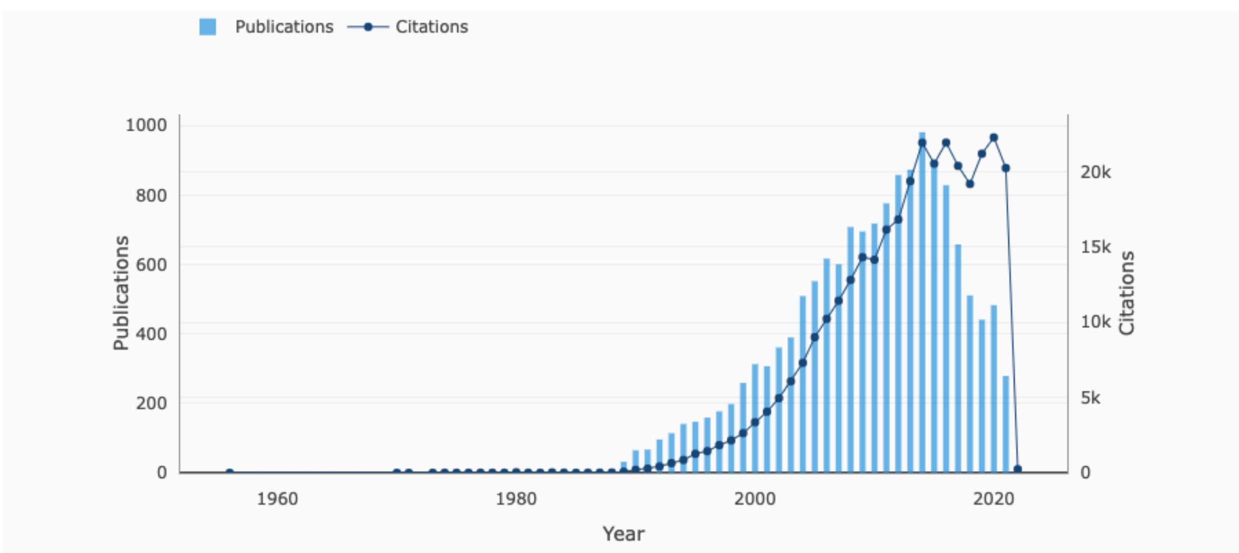
ABSTRACT

This paper summarizes librarian research on information visualization as well as general trends in the broader field, highlighting the most recent trends, important journals, and which subject disciplines are most involved with information visualization. By comparing librarian research to the broader field, the paper identifies opportunities for libraries to improve their information visualization support services.

INTRODUCTION

The technique of creating images to communicate facts is thousands of years old. Early examples include Ptolemy's second-century work *Geographia*, as well as Charles Minard's mid-nineteenth–century flow map, *Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813.* Information visualization as "an independent self-contained research field," however, is a contemporary development that emerged in the 1990s.[1] Microsoft Academic Graph shows the sharp rise of information visualization publications from 1990 on (see fig. 1).[2]

Figure 1. Information visualization publications (total 14,865) and citations (total estimated 330,234) according to Microsoft Academic Search.



About the Authors

Michael Groenendyk (corresponding author: michael.groenendyk@concordia.ca) is Digital Scholarship Librarian, Concordia University (Canada). **Tomasz Neugebauer** (tomasz.neugebauer@concordia.ca) is Digital Projects & Systems Development Librarian, Concordia University (Canada). © 2024.

This peer-reviewed work was submitted on 8 August 2023, accepted for publication on 13 November 2023, and published 18 March 2024.

https://doi.org/10.5860/ital.v43i1.16867

Due to the interdisciplinary nature of information visualization, as well as the variety of visualization methods used by researchers, understanding current research trends can be difficult. While many libraries now offer information visualization support services, little research has been done to determine how researchers exploring and using information visualization techniques can best be supported. By comparing information visualization research in library and information science (LIS)-focused journals to the larger research field, this paper highlights overlaps and differences and identifies opportunities for libraries to improve their information visualization support services.

LITERATURE REVIEW

In the 1990s, due to the enormous amount of information visualization research being published, researchers began to publish surveys summarizing existing literature. Typically, these surveys focused on a specific subject area or visualization method. One of the earliest examples of an information visualization survey is "Interactive High-Dimensional Data Visualization," which summarizes research involving high-dimensional data.[3] Other early examples include "From Visual Data Exploration to Visual Data Mining: A Survey" and "A Survey of Radial Methods for Information Visualization."[4]

The first widely scoped information visualization survey is "A Survey on Information Visualization: Recent Advances and Challenges," which attempts to describe the entirety of information visualization research as well as advances in the field.[5] This survey was followed by another far-reaching survey, "Survey of Surveys,"[6] which summarizes previous information visualization surveys and then classifies existing research into the five categories of data enhancement and transformation, visual mapping and structure, exploration and rendering, interactive analysis, and perception.[7] The survey of surveys notes the percentage of articles that fall into each category. It also lists the top publications and conferences that information visualization research appears in, as well as the academic disciplines most involved with the subject area.

More recent information visualization surveys include a survey of interactive 3D visualizations;[8] a survey of visualization methods used for building information modeling;[9] and a survey of visualization methods used for musical data.[10] Despite the increasing popularity of these surveys, though, so far, no survey of library and information studies (LIS) information visualization research has been undertaken.

Significant LIS information visualization publications include the book *Data Visualization: A Guide to Visual Storytelling for Libraries*, which describes how information visualization can be used for library assessment.[11] The book also provides information on creating visualization with Google Charts, Tableau, Excel, and programming languages such as R and Python.

Another influential LIS publication is the Canadian Association of Research Library's *Data Visualization Toolkit*.[12] The toolkit provides an overview of popular information visualization methods, with a focus on how these methods can be used for library assessment. The toolkit notes that the software programs most often used by librarians for information visualization are Excel, Tableau, Google Data Studio, and Google Charts.

Highly cited LIS articles typically involve library assessment. Examples include "How Data Visualization Supports Academic Library Assessment: Three Examples from The Ohio State University Libraries using Tableau,"[13] "Data Visualization and Rapid Analytics: Applying Tableau

Desktop to Support Library Decision-making,"[14] and "Using Data Visualization to Examine an Academic Library Collection."[15]

To better understand what LIS information visualization research has been done outside of the most influential publications, this paper seeks to summarize all published, LIS-related, information visualization research, highlighting the most popular research topics and the most frequently used visualization methods.

METHODOLOGY

The three major library indexes (Library, Information Science & Technology Abstracts (LISTA), Library & Information Science Abstracts (LISA), and Library Literature & Information Science Full Text) were searched for articles containing the subject keywords "information visualization" OR "data visualization". This search returned 1,631 results. The metadata for these articles was downloaded as a CSV file. Using the data cleaning software OpenRefine, duplicate articles, non-English articles, and non-articles such as editorials were removed. The deduped CSV contained metadata for 1,541 articles.

To compare library research to the broader information visualization field, the Scopus abstract and citation database was searched for peer-reviewed articles also containing the subject keywords "information visualization" OR "data visualization". This search returned 54,680 results. Results were further limited to using Scopus's subject keyword filter (limiting, again, to "information visualization" and "data visualization"), which decreased the number of articles to 47,958. The metadata for the articles was downloaded as a CSV file. Duplicate articles, non-English articles, and non-articles were removed, which left 16,659 articles.

For additional insight into information visualization research trends, the metadata for articles appearing in IEEE Visualization Conferences publications from 1990 to 2020 was downloaded through an open data website.[16]

Article metadata was analyzed using Python and the Python libraries pandas and NLTK (Natural Language Toolkit). The analysis determined the frequency that subject terms, author names, department affiliations, and keywords appeared in the literature.

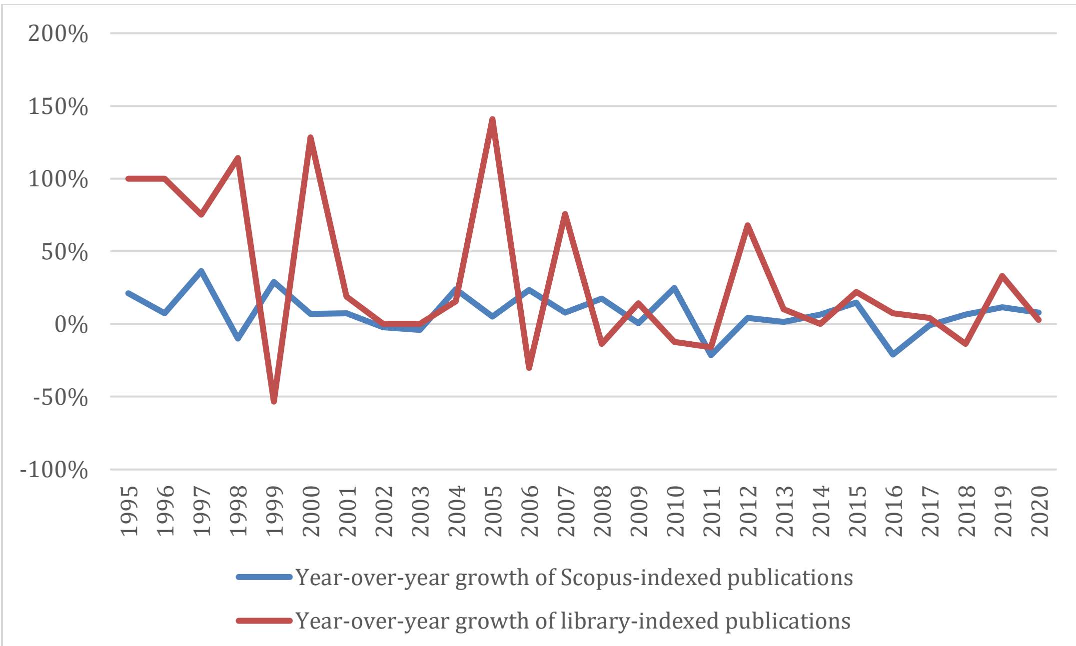
Using a custom Python script,[17] along with DBPedia Spotlight (a named entity detection tool used to identify concepts and the relationship of concepts in text), the IEEE Conference metadata was also analyzed to identify important concepts in article titles and abstracts. The resulting data from the script was the enriched IEEE Conference metadata with WikiData entities to show which concepts are most associated with LIS information visualization literature and which are most associated with IEEE conference publications (with these publications typically representing current, cutting-edge trends).

For the last part of the analysis, an Apache Pig script was used to create a dynamic co-concept graph that was then loaded into a Gephi tool.[18] Using a method previously developed by Neugebauer et al., an interactive visualization was created[19] that showed the association between WikiData concepts and IEEE conference publications. The code used to create these transformations and the visualizations has been made available on GitHub.[20] Appendix A contains a summary of the WikiData entities related to information visualization that were identified through this method and mentioned in the paper.

RESULTS

While the overall number of information visualization articles indexed by Scopus has significantly increased over the past 30 years (from 200 publications a year to 2,000), the year-over-year growth of these publications has remained fairly steady. While the number of library-index publications is much lower (1,541), the number of yearly publications has increased at a much more rapid rate. Year over year, this growth has widely varied too, and there have been years when the number of library-index publications doubled in size but also sharply fell off. (See fig. 2.)

Figure 2. Article publication year-over-year growth percentage (LIS indexed vs Scopus indexed).



Within Scopus, the most significant information visualization journals primarily feature research from authors with engineering and computer science department affiliations, such as *Computer Graphics Forum* and the IEEE publications. Library information visualization research appeared most often in medical and bibliometrics-related publications, as well as library-specific publications such as the *Journal of the American Society for Information Science & Technology.* The top five most popular journals for each index can be seen in table 1 and table 2.

Table 1. Library-indexed journals that publish the most information visualization research

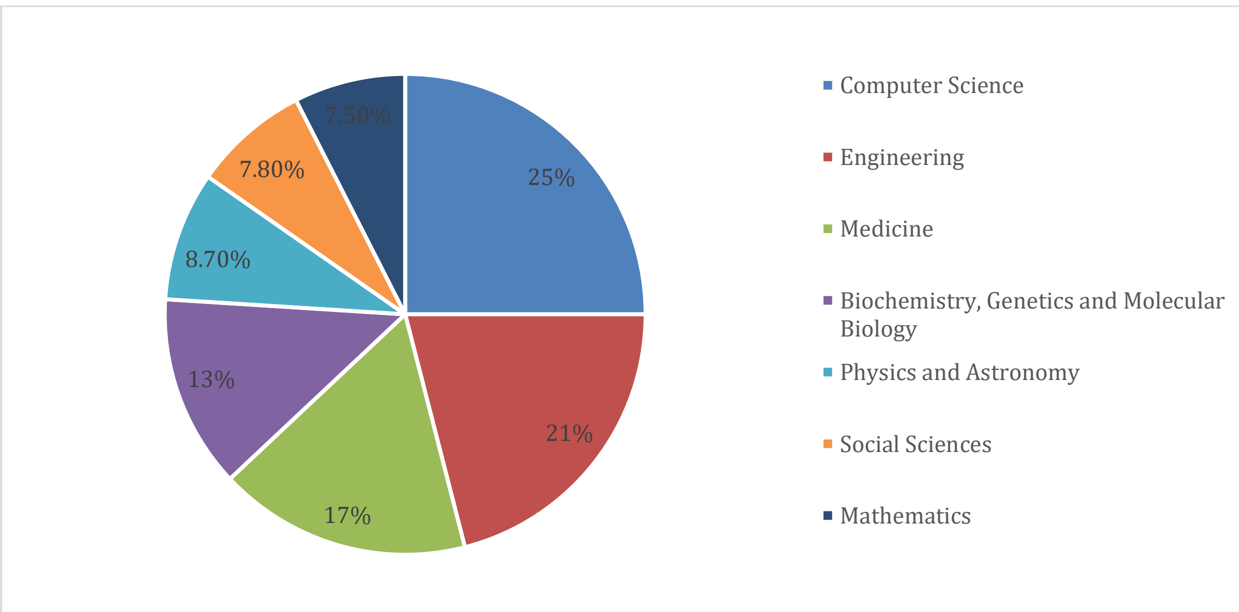
Ranking	Library-indexed journal
1	*Scientometrics*
2	*Journal of Medical Internet Research*
3	*Journal of the American Society for Information Science & Technology*
4	*Journal of the American Medical Informatics Association*
5	*Behaviour & Information Technology*

Table 2. Scopus-indexed journals that publish the most information visualization research

Ranking	Scopus-indexed journal
1	*Computer Graphics Forum*
2	*IEEE Transactions on Visualization and Computer Graphics*
3	*IEEE Computer Graphics and Applications*
4	*Proceedings of the International Conference on Information Visualisation*
5	*Journal of Visualization*

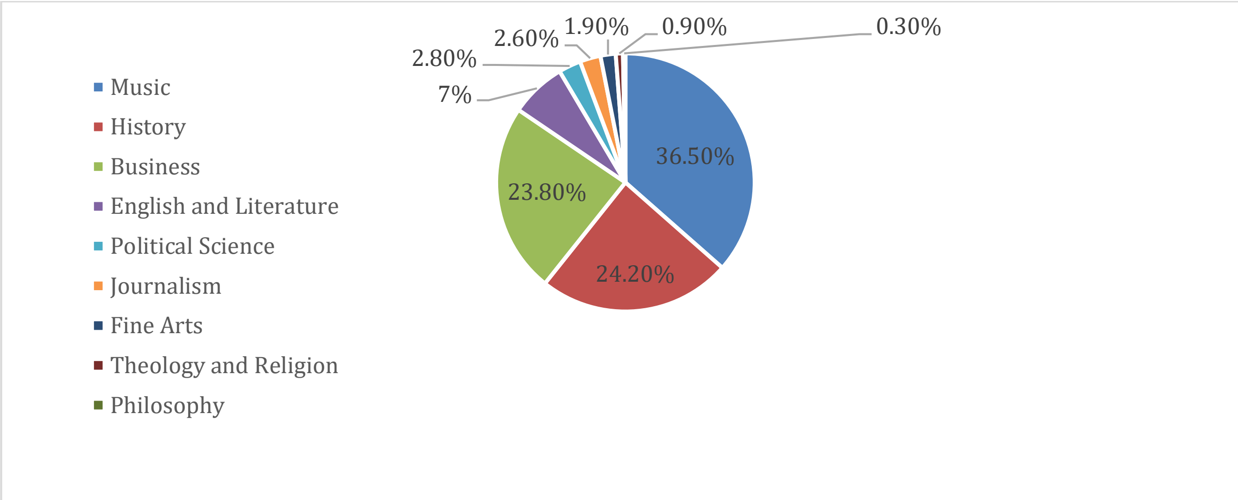
The subject areas that appear most frequently within Scopus-indexed articles (about information visualization) are computer science (25% of all articles), engineering (21%), and medicine (17%). The full subject area breakdown for Scopus indexed articles can be seen in figure 3.

Figure 3. Frequency of subject heading tags for Scopus-indexed articles.



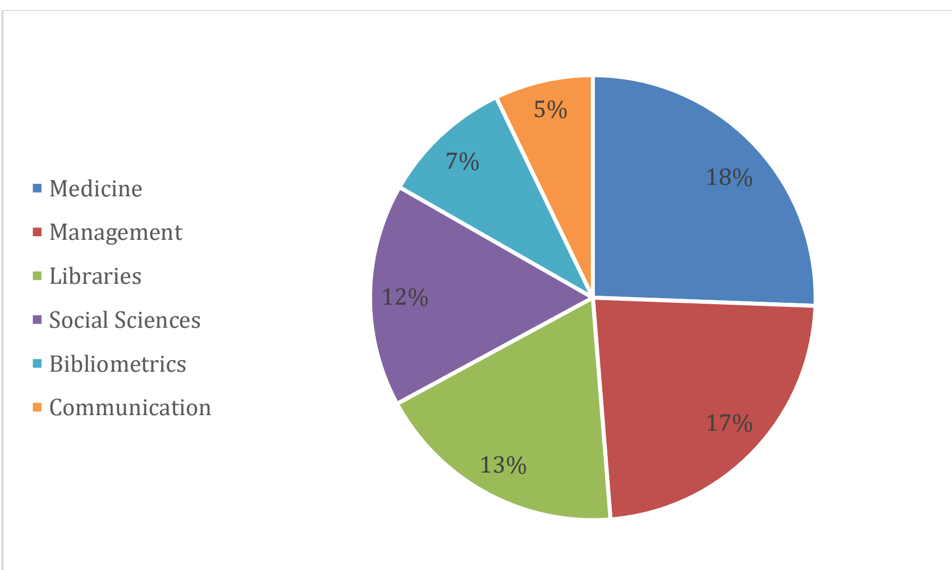
For sub-subjects within Scopus's social science subject heading, music represented 36.5% of all publications, followed by history with 24.2%, and business with 23.8%. The full breakdown of the social science sub-subject areas can be seen in figure 4.

Figure 4. Frequency of social science sub-subject tags for Scopus-indexed articles



The analysis of both Scopus and LIS-indexed article metadata demonstrated medicine's importance within information visualization research. The LIS-indexed metadata showed that most librarians involved with information visualization research are affiliated with medical and science libraries. Medical journals such as the *Journal of Medical Internet Research* and the *Journal of the American Medical Informatics Association* were among the top publishers of library-indexed, information visualization research. Looking at the subject keywords most associated with LIS-indexed articles (about information visualization), "medicine" emerged as the most common, closely followed by "management." The top six subject keywords can be seen in figure 5. References to the information visualization software programs VoxViewer and Tableau were also common. Twenty-one articles (2% of all articles) mentioned Tableau in their title or abstract, while 40 articles (3%) mentioned VosViewer.

Figure 5. Frequency of subject heading tags for LIS-indexed articles.



Within the Scopus-indexed articles (about information visualization), the most frequently used subject keyword was "graphics," which was used to describe 56% of all articles. Math, engineering, computer science, and medical terminology were also popular subject keywords. For library indexed articles, the most popular subject keyword (outside of "information visualization" itself) was "bibliometrics," which described 6% of all articles. Other popular keywords were "bibliographic," "data mining," "citation analysis," "research funding," "medical informatics," and "user interfaces." "User interfaces" is notable as it was the only subject keyword that appeared frequently in both the Scopus and library-indexed metadata.

A DBPedia Spotlight analysis of IEEE metadata makes it possible to get additional insight into the subject content of the conference literature, such as identifying core, foundational concepts for the discipline, with breadth across many conference publications. We identified concepts appearing in at least 40 publications and as being associated with information visualization conference publications published between 1990 and 2020. There are broad multidisciplinary concepts in the list, such as interactive visual analysis, virtual reality, and sensemaking. The largest number of terms are well known traditional concepts in computer graphics such as volume rendering, texture mapping, vector, and scalar fields. In addition, there are concepts related to artificial intelligence, such as machine learning, data mining, and dimensionality reduction.

To better understand the relationship between these concepts, a co-concept map was generated, which allowed us to identify the concepts with the highest betweenness centrality. These are the concepts that most frequently lie on the shortest path between others in the graph, and as such, are core connecting concepts. The central concepts in this sense included multidisciplinary applications of visualization such as in bioinformatics, CT scanning, and computer vision. Knowledge extraction and deep learning appeared in this list of most central concepts as well, underlining the connection between artificial intelligence and information visualization. The co-concept graph was also filtered by time to see what concepts have been present in the literature for a period of 29 years or longer, the time span from 1991 to 2019, inclusive. These long-standing concepts included areas of application, such as physics, geology, fluid mechanics, NASA, and microscopy, as well as foundational concepts from computer science for the field, such as data compression, polynomials, manifolds, parallel computing, k-d tree, and raster graphics.

To identify important new and trending topics in information visualization, a filter was created to show the most centrally connected concepts from publications between 2015 and 2020 that also had a relationship with other concepts of five years or less. The filter resulted in 24 concepts that are relatively new to the research area, but still have a higher betweenness centrality, suggesting that they are simultaneously new and structurally important to the overall field. These concepts included deep learning, formal grammar, convolutional neural networks, artificial intelligence, reinforcement learning, and generative adversarial network.

Within the abstracts of library-focused information visualization articles the most often mentioned keywords were "bibliometrics," "medical," "health," "service," and "citation." Software programs frequently mentioned are Tableau, Voyant, and Voxviewer. While mentions of JSON and Jupyter Notebooks were prevalent in IEEE publications, they were absent from library-indexed articles.

To gain insight into what concepts are most often associated with library information visualization articles, LIS indexed metadata was also analyzed using DBPedia Spotlight. Many of the resulting concepts were the same as those identified by Scopus. These concepts included information retrieval, bibliometrics, and data mining. Unique to LIS publications were the concepts of informatics, scientometrics, natural language processing, social network analysis, and medical imaging.

CONCLUSION

Based on this study's analysis of LIS-indexed, information visualization publications, librarians working within this area have primarily used information visualization for bibliometric and citation analysis research. When creating information visualizations, librarians prefer to use software such as Tableau and Voxviewer, which simplify the visualization process.

The DBPedia concept analysis of Scopus article abstracts confirmed an LIS association with concepts such as scientometrics, co-citation, health informatics, PubMed, and information theory. It also revealed that the information visualization topics LIS researchers are interested in include central topics in the IEEE publications such as text mining, cluster analysis, machine learning, multidimensional scaling, and natural language processing. Interestingly, the words *pandemic* and *coronavirus* appeared frequently in this analysis, showing that LIS information visualization research has found increased relevance during the COVID pandemic. Also of note is the concept of digital humanities, which figures prominently in LIS literature but does not appear in the top terms of either Scopus indexed articles or IEEE conference proceedings.

The analysis of IEEE conference proceedings, which looked at the concepts gaining importance, identified two concepts, deep learning and artificial intelligence, for which libraries could provide better support. The analysis also showed the importance of Project Jupyter to information visualization researchers. JSON is becoming increasingly important too, as are electronic health records and interoperability. Librarians might benefit from paying closer attention to these topics.

Possible bridges between libraries and the broader area of information visualization research are virtual reality and user-computer interfaces. The DBPedia Spotlight analysis showed some overlap between computer science, engineering, and LIS publications through concepts such as machine learning, virtual reality, data mining, and human-computer interaction. Many significant concepts in engineering research, though, such as computational fluid dynamics, polynomials, manifolds, data structures, flow visualization, and computer simulation, have no presence in LIS literature.

As many librarians are actively engaged in information visualization research, libraries might improve their information visualization support services by supporting other librarian researchers. LIS research involving medical/health informatics and geographic information systems demonstrates an opportunity for further collaboration with researchers from these departments. The important role of digital humanities within LIS literature might be an opportunity for libraries to connect digital humanities scholars with researchers in engineering and computer science who are commonly on the forefront of information visualization techniques. Libraries also have a unique opportunity to leverage their information visualization knowledge to showcase information visualization techniques with a broad public interest, such as the visualization of social media data from X (formerly Twitter) and Facebook, the COVID pandemic, or other trending topics.

Though many academic libraries have long embraced information visualization, they have mostly approached the subject from a digital humanities perspective, which, as this paper shows, only represents a sliver of the information visualization field. For libraries to improve their information visualization services, librarians need to become more familiar with research generated by the natural sciences, computer science, and engineering. The collaborative nature of libraries puts them in a strong position to encourage partnerships between researchers from different departments in an increasingly interdisciplinary research area. Encouraging collaboration between information visualization researchers from different departments is one of the most important services any library offering information support and training can provide.

APPENDIX A: DISCUSSED WIKIDATA ENTITIES RELATED TO INFORMATION VISUALIZATION IDENTIFIED USING THE DBPEDIA SPOTLIGHT ANALYSIS

The full dataset includes information visualization WikiData enriched IEEE publications[21] and LIS publications.[22]

WikiData label (English)	WikiData URL
artificial intelligence	https://www.wikidata.org/wiki/Q11660
bibliometrics	https://www.wikidata.org/wiki/Q603441
bioinformatics	https://www.wikidata.org/wiki/Q128570
cluster analysis	https://www.wikidata.org/wiki/Q622825
co-citation	https://www.wikidata.org/wiki/Q1785764
computational fluid dynamics	https://www.wikidata.org/wiki/Q815820
computer simulation	https://www.wikidata.org/wiki/Q925667
computer vision	https://www.wikidata.org/wiki/Q844240
convolutional neural networks	https://www.wikidata.org/wiki/Q17084460
SARS-CoV-2	https://www.wikidata.org/wiki/Q82069695
CT scan	https://www.wikidata.org/wiki/Q32566
data compression	https://www.wikidata.org/wiki/Q2493
data mining	https://www.wikidata.org/wiki/Q172491
data structure	https://www.wikidata.org/wiki/Q175263
deep learning	http://www.wikidata.org/wiki/Q197536
digital humanities	https://www.wikidata.org/wiki/Q1026962
dimensionality reduction	http://www.wikidata.org/wiki/Q16000077
electronic health records	http://www.wikidata.org/wiki/Q10871684
flow visualization	https://www.wikidata.org/wiki/Q5462121
fluid mechanics	https://www.wikidata.org/wiki/Q172145
formal grammar	https://www.wikidata.org/wiki/Q373045
generative adversarial network	https://www.wikidata.org/wiki/Q25104379
geology	https://www.wikidata.org/wiki/Q1069
health informatics	https://www.wikidata.org/wiki/Q870895

WikiData label (English)	**WikiData URL**
human-computer interaction	https://www.wikidata.org/wiki/Q207434
informatics	https://www.wikidata.org/wiki/Q4027615
information retrieval	https://www.wikidata.org/wiki/Q816826
information theory	https://www.wikidata.org/wiki/Q131222
interactive visual analysis	https://www.wikidata.org/wiki/Q17092520
Fast Healthcare Interoperability Resources	https://www.wikidata.org/wiki/Q19597236
JSON	https://www.wikidata.org/wiki/Q2063
Project Jupyter	https://www.wikidata.org/wiki/Q55630549
k-d tree	https://www.wikidata.org/wiki/Q309949
knowledge extraction	https://www.wikidata.org/wiki/Q1582085
machine learning	https://www.wikidata.org/wiki/Q2539
manifold	https://www.wikidata.org/wiki/Q203920
medical imaging	https://www.wikidata.org/wiki/Q931309
microscopy	https://www.wikidata.org/wiki/Q1074953
multidimensional scaling	https://www.wikidata.org/wiki/Q620538
NASA	https://www.wikidata.org/wiki/Q23548
natural language processing	https://www.wikidata.org/wiki/Q30642
pandemic	https://www.wikidata.org/wiki/Q12184
parallel computing	https://www.wikidata.org/wiki/Q232661
physics	https://www.wikidata.org/wiki/Q413
polynomial	https://www.wikidata.org/wiki/Q43260
PubMed	https://www.wikidata.org/wiki/Q180686
raster graphics	https://www.wikidata.org/wiki/Q182270
reinforcement learning	https://www.wikidata.org/wiki/Q830687
scalar field	https://www.wikidata.org/wiki/Q193460
scientometrics	https://www.wikidata.org/wiki/Q472342
sensemaking	https://www.wikidata.org/wiki/Q2063340
social network analysis	https://www.wikidata.org/wiki/Q7551269

WikiData label (English)	**WikiData URL**
text mining	https://www.wikidata.org/wiki/Q676880
texture mapping	https://www.wikidata.org/wiki/Q752176
vector field	https://www.wikidata.org/wiki/Q186247
virtual reality	https://www.wikidata.org/wiki/Q170519
volume rendering	https://www.wikidata.org/wiki/Q2165951

ENDNOTES

[1] Wolfgang Aigner et al., *Visualization of Time-Oriented Data*, Human-Computer Interaction Series (London: Springer, 2011), https://doi.org/10.1007/978-0-85729-079-3.

[2] Microsoft Academic Graph was discontinued as a publicly available resource on December 31, 2021.

[3] Andreas Buja, Dianne Cook, and Deborah F. Swayne, "Interactive High-Dimensional Data Visualization," *Journal of Computational and Graphical Statistics* 5, no. 1 (March 1, 1996): 78–99, https://doi.org/10.1080/10618600.1996.10474696.

[4] M. C. Ferreira de Oliveira and H. Levkowitz, "From Visual Data Exploration to Visual Data Mining: A Survey," *IEEE Transactions on Visualization and Computer Graphics* 9, no. 3 (July 2003): 378–94, https://doi.org/10.1109/TVCG.2003.1207445; Geoffrey M. Draper, Yarden Livnat, and Richard F. Riesenfeld, "A Survey of Radial Methods for Information Visualization," *IEEE Transactions on Visualization and Computer Graphics* 15, no. 5 (September 2009): 759–76, https://doi.org/10.1109/TVCG.2009.23.

[5] Shixia Liu et al., "A Survey on Information Visualization: Recent Advances and Challenges," *The Visual Computer* 30, no. 12 (December 1, 2014): 1373–93, https://doi.org/10.1007/s00371-013-0892-3.

[6] Liam McNabb and Robert S. Laramee, "Survey of Surveys (SoS)—Mapping the Landscape of Survey Papers in Information Visualization," *Computer Graphics Forum* 36, no. 3 (2017): 589–617, https://doi.org/10.1111/cgf.13212.

[7] Stuart Card, Jock Mackinlay, and Ben Shneiderman, *Information Visualization: Using Vision to Think.*, (Morgan Kaufmann: Cambridge, 1999).

[8] Erico de Souza Veriscimo, Joao Luiz Bernardes Junior, and Luciano Antonio Digiampietri, "Evaluating User Experience in 3D Interaction: A Systematic Review," in *XVI Brazilian Symposium on Information Systems*, SBSI'20 (New York, NY, USA: Association for Computing Machinery, 2020), 1–8, https://doi.org/10.1145/3411564.3411640.

[9] Paulo Ivson et al., "A Systematic Review of Visualization in Building Information Modeling," *IEEE Transactions on Visualization and Computer Graphics* 26, no. 10 (October 2020): 3109–27, https://doi.org/10.1109/TVCG.2019.2907583.

[10] R. Khulusi et al., "A Survey on Visualizations for Musical Data," *Computer Graphics Forum* 39, no. 6 (2020): 82–110, https://doi.org/10.1111/cgf.13905.

[11] Lauren Magnuson, *Data Visualization: A Guide to Visual Storytelling for Libraries*, (Rowman & Littlefield: Lanham, 2016).

[12] "CARL Data Visualization Toolkit," *Canadian Association of Research Libraries* (blog), November 21, 2019, https://www.carl-abrc.ca/measuring-impact/carl-data-visualization-toolkit/.

[13] Sarah Anne Murphy, "How Data Visualization Supports Academic Library Assessment: Three Examples from The Ohio State University Libraries Using Tableau," *College & Research Libraries News* 76, no. 9 (2015): 482–86.

[14] Sarah Anne Murphy, "Data Visualization and Rapid Analytics: Applying Tableau Desktop to Support Library Decision-Making," *Journal of Web Librarianship* 7, no. 4 (October 2013): 465–76, https://doi.org/10.1080/19322909.2013.825148.

[15] Jannette L. Finch and Angela R. Flenner, "Using Data Visualization to Examine an Academic Library Collection | Finch | College & Research Libraries," accessed October 27, 2020, https://doi.org/10.5860/crl.77.6.765.

[16] Petra Isenberg et al., "Vispubdata.Org: A Metadata Collection About IEEE Visualization (VIS) Publications," *IEEE Transactions on Visualization and Computer Graphics* 23, no. 9 (September 2017): 2199–2206, https://doi.org/10.1109/TVCG.2016.2615308.

[17] Tomasz Neugebauer, "Photomedia/CitationDataEnrichTransform/get-concepts," Python, December 8, 2021, https://github.com/photomedia/citationDataEnrichTransform/blob/main/get-concepts.py.

[18] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks," *Proceedings of the International AAAI Conference on Web and Social Media* 3, no. 1 (March 19, 2009): 361–62, https://ojs.aaai.org/index.php/ICWSM/article/view/13937.

[19] Tomasz Neugebauer, "Concepts in Information Visualization Conference Publications 1990-2020: Concept-concept Graph for IEEE Information Visualization Conferences," 2021, https://photomedia.ca/visualizations/InfoViz.

[20] Tomasz Neugebauer, "Photomedia/CitationDataEnrichTransform," May 14, 2021, https://github.com/photomedia/citationDataEnrichTransform.

[21] "Information Visualization Conference Publications 1990-2020 Enriched with WikiData Concepts," 2021, https://github.com/photomedia/citationDataEnrichTransform/blob/main/enriched-publications.json.

[22] "LIS Publications on Information Visualization enriched with WikiData Concepts," 2021, https://github.com/photomedia/citationDataEnrichTransform/blob/main/enriched-publications-LIS.json.