




Systematic reviews and tech mining: A methodological comparison with case study

Patricia F. Anderson¹  | Carol Shannon¹  | Skye Bickett² | Joanne Doucette³ | Pamela Herring⁴ | Andrea Kepsel⁵  | Tierney Lyons⁶ | Scott McLachlan⁷ | Lin Wu⁸

¹Taubman Health Sciences Library, University of Michigan, Ann Arbor, Michigan

²PCOM Library, Philadelphia College of Osteopathic Medicine—Georgia Campus, Suwanee, Georgia

³University Libraries, MCPHS University, Boston, Massachusetts

⁴Harriet F. Ginsburg Health Sciences Library, University of Central Florida College of Medicine, Orlando, Florida

⁵Michigan State University Libraries, Michigan State University, East Lansing, Michigan

⁶Cerebros Medical Systems, Jessup, Pennsylvania

⁷Ruskin College, University of Oxford, Oxford, UK

⁸Health Sciences Library, University of Tennessee Health Science Center Library, Memphis, Tennessee

Correspondence

Patricia F. Anderson Taubman Health Sciences Library, University of Michigan, Ann Arbor, MI.

Email: pfa@umich.edu

When the Medical Library Association identified questions critical for the future of the profession, it assigned groups to use systematic reviews to find the answers to these questions. Group 6, whose question was on emerging technologies, recognized early on that the systematic review process would not work well for this question, which looks forward to predict future trends, whereas the systematic review process looks back in time. We searched for new methodologies that were more appropriate to our question, developing a process that combined systematic review, text mining, and visualization techniques. We then discovered tech mining, which is very similar to the process we had created.

In this paper, we describe our research design and compare tech mining and systematic review methodologies. There are similarities and differences in each process: Both use a defined research question, deliberate database selection, careful and iterative search strategy development, broad data collection, and thoughtful data analysis. However, the focus of the research differs significantly, with systematic reviews looking to the past and tech mining mainly to the future.

Our comparison demonstrates that each process can be enhanced from a purposeful consideration of the procedures of the other. Tech mining would benefit from the inclusion of a librarian on their research team and a greater attention to standards and collaboration in the research project. Systematic reviews would gain from the use of tech mining tools to enrich their data analysis and corporate management communication techniques to promote the adoption of their findings.

KEYWORDS

research methodologies, systematic reviews, tech mining

1 | INTRODUCTION

In 2012, the Medical Library Association (MLA) identified a series of critical questions related to the future of the profession and assigned teams to perform systematic

reviews (ideally) to address them.¹ One question focused on identifying emerging technologies relevant to health sciences librarians, which we interpreted to mean technologies not only relevant to the performance of librarians' day-to-day tasks but also relevant to the people

they serve. When the question was assigned initially, MLA recognized that a systematic review might not be the appropriate methodology to answer it. Systematic reviews were developed to discover trends in the health sciences literature that could be used to determine a consensus around core clinical questions. This methodology has less application to questions that are not focused in a historical base in the research literature. Because our question focused on emerging technologies, which are more recent developments, the team quickly recognized that the systematic review methodology was inappropriate for addressing that research question, and we were asked to revise the question and methodology to address the underlying intent of the research question.

To honor the intent of the larger project, which was eventually named the MLA Systematic Review Project,² as we developed our independent methodology, we made every attempt to choose methods and tools that could be replicated by other librarians, were clearly documented, and that utilize open-access or open-source databases and tools. As we explored methodologies during the iterative development process, we discovered tech mining in competitive intelligence. Tech mining, which uses text mining, visualization, and communication tools, tends to be a less transparent and open process but is designed to be more forward looking. Following our discovery of tech mining, the team focused on a more intentional integration of those methods into the project, which now uses a blended methods approach, selecting elements from both systematic review and tech mining methodologies to create a custom approach to fit the project's audience, goals, and timeline.

To avoid confusion in our comparison, definitions of the most important concepts are listed below. For the purposes of this paper, it is particularly important to distinguish between *tech mining*, a competitive intelligence methodology, and *text mining*, a collection of textual data analysis tools and strategies.

Systematic reviews: "A review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review. Statistical methods (**meta-analysis**) may or may not be used to analyse and summarise the results of the included studies."³

Tech mining: A methodology developed in engineering and corporate research and development for forecasting areas of strategic importance. It uses text mining, visualization, and communication tools to harness "information about emerging technologies to inform technology management."⁴

Text mining: "Text data mining involves combing through a text document or resource to get valuable structured information. This requires sophisticated analytical tools that process text in order to glean specific keywords or key data points from what are considered relatively raw or unstructured formats. Text data mining is also known as text mining or text analytics."⁵

As context for these concepts, it is also helpful to remember the broader context of emerging technologies environmental scanning, of which tech mining is a part. Tech mining is one of the most crisply defined methodologies in this area⁶ and uses techniques closely related to systematic review methods. Other terms and methodologies used for these activities include competitive technical intelligence, disruptive or emerging technologies forecasting or foresight, forecasting and foresight technologies, horizon scanning, monitoring for anticipation, patent landscape reports, persistent forecasting, prediction horizon, technology futures or monitoring or roadmapping, and others.⁷⁻¹³

2 | CASE STUDY: IDENTIFYING EMERGING TECHNOLOGIES RELEVANT TO HEALTH SCIENCES LIBRARIANS

This is the first article of the project, focusing on methodologic context. We intend to write subsequent articles on phases 1 to 3.

Phase 1. Preparation

The team began this project by scanning articles to identify existing search strategies on emerging technologies in the health sciences. We selected the Varela-Lema et al search strategy as closest to the project goals. The team began by brainstorming to discover additional refinements of the assigned question and terms and concept gathering. This process was expanded through a separate modified Delphi and survey. These techniques served to clarify the scope of the project and identify additional relevant terms and concepts. The pool of collected terms and concepts next went through a clustering process, which elicited four subclusters.

Phase 2. Searches

To ensure the validity of the identified concept clusters, the team began by creating new search strategies for each of the concept groups. The second step was to

test the Varela-Lema et al search strategy, validate terms, and expand the strategy to include Medical Subject Headings (MeSH) and relevant journal titles. The resulting emerging technologies strategy was tested in combination with the concept cluster search strategies. At this point, the team created two additional variants of the base search strategy, one more sensitive and one more specific.

Phase 3. Analysis

The team ran the base search strategy itself in PubMed and exported the results for data analysis. To create a CSV file, we imported the PubMed search results into EndNote, deduplicated them, and then exported the results using a custom filter. The CSV file was imported into Excel for later data cleaning. Data cleaning and analysis were performed in a collection of open-access text mining tools.

3 | PHASE 1

3.1 | Question development

In 2008, the MLA Research Section conducted an initial Delphi study to develop the most important research questions for the profession.¹⁴ In 2011, the Research Agenda Committee of the Research Section completed a second Delphi study to refresh MLA's research agenda. It identified 15 questions to be studied by teams and answered with systematic reviews. The committee published the study in July 2012.¹ Team Six was formed from an international group of MLA members and assigned the following question:

*The explosion of information, expanding of technology (especially mobile technology), and complexity of healthcare environment present medical librarians and medical libraries opportunities and challenges. To live up with the opportunities and challenges, what kinds of skill sets or information structure do medical librarians or medical libraries are required to have or acquire so as to be strong partners or contributors of continuing effectiveness to the changing environment?*¹

To focus our research efforts, the team narrowed and restructured the scope of the question to (a) What are the emerging technologies and technological tools of greatest interest to health sciences librarians for their own professional work? and (b) How can health sciences librarians methodically identify emerging technologies and trends

important to their stakeholders and clients in a systematic and replicable way?

3.2 | Audience

The primary target audience for the results of this analysis is health sciences librarians. However, the study needed to identify not only emerging technologies relevant to the daily tasks and duties of librarians but also those relevant to those individuals served by health sciences libraries. Additionally, the term *health sciences librarian* covers a range of positions, from outreach librarians who support the general public to hospital librarians who provide direct support to clinicians and patients and academic health sciences librarians, who work with clinical researchers and bioinformatics specialists. Each area has its own focus, and what is important to one group might be merely a curiosity to another. Health sciences librarians not only need to know of new technologies to facilitate the work that they do but also those about which they might reasonably be asked for more information or which might develop as potential collaborations between the library and its stakeholders: clinicians, researchers, students, patients, caregivers, and the general public.

3.3 | Topic and scope extraction

The team generated terms for the search strategy in three stages. In the initial brainstorming phase, the project team utilized Mindmeister,¹⁵ an online collaborative mind mapping program,^{16,17} to identify tools and technologies of interest to health sciences librarians. Next, using a modified Delphi method, the project team initiated a search process, eliciting information from a combination of internal and external communities through survey and focus groups. This Delphi study was performed through two Twitter-based focus group chats. The participants were stakeholders: One group was composed of health sciences librarians and the second group of self-identified health care leaders, such as clinicians, patient advocates, hospital CEOs, and others. These two steps generated technology and topic lists, which were the basis of the online survey that followed. The survey asked self-identified stakeholders to list relevant emerging technologies and the resources they used to track or discover new trends. After the survey and focus groups were completed, members of the team reviewed the technology and topic lists that were generated, clustering these terms into related concept groups.

This process generated four concept clusters: communication and education from librarians for stakeholders, technologies used on or in the human body, technologies

used in health care delivery and public health, and technologies used in libraries and librarianship for professional tasks and products. Subgroups of the team worked on each category, generating additional terms within those focused areas and identifying technologies that should be retrieved in a well-developed search strategy on emerging technologies. The resulting expertise fed into the relevancy testing and validation of a base search strategy, which served as the foundation in combination of more specific concept groups and for refining for different levels of sensitivity and specificity, as needed.

Having realized that a traditional systematic review was inappropriate to discover trends in emerging technologies, we revised and combined methods to preserve the best of what was appropriate from systematic review methodologies, while incorporating useful elements from tech mining techniques. Systematic review methodologies look back in time, compiling the best of the research on a narrow topic. Emerging technologies, however, combine a breadth of scope with the need to forecast the future, not the past. These very different temporal foci largely distinguish the two methodologies examined in this paper.

4 | PHASE 2

4.1 | Search strategy development

Because this is a preliminary study proposing and testing a novel methodology, the team chose to pilot the strategy in only one database. We selected PubMed in part because it was created and is maintained by the US National Library of Medicine (NLM), is freely available to the world, and is widely used by both health sciences librarians and their stakeholders.

Beginning with the validation of a PubMed search strategy for emerging technologies published by Varela-Lema et al,¹⁸ the team initially identified search concept groups as *Emerging* combined with *Technologies*. Following our testing, the concept groups were modified to ["Emerging Technologies" OR (Emerging AND Technologies)], with the other terms from the original search modified substantially through free-text term additions, deletions, expansions, corrections, and the addition of MeSH where appropriate. The validation process included testing terms individually for sensitivity and specificity, then repeating the testing as the concept clusters were developed, refining at each step of the process and defining criteria for the inclusion or exclusion of terms.

The team developed a search strategy to explicitly identify and include emerging technologies in the health sciences. A problem with only searching structured

vocabulary, such as MeSH, for emerging technologies terms is that once the database producer identifies and adds these subject terms to a literature database, the technologies might no longer be emerging: When a concept is emerging, there is no structured vocabulary for it. Using unstructured vocabulary helps to address that issue, but only retrieves articles where the authors intentionally identify the novelty of their work in the title or abstract. To support the discovery of unknown terms, concepts, or technologies that are emerging, we identified relevant journals for keyword searching.

We selected specific journal titles with a high specificity for the topic by searching the NLM catalog using relevant terms from the subgroups and emerging technologies base search. The team scanned the tables of contents and abstracts of the prior three issues or years for inclusion and exclusion criteria, which were identified for the level of specificity required for terms and titles. The search results were limited to 1 year (January-December 2016) to constrain the size of the data file for this proof-of-concept analysis. They were also limited to English language to help avoid the challenges of trying to apply text mining across multiple languages whose terms may not be equivalent.

The team developed an initial search strategy for 3D printing,¹⁹ one of eight search strategies ultimately created for this question. In the future, we hope to expand the dataset for analysis to 3 years and also to validate the other seven search strategies (one for each of the four concept groups, plus three variations on each). We will create both focused and broad strategies to accommodate the needs of specific topic clusters within emerging technologies that we have identified as subsets of interest to our stakeholder populations.

5 | PHASE 3

5.1 | Data export, cleaning, and analysis

In systematic reviews, data have multiple layers: You begin with the full record and then move on to the full-text of the article, but the initial analysis is performed on a subset of the citation, usually the title and abstract. In tech mining, the complete record is downloaded and subdivided to accommodate different types of analyses. In our project, we used elements of both processes, but they were simplified to make the project feasible within the resources available to us. We downloaded the complete records but immediately set them aside, following extraction of the fields we intended to analyze: title, abstract, keywords, and MeSH terms. That limitation meant that we were not doing certain portions of the standard tech mining analysis, such as coauthorship

patterns, impacts, knowledge transfer analysis, and trend analysis. The portion we used did make it possible for us to do the cluster analysis and the core sources analysis, which we then integrated into the development of the search strategy.

5.1.1 | Data export

After the team ran the final search strategy in PubMed, we exported the resulting list in the MEDLINE format to create a TXT file. In our initial attempts at data analysis, we discovered that no one on the team had access to a computer large enough to run the analysis on 3 years of data, so we limited this preliminary analysis to a single year.

To support the proposed text mining analysis of this dataset using Voyant or OpenRefine, we needed to build a CSV file. Initially, we used FLink,²⁰ an NLM-supported tool for generating CSV files from PubMed search results. Because the program could handle only 10 000 records and could not produce a CSV file with abstracts (which were essential to our analysis), the team next turned to EndNote. We exported the PubMed search results as a TXT file, imported them to EndNote, deduplicated the records, and then exported them as a CSV file, using a custom output style that we created.

5.1.2 | Data cleaning

The resulting 162 339 records were downloaded to Excel, and all fields except PMID, Title, Abstract, and MeSH or Keywords were deleted. The remaining content was cleaned by removing punctuation (using nested SUBSTITUTE functions) and changing all text to lowercase (using the LOWER function). During the punctuation removal process, the team also considered topics such as the separation of MeSH and subheadings by removal of the “/” characters, the importance of “.” characters in numerical values, and whether or not to keep numeric data.

In systematic reviews, data cleaning is primarily a deduplication of the records in the dataset. In tech mining, data cleaning is considerably more complicated to support a very different type of data analysis. This stage is where the methods for our project began to diverge substantially from classic systematic review methodologies. For the analysis of this dataset and to support the extraction of key concepts for strategic planning, the team adopted a modified and limited analysis that is more related to tech mining. Our goal was to make our methodology transparent, available, and modifiable by future researchers, as well as to support the concept of open intelligence systems.²¹ Although there are commercial

tools, such as Web of Science or VantagePoint, we chose only analytical tools that are similarly open access and have a low barrier to entry for librarians for this step of the process.

This philosophy meant that we would make minimal use of tools that require programming skills, such as R, GREP, and Python, which would have been the expected choices. The tools that we selected for our preliminary data analysis were Voyant Tools²² for collocation and stop list refinement to identify words or concept clusters within the data; OpenRefine²³ to expand the data analysis through data cleaning and refinement; and AntConc²⁴ using collocation, Corpus: Keyword in Context, and word frequency to dive deep into the specific terms or phrases of interest. Voyant is web based, requires that the dataset is open to the public, and offers a wide range of very visual tools for identifying patterns in textual data, thus making it possible to read and analyze texts in new ways. OpenRefine helps with messy data: to clean it; to change it from one form to another, if needed; and to combine it with other data. Using these tools in sequence allowed us to discover trends in one tool for deeper analysis or exploration in another tool.

5.1.3 | Analysis

The fields that we chose to analyze included structured vocabulary (MeSH) and unstructured vocabulary (title, abstract, keywords). Searching these fields in a literature database is a way to overcome the lack of a structured vocabulary for emerging technology concepts and to address database delays in cataloging, while at the same time allowing for the possibility of identifying unexpected patterns in structured vocabulary use.

6 | DISCUSSION

6.1 | Methodologic comparison

In this project, the similarities of systematic reviews and tech mining strengthened it; however, we needed to be aware of the key differences. These primarily fell into the areas of target audience, data sources, and typical analysis. These distinctions are detailed in Table 1.

For both methodologies, the audience of the resulting analyses are decision makers. For systematic reviews, the decision makers may range from frontline clinicians to hospital administrators and funders. The audiences for tech mining follow a structure similar to that of corporate R&D, but with a stronger focus on funding and strategy. Data sources are selected, and the inclusion criteria used reflect those audiences and purposes, with tech mining pulling data largely from the patent and engineering

TABLE 1 Comparative overview of systematic review and tech mining methods

Systematic Review	Tech Mining
Primary purposes	
To scan, synthesize, and distill large quantities of information for key stakeholders in an efficient, well-documented, replicable way.	To scan, synthesize, and distill large quantities of information for key stakeholders in an efficient way.
To do so in a way that eliminates or minimizes bias, and supports effective decision making to reduce harm (physical, psychological, or economic) to health care participants and/or organizations.	To do so in a way that supports effective decision making to maximize innovation, competitive intelligence, and competitive advantage for organizations.
To utilize the information discovered to identify information trends and patterns, information gaps, and research gaps in the evidence base for the purpose of directing future research.	To utilize the information discovered to identify information trends and patterns for the purpose of identifying organizational opportunities, needs, and directing future research.
Methodology guidance and standards	
Question/topic: PICO/T (O'Connor et al, 2011; Becker and Oxman, 2011) Reporting: PRISMA Reporting: MOOSE AHRQ: Methods guide Cochrane handbook for systematic reviews of interventions (Higgins & Green, 2011) Cochrane: MECIR (methodological expectations) Recommendations: GRADE Recommendations: EPICOT Others	Porter & Cunningham, 2005
Registries of topics, protocols, reviews	
Cochrane collaboration Campbell collaboration PROSPERO	
Methods and process	
Topic identification, including audience, motivating challenge, goals/outcomes, timeline (O'Connor et al, 2011; Becker & Oxman 2011)	Topic identification, including audience, motivating challenge, goals/outcomes, timeline
Inclusion/exclusion criteria (O'Connor et al, 2011)	
Query formulation	Query formulation
Search strategy: Draft, design, test, validate, refine, revise, update, repeat (O'Connor et al, 2011)	Search strategy: draft, design, test, validate, refine, revise
Select databases and information sources (Becker, Oxman 2011)	Select databases and information sources
Determine endpoint (iteration and redundancy in results)	Determine endpoint (principal components analysis [PCA])
Export data from finalized search	Export data from finalized search
Deduplication of dataset	
	Data cleaning: deduplication, clustering, irrelevancy, text mining, and other automated techniques to support these
Analysis	
Select studies for inclusion (Higgins & Deeks, 2011b)	
Extract data from included studies (Higgins, et al, 2011b, 7.6.3)	
Evaluate risk of bias (Higgins et al, 2011b, 7.6.2)	
Assess quality of evidence (Higgins et al, 2011a, 8.5.1)	
Interpret and summarize results (Schünemann et al, 2011a & 2011b)	Inductive analysis (identify patterns in the data, keywords, clusters, hypothesis generation)

(Continues)

TABLE 1 (Continued)

Systematic Review	Tech Mining
	Deductive analysis (validation, hypothesis refinement, investigation)
	Social network analysis (coauthor networks, cocitation analysis, related bibliometric analyses)
	Modeling (qualitative, quantitative, stochastic, probability)
	Trends (time series description, forecasting, growth models, identifying novelty)
	Analytic techniques (selected): <ul style="list-style-type: none"> • Bibliometric analysis • Hierarchical clustering • Semantic network analysis • Mapping of technology evolution mathematical modeling • Statistical analysis • Semantic TRIZ • Principal component analysis (PCA) • Subject-action-object (SAO) modeling • Ontological modeling • Web scraping • Ontology modeling • Advanced bibliometrics • Semantic the theory of inventive problem solving (TRIZ) • Sentiment analysis
Dissemination and follow-up	
Present/publish findings	Present/publish findings
Typical presentation of findings: Conference poster Conference presentation Peer-reviewed article	Typical presentation of findings: Internal or informal Action points Alerts Benchmarks Gap analysis Innovation indicators Management presentation <i>One pagers</i> Scorecard Technology/competitor profiles Technology roadmap Visualizations External or formal Conference poster Conference presentation Peer-reviewed article Others
Incorporate into clinical guidelines, as appropriate	
Reassess at predetermined intervals	
Repeat study when changes in evidence base warrant an update	Analyze findings in context of competitors, chart comparisons
	Assess utilization of findings in decision support, resource allocation, research and development directions, and more.
	Assess validity (internal, statistical conclusion, construct, external)

(Continues)

TABLE 1 (Continued)

Systematic Review	Tech Mining
Primary data sources	
Topical databases MEDLINE/PubMed EMBASE Cochrane databases CINAHL Other health bibliographic databases	Topical databases Patent databases (USPTO, EPO, JPO, etc) Engineering bibliographic databases (INSPEC, COMPENDEX, etc) General purpose bibliographic databases (ISI WOS, Scopus, etc)
Grey literature and related resources Conference proceedings Clinical trial registries Grey literature resources	Grey literature and related resources Web data R&D (expenditures) data Social media (networks, blogs) data Geospatial data Commercial, market, and Business data Annual reports and internal documents Funding and awards data
Additional content derived from initial dataset Hand searching Citation mining of textbooks, sentinel articles, and related sources.	Additional content derived from initial data set
Others	Others
Target audiences	
Clinicians Decision makers Health care advocates Health care executives Insurers and health care payers Information professionals and librarians Managers Patients and caregivers Policy makers Researchers Risk assessment and compliance Students Others	Strategic planners R&D managers Researchers, inventors, project managers New product developers and designers Procurement Process managers Product managers Product service managers Marketing experts Information professionals and librarians IP managers and specialists Others

and technical databases,^{4,10,25-27} while systematic reviews extract data from a variety of health, life science, and social science databases, depending on the specific topic.^{28,29} Both of these bibliometric methods appropriately focus on data sources that match the domain of the question and target audience.

A key difference in the data selection process is that in systematic reviews, the researcher attempts to derive a consensus from the data, while in tech mining, the focus is on forecasting and scanning to identify emerging topics and trends. This distinction is made explicit in the actual data, which is derived differently from the search results for the two methodologies. For systematic reviews, the search results provide a collection of citations, which are assessed individually, ultimately leading to a smaller

selection of articles. In tech mining, the search results are processed very differently, with selected fields extracted and merged into a text corpus for the text mining analysis. The fields selected for data analysis in tech mining will vary depending upon the type of analysis being performed. For this preliminary study, we selected the title, abstract, keywords, and MeSH terms from journal article records. A more robust tech mining analysis would have also included the authors and journal titles.

Time limits for systematic reviews tend to be long, going back to the initial publication of the concept being studied in the research literature and including a comprehensive stream of data. In tech mining, the timeframe for the data is more limited, representing a sampling of years that may or may not be contiguous, and focusing on

newer content, with defined constraints similar to those of rapid reviews,³⁰ yet with even more restrictions.⁴ For data analysis, systematic reviews have very clearly defined protocols that have evolved into guidelines for both analysis and presentation. A great deal of the work falls on the shoulders of the expert reviewers, who manually examine each title and select articles to meet the inclusion and exclusion criteria. In tech mining, the analysis is vastly different, depending instead on an automated analysis with a variety of text mining tools and other instruments to identify significant patterns.

The greatest methodological similarities come early in the process for each: how the topics of interest are identified, questions are formulated, and the focus on enrichment and diversity of terminology is used in searching. The approach to developing search strategies is somewhat different between the two, as are the standards related to the process. For systematic reviews, most likely because of the focus on health care and the governmental and legal impacts on the use of the findings, there has been a substantial development of formal standards in the methodology. This process has resulted in a focus on bias reduction, replicability of the study, and mandatory inclusion of librarians and information professionals in the process stages most closely tied to their areas of

expertise: question formulation and search strategy development. Information professionals and librarians are also frequently included in other aspects of systematic reviews and are expected to be included in the project team. In contrast, the tech mining methodology may or may not include librarians and information professionals on their research teams, even though Porter and Cunningham recommend this in their textbook⁴ and in their later publications and presentations on tech mining.

In this project, the team's overall approach was to utilize the basics from systematic review methodologies, with some influence from tech mining for the development of the search for a health care-specific audience, while shifting to the tech mining approach exclusively for the data extraction and analysis. The patterns of overlap from the two approaches is shown in Table 2.

6.2 | Findings

While we will report the results from this project with a more complete analysis later, our work thus far has provided some intriguing preliminary findings, indicating that this fusion of methodologies may be useful to others.

The groupings extracted from our initial, modified Delphi study were organized in five main areas of

TABLE 2 Case study: combining elements of systematic review and tech mining methods

SR	TM	
Methods and Process		
X	X	Topic identification, including audience, motivating challenge, goals/outcomes, timeline
	X	Topic clustering
X		Inclusion/exclusion criteria
X	X	Query formulation
—	—	Select terms, concepts, journals according to replicable inclusion/exclusion criteria as well as sensitivity/specificity test results, assessing irrelevancy
X	X	Search strategy: draft, design, test, validate, refine, revise, update, repeat
X	X	Select databases and information sources
X	X	Determine endpoint
X	X	Export data from finalized search
X		Deduplication of dataset
	X	Data cleaning (field elimination, stop word sets to improve relevancy, etc)
Analysis		
	X	Inductive analysis (text mining to identify patterns in the data, keywords, clusters, hypothesis generation)
	X	Trends (time series description, forecasting, growth models, identifying novelty)
Primary data sources		
X		MEDLINE/PubMed
Primary target audiences		
X		Clinicians; information professionals and librarians; patients and caregivers; researchers; students

Abbreviations: SR, systematic review methods; TM, tech mining methods. Dash, not directly taken from either SR or TM methods.

interest: the four concept clusters previously mentioned (information, public health, education, and the body), with technology as the primary search concept to be combined with the others. The next step, text mining analysis, made clear that there were additional clusters of content, especially new methodologies (such as big data and data visualization) and emerging interdisciplinary trends (such as precision medicine). As we had not discovered that content through the original process, this shows the potential benefit of text mining for unearthing unknown areas of relevance. A complete tech mining analysis may reveal even more. The analysis was otherwise in concordance with the clusters from the Delphi study, which served to validate both methods.

The primary areas of the body that were strongly represented in the data included blood* (7424), bone (4704), brain* (5645), and urin* (3101). Related concepts that were strongly represented in the dataset included cancer* (10 100), diagnostic* (6433), treatment (12 189), and biomark* (6377). The three top technologies that arose from the text mining process were robot* (2773), simulat* (8383), and 3D* technologies (6229), especially 3D print* (3604). All three were being used most heavily in surgery. Simulations were also prominent in education and training. While this degree of detail is still fairly high level and broad, it also makes clear the potential for the use of these methods, especially in larger dataset, to identify more detail than could be realistically or replicably identified for our topic through a traditional systematic review method or our modified Delphi study.

7 | CONCLUSION

Systematic reviews and tech mining feature a number of similarities in that they both require specific question creation processes, database selection, search strategy development, data collection from the search results, and analysis of that data according to the recommended practices for the specific methodology. Systematic reviews focus on historic trends, while tech mining instead skims the recent past and focuses on the future. Thus, while there is overlap in the methodologies, some variation arises depending on whether the desired outcome is past trends or future predictions.

The skills needed for systematic reviews are similar to those for tech mining, and librarians experienced with the former can make valuable contributions to the latter. Both methodologies involve searching multiple resources, resulting in large bodies of information that must be analyzed for their applicability to the research question at hand. Much as a librarian should be involved with designing and executing search strategies for systematic reviews,

they should also be consulted for assistance with tech mining and ideally be an active member of the research team.

The cultural context of the communities in which these project teams operate has also had significant impact on the evolution of these two methodologies. The level of confidentiality around R&D may have served to minimize the inclusion of engineering librarians in tech mining projects. While health sciences librarians have expertise with patient privacy and HIPAA regulations comparable to R&D, the focus of their work has been in extracting information patterns from the current and past literature rather than forecasting. There is much to learn from each of these methods, and the strengths of each can empower the other. Tech mining experts should explore standards, increase collaboration, and better integrate librarians and information professionals into their teams. Systematic review experts should explore the tools and techniques used in tech mining to extract even richer information from their dataset and use corporate management communication techniques to expand the influence and adoption of their findings.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

ORCID

Patricia F. Anderson  <http://orcid.org/0000-0001-6348-2324>

Carol Shannon  <http://orcid.org/0000-0003-3193-6963>

Andrea Kepsel  <http://orcid.org/0000-0002-9651-6700>

REFERENCES

1. Eldredge JD, Ascher MT, Holmes HN, Harris MR. The new Medical Library Association research agenda: final results from a three-phase Delphi study. *J Med Libr Assoc.* 2012;100(3):214-218.
2. Eldredge JD, Holmes H, Ascher MT. MLA research agenda. Systematic Review Project. Team Updates Presentation. MLA Annual Meeting Supplement. May 17, 2015. 2015; Austin, TX.
3. Cochrane Community. Cochrane Glossary. <http://community.cochrane.org/glossary#letter-S>. Accessed December 16, 2017.
4. Porter AL, Cunningham SW. *Tech Mining: Exploiting New Technologies for Competitive Advantage*. Hoboken, NJ: John Wiley & Sons, Inc.; 2005.
5. Janalta Interactive I. Technopedia. <https://www.techopedia.com/definition/13859/text-data-mining>. Accessed December 16, 2017.
6. Kayser V, Blind K. Extending the knowledge base of foresight: the contribution of text mining. *Technol Forecast Soc Chang.* 2017;116:208-215.

7. Brockley EM. *Emerging Technology Intelligence: Scanning and Monitoring for Strategic Planning*. Cambridge, MA: Management of Technology Program, Massachusetts Institute of Technology; 2004.
8. Firat A, Woon W, Madnick S. *Technological Forecasting—A Review (MIT Working Paper CISL #2008-15)*. Cambridge, MA: Massachusetts Institute of Technology, Composite Information Systems Laboratory (CISL); 2008.
9. Loveridge D. *The Art and Science of Anticipating the Future*. London: Routledge; 2008.
10. Mikova N, Solokova A. Selection of information sources for identifying technology trends: a comparative analysis. (HSE working papers WP BRP 25/STI/2014). Moscow: National Research University Higher School of Economics; 2014.
11. Sun F, Schoelles K. AHRQ Health Care Horizon Scanning System: A Systematic Review of Methods for Health Care Technology Horizon Scanning (AHRQ Publication No. 13-EHC104-EF). Rockville, MD: Agency for Healthcare Research and Quality; 2013.
12. Sylak-Glassman EJ, Williams SR, Gupta N. *Current and Potential Use of Technology Forecasting Tools in the Federal Government. (IDA Document D-5735)*. Alexandria, VA: Institute for Defense Analyses; 2016.
13. Zhang Y, Zhou X, Porter AL, Vicente Gomila JM. How to combine term clumping and technology roadmapping for newly emerging science & technology competitive intelligence: “problem & solution” pattern based semantic TRIZ tool and case study. *Scientometrics*. 2014;101(2):1375-1389.
14. Eldredge JD, Harris MR, Ascher MT. Defining the Medical Library Association research agenda: methodology and final results from a consensus process. *J Med Libr Assoc: JMLA*. 2009;97(3):178-185.
15. Hollauf M. MindMeister Wins “Best User Centered Design” Award. *Mindmeister FOCUS Blog*. 2007;2018. <https://www.mindmeister.com/blog/product-news/2007/07/02/mindmeister-wins-best-user-centered-design-award/>
16. Anjomshoaa A, Sao KV, Tjoa AM, Weippl E, Hollauf M. Context Oriented Analysis of Web 2.0 Social Network Contents - MindMeister Use-Case. In: Nguyen NT, Le MT, Swiatek J, eds. *Asian Conference on Intelligent Information and Database Systems (ACIIDS 2010)*. Berlin: Springer; 2010.
17. Michinov N. The Use of Electronic Brainstorming for Collecting Ideas in Scientific Research Teams: A Challenge for Future Online Research. In: Nunes Silva C, ed. *Online Research Methods in Urban Planning Studies: Design and Outcomes*. Hershey, PA: IGI Global; 2012:157-172.
18. Varela-Lema L, Punal-Riobóo J, Acción BC, Ruano-Ravina A, García ML. Making processes reliable: a validated pubmed search strategy for identifying new or emerging technologies. *Int J Technol Assess Health Care*. 2012;28(4):452-459.
19. Anderson PF, Bickett S, Doucette J, et al. *Identifying Emerging Technologies Through Text-Mining of PubMed Search Results: New Potential for Extracting Meaning*. Seattle: Medical Library Association Annual Meeting; 2017.
20. NCBI. FLink: How to Save PubMed Search Results as a Comma-Separated Value (CSV) File and Import Them Into a Spreadsheet; 2016.
21. Patton KM. The role of scanning in open intelligence systems. *Technol Forecast Soc Chang*. 2005;72(9):1082-1093.
22. Sinclair S, Rockwell G. Voyant Tools Team. Voyant Tools [web application]. 2012.
23. OpenRefine [computer program]. Version 2.7. 2017
24. AntConc [computer program]. Version 3.4.4. Tokyo: Waseda University; 2016.
25. Hopkins MM, Siepel J. Just how difficult can it be counting up R&D funding for emerging technologies (and is tech mining with proxy measures going to be any better)? *Tech Anal Strat Manag*. 2013;25(6):655-685.
26. Rodríguez-Salvador M, Rio-Belver RM, Garechana-Anacabe G. Scientometric and patentometric analyses to determine the knowledge landscape in innovative technologies: the case of 3D bioprinting. *PLoS ONE*. 2017;12(6):e0180375. <https://doi.org/10.1371/journal.pone.0180375>
27. Vicente-Gomila JM, Palli A, de la Calle B, Artacho MA, Jimenez S. Discovering shifts in competitive strategies in probiotics, accelerated with TechMining. *Scientometrics*. 2017;111(3):1907-1923.
28. Lefebvre C, Manheimer E, Glanville J, on behalf of the Cochrane Information Retrieval Methods Group. Searching for Studies. In: Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. London: The Cochrane Collaboration; 2011.
29. Vassar M, Yerokhin V, Sinnett PM, et al. Database selection in systematic reviews: an insight through clinical neurology. *Health Inf Libr J*. 2017;34(2):156-164.
30. Tricco AC, Antony J, Zarin W, et al. A scoping review of rapid review methods. *BMC Med*. 2015;13(1):224.

How to cite this article: Anderson PF, Shannon C, Bickett S, et al. Systematic reviews and tech mining: A methodological comparison with case study. *Res Syn Meth*. 2018;9:540–550. <https://doi.org/10.1002/jrsm.1318>