

## **Quick Understanding Our Engineering Faculty Research Needs Using Topic Modeling**

**Ms. Qianjin Zhang, University of Iowa**

Qianjin (Marina) Zhang is the Engineering & Informatics Librarian at the Lichtenberger Engineering Library, The University of Iowa. As a subject librarian, she manages collection and provides instruction, reference and consultation services for the engineering faculty and students. Her work also focuses on data management education and outreach to engineering students through presenting Data Management topic to an Engineering Ethics course and library workshops. She holds a MA in Information Resources & Library Science from The University of Arizona (Tucson, AZ), and a BS in Biotechnology from Jiangsu University of Science and Technology (Zhenjiang, China).

**Ms. Kari Kozak, University of Iowa**

Kari Kozak is the Head of the Lichtenberger Engineering Library at The University of Iowa. She provides instruction, reference, and consultation services to student, faculty, and staff within the departments and research centers in the College of Engineering as well as the Department of Computer Science. Kari holds bachelor's degrees in Meteorology and Environmental Studies from Iowa State and a master's degree in Library Science from the University of North Carolina – Chapel Hill. Before coming to the University of Iowa in November of 2008, she worked at Texas A & M University as a Science & Engineering Librarian.

## **Quick Understanding Our Engineering Faculty Research Needs Using Topic Modeling**

### **Abstract**

As engineering librarians, we recognize that understanding our faculty research needs is an ongoing endeavor. It is a continuing learning process throughout our time serving engineering faculty with diverse research interests. However, the time-intensive learning process may not efficiently help engineering librarians quickly develop an overall view of the changing and evolving departments. It's also challenging for early-career librarians who are new to engineering librarianship or do not have relevant subject background.

In order to tackle the problem, the authors explored research topics of our faculty's work using a topic modeling technique called Latent Dirichlet Allocation (LDA) which is a type of statistical topic model and a machine learning algorithm for discovering the research topics from text data. We retrieved thousands of bibliographic records of faculty publications as the text data, particularly for the title, abstract and keywords, from Web of Science, removed duplicates and cleaned up the data. Next, we ran the data through the LDA model. The model sorted the data into several groups of related words forming our research topics. As a result, we determined the optimal research topic number of 25 and interpreted the research topics based on the visualization of the LDA results.

In conclusion, our experiment with the LDA approach helped us quickly develop an understanding of faculty research interests, would provide good evidence from which to make decisions on collection management, reference and library instruction, and show the possibility of academic libraries to make use of data and data science techniques in the era of big data.

### **Introduction**

Liaison librarians face the challenge of learning faculty research and teaching needs in a timely manner. Wood and Griffin gave an overview of the current approaches including website analysis, interview, course syllabus analysis and large-scale surveys [1]. Department websites, especially faculty online profiles, are great sources for research interests, courses taught and awarded grants. Interviewing faculty members is also a great way to learn more about faculty research and teaching needs while building faculty-librarian relationships. Although course syllabi can provide a great snapshot of course information and types of assignments and even reveal information-seeking behavior which faculty require for their students, accessing course syllabi may need permission and assistance from the departments [1]. Large-scale surveys of faculty may be a great option as they can provide rich information on faculty's perspectives. For example, the University of Iowa participated in the Ithaca S+R Faculty Survey in 2015 and the survey results informed us of our faculty's perspectives and preferences on the library services and collections [2]. But the drawback of large-scale surveys is the impossibility of gathering continual data to reflect the evolving needs in addition to typical low response rates and problems obtaining representative samples [1]. The current approaches may be good enough if we have sufficient time, staffing and solid faculty-librarian relationships. However, in reality, adoption of the current approaches may be difficult for liaison librarians who serve a large number of faculty members and have many other job duties.

As the University of Iowa College of Engineering has experienced significant growth in research, we two engineering librarians see the necessity of analyzing faculty research needs. Since we serve about one hundred faculty members and are fully occupied with many other duties, it will take tremendous time and effort to conduct surveys, interviews or even review faculty online profiles. Instead of using the current approaches, we explored a topic modeling technique called Latent Dirichlet Allocation (LDA) as an

alternative approach to extract research topics from the bibliographic records of faculty publications. LDA is a type of statistical topic modeling and a machine learning algorithm for discovering topics from documents. Basically, LDA assumes each document as a mixture of topics that split out words with certain probabilities and then automatically generate potential topics based on the assumption. Although topic modeling is widely used in data science, there is limited literature about adopting such a technique to solve library-relevant problems. Several use cases were found. Fang and her team successfully applied topic modeling to discovering research topics and trends in six top accounting journals [3]. Zuo and Zhao adopted topic modeling to extract research topics from faculty publications when measuring three multidisciplinary areas [4]. Cain also discussed some potential applications of topic modeling for processing library and archival collections [5]. Inspired by the previous studies, we saw the possibility of applying topic modeling to solving our current dilemmas.

The contributions of this paper were two-fold. First, we applied topic modeling as an alternative approach to rapidly understanding our engineering faculty research needs. Second, we showed the possibility of academic libraries to make use of data and data science techniques to better serve our patrons in the era of big data.

## Methods

First of all, we gathered the bibliographic records of our engineering faculty research papers for the past 20 years, between 1998 and 2018, from Web of Science. We searched the Address field by entering department abbreviations as shown in Table 1 and retrieved a total of 5641 bibliographic records.

Table 1. Searching the Address Field in Web of Science

Full Department Name	Search Strategy
Biomedical Engineering	univ iowa SAME biomed* eng*
Chemical and Biochemical Engineering	univ iowa SAME chem* & biochem* eng*
Civil and Environmental Engineering	univ iowa SAME civil* & environ* eng*
Electrical and Computer Engineering	univ iowa SAME elect* & comp* eng*
Mechanical and Industrial Engineering	univ iowa SAME mech* & ind* eng*
IIHR Hydrosience and Engineering	univ iowa SAME iihr hydrosci* & eng*
Center for Computer Aided Design	univ iowa SAME ccad

Since we defined research papers as journal articles and conference proceedings, we limited the document types to “articles” and “proceedings paper” and the languages to “English”. We also removed 619 duplicates and excluded 123 records that did not have an abstract because abstracts contained important and descriptive information about research. Duplicates were managed using EndNote reference management software. As a result, we obtained a total of 4899 bibliographic records as text data.

In the second step of data cleaning, we removed publishers’ copyright information in the abstract using regular expressions because they were distracting and irrelevant to research topics. We also combined the title, abstract and keywords of each record into a single document, resulting in a total of 4899 documents that were also called word sets. But the cleaned text data was not ready for the LDA processing.

Table 2. Extended Stopwords

---

method, methods, problem, problems, paper, papers, propose, proposes, proposed, show, shows, showed, present, presents, presented, result, results, resulting, resulted, suggest, suggested, suggests, study, studied, studies, test, tested, tests, conclusion, conclusions, conclude, concluded, concludes, concluding, significantly, significant, science, sciences, engineering, approach, approaches, approaching, approached, report, reports, reported, publishing, publisher, publishes, published, model, models, modeling
--

---

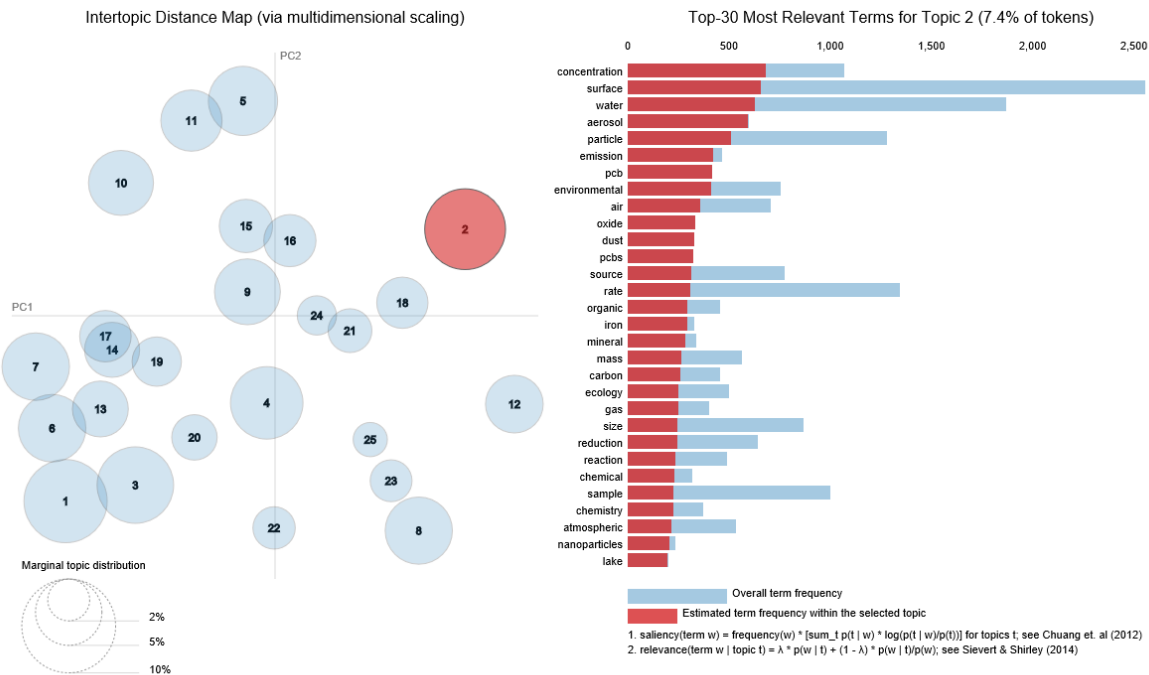
The third step was to preprocess the text data including tokenizing each sentence into words, removing punctuation, numbers and stopwords, and normalizing words. Stopwords were frequently occurring words but meaningless to research topics, as shown in Table 2. When normalizing words, we could do either stemming (reducing words to their root form) or lemmatization (changing words in third person to first person or verbs in past and future tenses into present tense) [6]. We chose lemmatization rather than stemming because lemmatization would provide semantic meanings for various formats of words. After preprocessing, each document was broken into a word set, resulting in a total of 4899 word sets.

The last step was to set an appropriate value for the number of research topics as a parameter and the LDA model sorted the total of 4899 word sets. The LDA would process each word set through pulling words to cluster particular words that were more likely to appear together. The groups of particular words were interchangeable with research topics in this paper. Since the optimal research topic number was not yet determined, the LDA was run 20 times with different values ranged from 15 to 34 for the number of research topics. Therefore, the LDA results with different research topic numbers were generated and saved for further examination. We used gensim LDA packages for this study and adopted Zuo’s Python codes for the steps of preprocessing and model building [7, 8].

### Results and Analysis

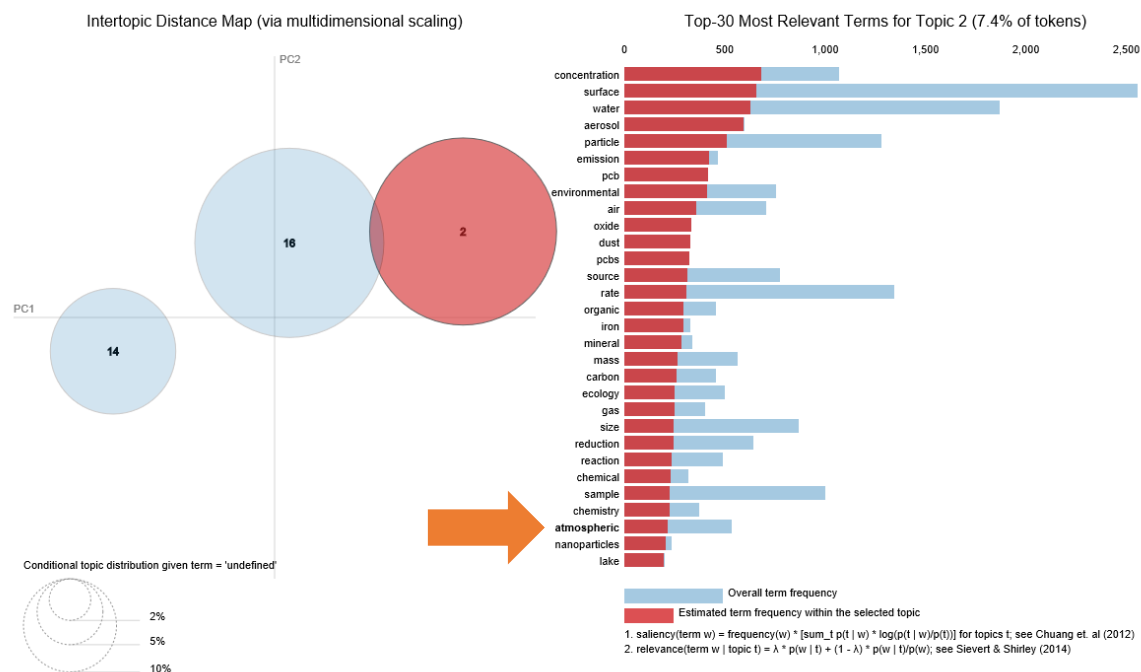
The LDA generated different research topics where each topic was a combination of words and their relative probabilities to that topic, as shown in Appendix 1, Column Top 30 Words. In order to better interpret the research topics and find the optimal research topic number, we adopted pyLDavis, an interactive LDA visualization python package, to create an overall view of the LDA results on how the research topics related to each other while allowing for an examination on how the most relevant words were associated with each individual topic.

Figure 1. Research Topic Visualization using pyLDavis



The topic visualization was composed of two panels as shown in Figure 1. The left panel of the visualization presented an entire view of the topic model. In this view, the areas of the circles represented an overall prevalence of each research topic and the research topics were sorted in decreasing order of prevalence. The distance between the centers of the circles indicated how the research topics related to each other. Correspondingly, the right panel of the visualization was a histogram of the top 30 most relevant words occurring in the highlighted research topic on the left panel. In addition, the visualization allowed for exploring topic-word relationships. To be specific, when the word “atmospheric” on the right panel was highlighted in Figure 2, the left panel revealed how the word was distributed across the research topics. The word “atmospheric” could be found in the topics of toxicology in environment (Topic 2 in Appendix 1), precipitation analysis (Topic 14 in Appendix 1) and climate change (Topic 16 in Appendix 1).

Figure 2. The Conditional Distribution of One Chosen Word over Multiple Research Topics



In the examining phase, we compared the visualization of the LDA results with different research topic numbers. As the research topic number gradually increased from 15 to 34, new research topics appeared, some ambiguous research topics became clear and the results tended to be consistent. We chose 25 as the optimal research topic number because the results were consistent and the visualized LDA results were well clustered with fewer overlapping circles as shown in Figure 1. A total of 25 interpreted topics were listed in Appendix 1, Column Topic.

## Discussion

Using the LDA approach, our study identified a total of 25 research topics from our faculty publications for the past 20 years. We noticed that some research topics were clearly distinguished from each other because those research areas had been in existence for several years, resulting in a large number of publications. For example, the topic of toxicology in environment (Topic 2 in Appendix 1) reflected a

large research program on air pollutants that had been running for twelve years. The results also gave us an insight into some research areas which we had never recognized. An example would be the topic of lung imaging (Topic 3 in Appendix 1) which was aligned with a research laboratory studying pulmonary imaging. Since many words occurring in the topics were technical jargons, we referred to the bibliographic records and figured out the meanings. For example, the word of “oct”, as shown in Top 30 Words for Topic 1 in Appendix 1, was an abbreviation for “optical coherence tomography” imaging rather than “October”, revealing a cross-disciplinary research in such biomedical imaging area. As we gradually learned the high frequency words, we began to create a great understanding of the research projects that our faculty were undertaking.

Admittedly, our study had some limitations. First, we used the bibliographic records (the title, abstract and keywords) of faculty publications and excluded low-quality records that did not have an abstract, which might introduce bias to the text data for analysis. We might get better results if we could use full text of faculty publications rather than the bibliographic records. However, it would introduce new difficulties such as acquiring PDF files and converting PDF files to plain text files. Second, with using only one database (Web of Science) to pull all the bibliographic records, this may not be an all-inclusive list of every item published by the faculty. The nature of some cross-disciplines is also a limitation. It might partially account for the difficulty in interpretation for some research topics, especially a few slightly overlapping topics. Furthermore, this approach may miss capturing some real-time research interests because of scholarly publication delays and dramatic changes within the university or college. Faculty may have moved on from what they previously completed while their papers are under editorial and publishing processes. University and college initiatives can quickly change research foci and new faculty may bring new research interests to the university and college.

## Conclusion

We implemented the LDA approach to identify a total of 25 research topics from the bibliographic records of faculty publications for the past 20 years. The study results helped us develop a quick understanding of our faculty research needs and would provide good evidence from which to make decisions on collection management, reference and library instruction. Our experiment with the LDA approach also showed the possibility of academic libraries, especially for liaison librarians, to make use of data and data science techniques to better serve the patrons at large. In addition, although our study focused on engineering faculty research needs, the LDA approach could definitely be applied to solving problems in understanding faculty research and teaching needs in many other disciplines, analyzing library data and beyond. For example, the LDA approach could be used to abstract research topics from theses and dissertations, which would discover the research needs of graduate students. The other example would be using LDA approach to analyze library online chat transcripts to identify frequently asked questions. Finally, we were optimistic about topic modeling as a powerful tool to demystify large volumes of text data and its potential in tackling the problems confronting academic librarians.

## References

- [1] N. B. Wood and M. Griffin, "Liaison Librarians in the Know: Methods for Discovering Faculty Research and Teaching Needs," in *Proceedings of the Charleston Library Conference*, Charleston, SC, 2016, pp. 326-333.
- [2] Ithaka S+R, "University of Iowa Faculty Survey: Report of Findings," 2015.
- [3] D. B. Fang, H. X. Yang, B. J. Gao, and X. J. Li, "Discovering research topics from library electronic references using latent Dirichlet allocation," *Library Hi Tech*, Article vol. 36, no. 3, pp. 400-410, 2018.

- [4] Z. Y. Zuo and K. Zhao, "The more multidisciplinary the better? - The prevalence and interdisciplinarity of research collaborations in multidisciplinary institutions," *Journal of Informetrics*, vol. 12, no. 3, pp. 736-756, Aug 2018.
- [5] J. O. Cain, "Using Topic Modeling to Enhance Access to Library Digital Collections," *Journal of Web Librarianship*, Article vol. 10, no. 3, pp. 210-225, 2016.
- [6] S. Li. (2018). *Topic Modeling and Latent Dirichlet Allocation (LDA) in Python*. Available: <https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24>
- [7] Z. Y. Zuo. (2017). *Topic Model Preprocessor*. Available: <https://github.com/zhiyzuo/python-topic-model-preprocessor>
- [8] Z. Y. Zuo. (2018). *Text Analytics in Python*. Available: <https://github.com/zhiyzuo/python-tutorial/blob/master/7-Text-Analytics-in-Python.ipynb>

### Appendix 1. A Total of 25 Research Topics Generated by the LDA Model

No.	Topic	Top 30 Words
1	Optical Coherence Tomography (OCT) Imaging	0.040*"image" + 0.028*"segmentation" + 0.011*"base" + 0.010*"surface" + 0.009*"analysis" + 0.009*"graph" + 0.009*"retinal" + 0.009*"automate" + 0.008*"vessel" + 0.008*"feature" + 0.008*"oct" + 0.007*"shape" + 0.007*"algorithm" + 0.007*"layer" + 0.006*"detection" + 0.006*"segment" + 0.006*"object" + 0.006*"thickness" + 0.006*"tomography" + 0.005*"set" + 0.005*"tree" + 0.005*"optical" + 0.005*"computer" + 0.005*"region" + 0.005*"volume" + 0.005*"coronary" + 0.004*"ultrasound" + 0.004*"patient" + 0.004*"medical" + 0.004*"scan"
2	Toxicology in Environment	0.015*"concentration" + 0.014*"surface" + 0.013*"water" + 0.013*"aerosol" + 0.011*"particle" + 0.009*"emission" + 0.009*"pcb" + 0.009*"environmental" + 0.008*"air" + 0.007*"oxide" + 0.007*"dust" + 0.007*"pcbs" + 0.007*"source" + 0.007*"rate" + 0.006*"organic" + 0.006*"iron" + 0.006*"mineral" + 0.006*"mass" + 0.006*"carbon" + 0.005*"ecology" + 0.005*"gas" + 0.005*"size" + 0.005*"reduction" + 0.005*"reaction" + 0.005*"chemical" + 0.005*"sample" + 0.005*"chemistry" + 0.005*"atmospheric" + 0.004*"nanoparticles" + 0.004*"lake"
3	Lung Imaging	0.031*"image" + 0.028*"lung" + 0.015*"airway" + 0.014*"registration" + 0.012*"compute" + 0.011*"tomography" + 0.011*"pulmonary" + 0.010*"volume" + 0.009*"base" + 0.007*"ventilation" + 0.007*"scan" + 0.006*"regional" + 0.006*"measure" + 0.006*"subject" + 0.005*"difference" + 0.005*"dose" + 0.005*"tissue" + 0.005*"motion" + 0.005*"imaging" + 0.005*"compare" + 0.005*"disease" + 0.004*"function" + 0.004*"human" + 0.004*"medical" + 0.004*"inverse" + 0.004*"high" + 0.004*"quantitative" + 0.004*"average" + 0.004*"reconstruction" + 0.004*"respiratory"
4	Biomechanics	0.016*"stress" + 0.010*"motion" + 0.010*"load" + 0.010*"patient" + 0.009*"contact" + 0.008*"spine" + 0.007*"human" + 0.006*"muscle" + 0.006*"aneurysm" + 0.006*"force" + 0.006*"analysis" + 0.006*"property" + 0.006*"joint" + 0.006*"element" + 0.006*"increase" + 0.006*"mechanical" + 0.005*"wall" + 0.005*"screw" + 0.005*"pressure" + 0.005*"spinal" + 0.005*"level" + 0.005*"tissue" + 0.005*"biomechanical" + 0.004*"fracture" + 0.004*"finite" + 0.004*"lumbar" + 0.004*"low" + 0.004*"measure" + 0.004*"effect" + 0.004*"compare"
5	Hydrology and Water Resources	0.055*"flow" + 0.016*"velocity" + 0.014*"bed" + 0.014*"structure" + 0.013*"channel" + 0.011*"large" + 0.011*"simulation" + 0.010*"water" + 0.010*"transport" + 0.009*"layer" + 0.009*"sediment" + 0.009*"shear" + 0.008*"scale" + 0.008*"eddy" + 0.008*"turbulence" + 0.008*"effect" + 0.007*"dimensional" + 0.007*"field" + 0.006*"turbulent" + 0.006*"numerical" + 0.006*"river" + 0.005*"scour" + 0.005*"particle" + 0.005*"condition" + 0.005*"depth" + 0.005*"region" + 0.005*"stream" + 0.005*"wall" + 0.005*"vortex" + 0.005*"distribution"
6	Wind Turbine	0.021*"data" + 0.021*"algorithm" + 0.011*"time" + 0.011*"driver" + 0.009*"wind" + 0.009*"image" + 0.009*"matrix" + 0.008*"performance" + 0.008*"low" + 0.006*"recovery" + 0.006*"learn" + 0.006*"sample" + 0.006*"drive" + 0.006*"signal" + 0.006*"high" + 0.005*"scheme" + 0.005*"prediction" + 0.005*"mri" + 0.005*"rank" + 0.005*"machine" + 0.005*"reconstruction" + 0.005*"structure" + 0.005*"cluster" + 0.004*"sense" + 0.004*"power" + 0.004*"turbine" + 0.004*"technology" + 0.004*"sparse" + 0.004*"speed" + 0.004*"measurement"
7	Reliability-Based Design	0.031*"design" + 0.027*"optimization" + 0.020*"base" + 0.018*"system" + 0.018*"reliability" + 0.015*"analysis" + 0.012*"variable" + 0.012*"sensitivity" + 0.011*"data" + 0.011*"process" + 0.010*"energy" + 0.009*"function" + 0.009*"input" + 0.009*"performance" + 0.007*"probability" +

	Optimization (RBDO)	0.007*"algorithm" + 0.007*"develop" + 0.007*"failure" + 0.007*"rbdo" + 0.006*"random" + 0.006*"point" + 0.006*"constraint" + 0.005*"sample" + 0.005*"distribution" + 0.005*"stochastic" + 0.005*"numerical" + 0.005*"uncertainty" + 0.005*"measure" + 0.005*"simulation" + 0.005*"control"
8	Gene Therapy	0.054*"cell" + 0.018*"gene" + 0.012*"expression" + 0.012*"human" + 0.011*"protein" + 0.006*"mouse" + 0.006*"sequence" + 0.006*"tissue" + 0.006*"culture" + 0.006*"growth" + 0.005*"exon" + 0.005*"cancer" + 0.005*"biology" + 0.005*"genome" + 0.005*"splice" + 0.004*"function" + 0.004*"disease" + 0.004*"airway" + 0.004*"molecular" + 0.004*"type" + 0.004*"delivery" + 0.004*"specific" + 0.004*"rna" + 0.004*"increase" + 0.004*"endothelial" + 0.004*"vitro" + 0.004*"analysis" + 0.004*"express" + 0.004*"fibrosis" + 0.004*"tumor"
9	Flood and Water Management	0.021*"flood" + 0.018*"water" + 0.016*"river" + 0.011*"iowa" + 0.010*"event" + 0.010*"scale" + 0.009*"soil" + 0.008*"watershed" + 0.007*"resource" + 0.007*"discharge" + 0.007*"system" + 0.007*"peak" + 0.007*"area" + 0.006*"time" + 0.006*"basin" + 0.006*"network" + 0.006*"data" + 0.006*"stream" + 0.006*"state" + 0.006*"management" + 0.005*"process" + 0.005*"rainfall" + 0.005*"environmental" + 0.005*"nitrate" + 0.005*"year" + 0.005*"increase" + 0.005*"flow" + 0.005*"hydrologic" + 0.005*"runoff" + 0.005*"drainage"
10	Fluid Dynamics	0.018*"particle" + 0.015*"interface" + 0.014*"simulation" + 0.014*"equation" + 0.013*"fluid" + 0.013*"element" + 0.012*"dynamic" + 0.012*"flow" + 0.010*"boundary" + 0.009*"finite" + 0.009*"material" + 0.009*"phase" + 0.009*"interaction" + 0.008*"numerical" + 0.008*"mesh" + 0.008*"grid" + 0.007*"couple" + 0.007*"structure" + 0.007*"set" + 0.007*"level" + 0.006*"computational" + 0.006*"valve" + 0.006*"field" + 0.006*"base" + 0.006*"analysis" + 0.006*"mechanic" + 0.006*"solid" + 0.005*"surface" + 0.005*"formulation" + 0.005*"force"
11	Hydraulics	0.019*"wave" + 0.015*"surface" + 0.015*"simulation" + 0.011*"flow" + 0.010*"ship" + 0.009*"large" + 0.009*"vortex" + 0.008*"effect" + 0.008*"cylinder" + 0.008*"number" + 0.008*"high" + 0.008*"free" + 0.008*"resistance" + 0.008*"experimental" + 0.008*"wake" + 0.007*"dynamic" + 0.007*"current" + 0.006*"water" + 0.006*"motion" + 0.006*"speed" + 0.006*"scale" + 0.006*"grid" + 0.006*"cfd" + 0.005*"data" + 0.005*"interaction" + 0.005*"fluid" + 0.005*"reynolds" + 0.005*"case" + 0.005*"force" + 0.005*"hull"
12	Biodegradation	0.011*"plant" + 0.011*"heparin" + 0.010*"activity" + 0.009*"degradation" + 0.009*"sulfate" + 0.008*"bacteria" + 0.008*"product" + 0.008*"acid" + 0.007*"poplar" + 0.007*"transformation" + 0.006*"strain" + 0.006*"rdx" + 0.005*"enzyme" + 0.005*"metabolite" + 0.005*"reaction" + 0.005*"pseudomonas" + 0.005*"biodegradation" + 0.005*"culture" + 0.004*"environmental" + 0.004*"microbial" + 0.004*"compound" + 0.004*"high" + 0.004*"bind" + 0.004*"soil" + 0.004*"degrade" + 0.004*"tissue" + 0.004*"microbiology" + 0.004*"ecology" + 0.004*"molecular" + 0.004*"contaminate"
13	Circuit Design	0.055*"fault" + 0.026*"circuit" + 0.019*"sequence" + 0.018*"set" + 0.017*"scan" + 0.016*"procedure" + 0.015*"generation" + 0.013*"base" + 0.013*"detection" + 0.013*"path" + 0.012*"delay" + 0.011*"vector" + 0.010*"input" + 0.010*"time" + 0.009*"number" + 0.009*"detect" + 0.008*"target" + 0.008*"transition" + 0.008*"compaction" + 0.008*"sequential" + 0.008*"coverage" + 0.007*"reduce" + 0.007*"order" + 0.007*"demonstrate" + 0.006*"application" + 0.006*"computer" + 0.006*"experimental" + 0.006*"state" + 0.006*"pattern" + 0.006*"functional"
14	Precipitation Analysis	0.029*"rainfall" + 0.027*"data" + 0.022*"radar" + 0.019*"error" + 0.016*"measurement" + 0.016*"estimate" + 0.011*"uncertainty" + 0.011*"scale" + 0.011*"spatial" + 0.010*"base" + 0.009*"rain" + 0.009*"product" + 0.008*"sample" + 0.008*"precipitation" + 0.007*"bias" + 0.007*"estimation" + 0.007*"gauge" + 0.007*"time" + 0.007*"resolution" + 0.007*"field" + 0.007*"analysis" + 0.006*"calibration" + 0.006*"observation" + 0.006*"variability" + 0.005*"network" + 0.005*"temporal" + 0.005*"author" + 0.005*"distribution" + 0.005*"satellite" + 0.005*"reflectivity"
15	Mechanic Properties	0.020*"material" + 0.011*"crack" + 0.010*"growth" + 0.010*"alloy" + 0.008*"thermal" + 0.008*"deformation" + 0.008*"steel" + 0.008*"surface" + 0.008*"fatigue" + 0.008*"simulation" + 0.008*"mechanical" + 0.008*"fracture" + 0.008*"strain" + 0.008*"field" + 0.007*"temperature" + 0.007*"porosity" + 0.007*"metal" + 0.007*"laser" + 0.007*"stress" + 0.007*"damage" + 0.007*"effect" + 0.007*"cast" + 0.007*"analysis" + 0.006*"solidification" + 0.006*"property" + 0.006*"process" + 0.006*"grain" + 0.006*"heat" + 0.005*"load" + 0.005*"develop"
16	Climate Change	0.012*"climate" + 0.011*"atmospheric" + 0.011*"forecast" + 0.010*"precipitation" + 0.009*"state" + 0.009*"united" + 0.009*"north" + 0.008*"tropical" + 0.008*"pacific" + 0.007*"region" + 0.007*"increase" + 0.007*"cyclone" + 0.007*"meteorology" + 0.006*"frequency" + 0.006*"impact"



		+ 0.006*"rainfall" + 0.006*"event" + 0.006*"resolution" + 0.006*"high" + 0.006*"large" + 0.005*"extreme" + 0.005*"storm" + 0.005*"sea" + 0.005*"temperature" + 0.005*"activity" + 0.005*"seasonal" + 0.005*"regional" + 0.005*"ocean" + 0.005*"variability" + 0.005*"lead"
17	Internet of Things	0.035*"system" + 0.018*"identification" + 0.015*"algorithm" + 0.015*"nonlinear" + 0.011*"source" + 0.011*"estimation" + 0.010*"linear" + 0.009*"parameter" + 0.008*"device" + 0.007*"localization" + 0.006*"base" + 0.006*"estimate" + 0.006*"control" + 0.006*"convergence" + 0.006*"signal" + 0.006*"function" + 0.006*"time" + 0.006*"blind" + 0.005*"light" + 0.005*"noise" + 0.005*"temperature" + 0.005*"order" + 0.005*"square" + 0.005*"measurement" + 0.005*"input" + 0.005*"structure" + 0.005*"physic" + 0.005*"bound" + 0.004*"condition" + 0.004*"band"
18	Polymers and Advanced Materials	0.027*"polymer" + 0.016*"polymerization" + 0.013*"system" + 0.011*"monomer" + 0.010*"rate" + 0.010*"photopolymerization" + 0.010*"crystal" + 0.010*"liquid" + 0.009*"increase" + 0.009*"property" + 0.008*"phase" + 0.007*"kinetics" + 0.007*"hydrogel" + 0.007*"structure" + 0.007*"conversion" + 0.006*"material" + 0.006*"component" + 0.006*"clay" + 0.006*"behavior" + 0.005*"acrylate" + 0.005*"concentration" + 0.005*"effect" + 0.005*"thiol" + 0.005*"order" + 0.005*"film" + 0.005*"spectroscopy" + 0.005*"llc" + 0.005*"caffeine" + 0.005*"group" + 0.005*"light"
19	Human Vibration	0.026*"channel" + 0.012*"frequency" + 0.011*"system" + 0.011*"vibration" + 0.011*"wear" + 0.011*"signal" + 0.009*"head" + 0.009*"filter" + 0.008*"time" + 0.008*"rate" + 0.007*"total" + 0.007*"receiver" + 0.007*"power" + 0.006*"design" + 0.006*"transmit" + 0.005*"feedback" + 0.005*"analysis" + 0.005*"user" + 0.005*"communication" + 0.005*"bit" + 0.005*"base" + 0.005*"cod" + 0.005*"distribute" + 0.005*"optimum" + 0.005*"wireless" + 0.004*"array" + 0.004*"damage" + 0.004*"body" + 0.004*"transmission" + 0.004*"antenna"
20	Robotics	0.026*"sensor" + 0.021*"composite" + 0.016*"control" + 0.014*"network" + 0.013*"agent" + 0.010*"disease" + 0.009*"distribute" + 0.008*"fiber" + 0.008*"matrix" + 0.008*"system" + 0.008*"node" + 0.006*"effective" + 0.006*"time" + 0.006*"group" + 0.005*"oscillator" + 0.005*"base" + 0.005*"formation" + 0.004*"matter" + 0.004*"tensor" + 0.004*"law" + 0.004*"target" + 0.004*"concrete" + 0.004*"measure" + 0.004*"reinforce" + 0.004*"robot" + 0.004*"communication" + 0.004*"demonstrate" + 0.004*"pulse" + 0.004*"achieve" + 0.004*"unit"
21	Construction Technology	0.011*"membrane" + 0.009*"asphalt" + 0.008*"fish" + 0.006*"protein" + 0.006*"surfactant" + 0.006*"concentration" + 0.006*"blood" + 0.006*"pavement" + 0.006*"solution" + 0.006*"surface" + 0.006*"material" + 0.006*"design" + 0.006*"time" + 0.005*"increase" + 0.005*"temperature" + 0.005*"condition" + 0.005*"release" + 0.005*"performance" + 0.005*"interaction" + 0.005*"high" + 0.005*"bond" + 0.005*"determine" + 0.005*"content" + 0.005*"system" + 0.005*"effect" + 0.005*"range" + 0.004*"cell" + 0.004*"recycle" + 0.004*"warn" + 0.004*"disk"
22	Neuroscience	0.036*"bone" + 0.019*"age" + 0.018*"brain" + 0.012*"trabecular" + 0.011*"magnetic" + 0.010*"mri" + 0.010*"measure" + 0.010*"resonance" + 0.008*"participant" + 0.007*"rod" + 0.007*"high" + 0.007*"image" + 0.006*"year" + 0.006*"analysis" + 0.006*"adult" + 0.006*"control" + 0.006*"volume" + 0.005*"association" + 0.005*"compare" + 0.005*"plate" + 0.005*"individual" + 0.005*"vivo" + 0.005*"distal" + 0.005*"longitudinal" + 0.005*"child" + 0.005*"neuroscience" + 0.005*"difference" + 0.005*"neurology" + 0.005*"density" + 0.005*"group"
23	Biological and Pharmaceutical	0.011*"mutation" + 0.011*"void" + 0.010*"gene" + 0.009*"genetic" + 0.009*"variant" + 0.007*"identify" + 0.007*"collapse" + 0.006*"effect" + 0.006*"trenbolone" + 0.006*"beta" + 0.006*"nom" + 0.005*"disease" + 0.005*"spot" + 0.005*"association" + 0.005*"analysis" + 0.005*"alpha" + 0.005*"energetic" + 0.004*"factor" + 0.004*"phenotype" + 0.004*"growth" + 0.004*"high" + 0.004*"disorder" + 0.004*"sequence" + 0.004*"risk" + 0.004*"loss" + 0.004*"variation" + 0.004*"insulin" + 0.004*"syndrome" + 0.004*"level" + 0.004*"locus"
24	Tissue Engineering	0.011*"tissue" + 0.011*"strength" + 0.010*"flame" + 0.008*"fuel" + 0.007*"scaffold" + 0.007*"combustion" + 0.007*"research" + 0.007*"fabrication" + 0.007*"uhmwpe" + 0.006*"high" + 0.005*"pmma" + 0.005*"hydrogen" + 0.005*"material" + 0.005*"pressure" + 0.005*"product" + 0.005*"system" + 0.005*"mixture" + 0.005*"aod" + 0.004*"monsoon" + 0.004*"low" + 0.004*"cement" + 0.004*"hybrid" + 0.004*"modis" + 0.004*"layer" + 0.004*"interfacial" + 0.004*"improve" + 0.004*"organ" + 0.004*"release" + 0.004*"base" + 0.004*"mode"
25	Orthopedics	0.027*"cartilage" + 0.018*"injury" + 0.015*"impact" + 0.014*"articular" + 0.011*"joint" + 0.011*"osteoarthritis" + 0.009*"fracture" + 0.007*"high" + 0.007*"cell" + 0.007*"frequency" + 0.006*"damage" + 0.005*"powder" + 0.005*"animal" + 0.005*"chiral" + 0.005*"knee" + 0.005*"severity" + 0.005*"filament" + 0.005*"post" + 0.004*"risk" + 0.004*"chondrocytes" +

---

0.004\*"drill" + 0.004\*"area" + 0.004\*"ptoa" + 0.004\*"death" + 0.004\*"traumatic" + 0.004\*"esrp" +  
0.004\*"progenitor" + 0.004\*"analysis" + 0.004\*"effect" + 0.004\*"energy"

---