Using Open Calais to Identify the Research Areas of Engineering Faculty

Teresa U. Berry
Research Assistance Coordinator & Science Librarian
John C. Hodges Library
University of Tennessee, Knoxville
tberry0@utk.edu

Jeanine M. Williamson
Engineering Librarian
John C. Hodges Library
University of Tennessee, Knoxville
jwilliamson@utk.edu

Abstract

To meet the research, teaching, and learning needs of their users, academic librarians, particularly those functioning as subject liaisons, are expected to know the institution's curriculum and research areas so that they can help shape library strategies to meet those needs and to connect users to the library's resources and services. The present study investigated the use of Refinitiv's free web demo, Open Calais, as a text mining tool to help learn about the research areas in the University of Tennessee's Tickle College of Engineering. We investigated the following research questions: What interdisciplinary research areas in the College does Open Calais reveal? What are the differences in Open Calais' tagging of Scopus and web pages? What terms were uncovered by Open Calais that were unexpected by the subject librarian? The results showed a mixed picture of the usefulness of Open Calais for learning the research areas of the College of Engineering.

Introduction

To meet the research, teaching, and learning needs of their users, academic librarians, particularly those functioning as subject liaisons, are expected to know the institution's curriculum and research areas so that they can help shape library strategies to meet those needs and to connect users to the library's resources and services. The roles of liaison librarians are evolving from the traditional trifecta of reference, collection development, and library instruction to roles that emphasize engagement in all aspects of the research process, such as data management and scholarly publishing. This shift does not preclude the need to be familiar with the research being conducted on campus. For example, collection development may rely more on

data-driven acquisitions models and on-demand purchasing, but a liaison's knowledge of faculty research interests still comes into play when creating approval plan profiles and demand-driven acquisition plan parameters. The growing emphasis on engagement asks subject liaisons to develop deeper relationships with faculty. Perhaps Whatley (2009) put it best when she said that "building relationships is becoming the essence of what it is to be a liaison librarian" and helps tie the library's services to the university's mission. However, Vine (2018) admits that liaisons found that "research engagement is harder" and that reaching out to faculty to learn about their research areas was "daunting." Regardless, Díaz and Mandernach (2017) found that one of the key characteristics of building good relationships is knowing your users and their disciplines.

Subject liaisons often use a variety of methods to understand the research and teaching needs of researchers on campus, ranging from indirect approaches, such as examining faculty research publications and web pages, to direct methods, such as individual interviews. However, liaisons sometimes find the information difficult to glean due to an overwhelming number of publications from prolific researchers. Web sites can be unhelpful because they lack standardization and are often incomplete or outdated (Wood & Griffin 2016). Curriculum and faculty profile mapping methods can help liaisons become familiar with user research and teaching needs, but the effort is time intensive (Miller 2019). To cope with the amount of information and time constraints, librarians have begun to turn to text mining methods to inform their professional work.

Feldman and Sanger (2007) define text mining as analyzing unstructured textual data, exploring patterns, and extracting information from a collection of documents. Several studies have used text mining techniques to gain insight into the research interests of an academic department or institution. Hendrigan (2019) explored the feasibility of using Voyant Tools, a free web-based application, to examine article titles from Web of Science citations published by two engineering departments. She was able to use the results to familiarize herself with terminology and to inform collection development decisions. Gao and Wallace (2017) analyzed the word frequencies in titles from articles and conference papers published by university researchers over a ten-year period and were able to identify the research areas and trends for individual departments. They suggested that this type of study can be used to inform collection development decisions, especially for interdisciplinary areas of research. Gao (2017) also conducted a similar analysis for faculty in a single academic unit and discovered that including abstracts in a text analysis provided higher term frequencies and, thus, more reliable results than mining only article titles. She also suggested that analyzing by two-word phrases described areas of research more accurately and were more meaningful. Scalfani (2017) took a broader approach and used text analysis to examine over a hundred years of the chemistry thesis and dissertation titles from nine southeastern universities to discover chemistry research trends over time.

While Open Calais may not be as well-known as other text mining tools, a few studies did use Open Calais to extract information. To aid discovery of geo-reference information in digital content. Powell et al. (2010) found that Refinitiv's text mining web application Open Calais was a useful semantic analysis tool for extracting place names from titles and abstracts although the level of granularity may be insufficient for some users. In another study to improve subject access points to archival collections, Zeng et al. (2014) also used Open Calais to analyze the text in finding aids to suggest potential subject headings as well as tagging names of people, corporate entities, or geographic places. They also applied the same technique to titles, abstracts, introductions, and keywords of philosophy theses and dissertations. They found that whereas the tags generated by Open Calais helped identify names and descriptive terms, the subject areas were often defined too broadly to be useful.

Background

The University of Tennessee, Knoxville, is a land-grant institution with approximately 29,000 students and is classified by the Carnegie Foundation as a Doctoral University with Very High Research Activity. The Tickle College of Engineering is comprised of seven departments and seven research centers (Table 1). The College enrolls 4,500 students and has 186 full-time instructional faculty. The University of Tennessee has close ties with the Oak Ridge National Laboratory, and many College of Engineering researchers collaborate with Oak Ridge National Laboratory through a trio of interdisciplinary institutes: The Joint Institute for Computational Sciences, The Joint Institute for Advanced Materials, and the Shull Wollan Center, a Joint Institute for Neutron Sciences. The Tickle College of Engineering also offers biosystems engineering degrees with the Herbert College of Agriculture and several graduate degrees in industrial and systems engineering, mechanical engineering, and flight test engineering through the University of Tennessee Space Institute.

Table 1. Tickle College of Engineering departments and centers

Departments	Research Centers
Chemical and Biomolecular Engineering (CBE)	Center for Materials Processing
Civil and Environmental Engineering (CEE)	Center for Transportation Research
Electrical Engineering and Computer Science (EECS)	CURENT
Industrial and Systems Engineering (ISE)	Innovative Computing Laboratory
Materials Science and Engineering (MSE)	The Institute for a Secure and Sustainable Environment
Mechanical, Aerospace and Biomedical Engineering (MABE)	Reliability and Maintainability Center
Nuclear Engineering (NE)	Scintillation Materials Research Center

The University of Tennessee Libraries has a mix of subject and functional liaisons to support the learning, research, and teaching needs of the academic departments on campus. The engineering librarian is responsible for supporting the needs of the faculty, staff, and students associated with the seven departments in the Tickle College of Engineering as well as those with the University of University of Tennessee Space Institute. With this portfolio, she has the highest number of faculty, staff, and students of any other subject liaison at the Libraries. Although she has been an engineering librarian at the University of Tennessee for several years, it is still challenging to know the research areas of such a large user population with a wide range of interdisciplinary interests. She decided to investigate using Open Calais as a means of learning research areas more efficiently.

The Refinitiv Intelligent Tagging Demo, also known as Open Calais (https://permid.org/onecalaisViewer), is a free text mining tool that "uses natural language processing, text analytics and data-mining technologies to derive meaning from vast amounts of unstructured content." Entities are automatically identified in the text and tagged in several categories including the two categories used in this study, Technology and Social Tags. The tags are derived from multiple sources, such as the International Press Telecommunications Council (IPTC) taxonomy and, in the case of Social Tags, the Wikipedia folksonomy (Refinitiv [date unknown]). The present study makes observations about the nature of the tags found by Open

Calais, as well as evaluating its overall effectiveness in providing data useful to librarians attempting to learn the research areas of engineering faculty.

Research Questions

The overall aim of the study was to determine whether Open Calais was a useful tool for discovering faculty members' research interests in the Tickle College of Engineering. This aim incorporated several sub-questions:

1. What interdisciplinary research areas in the College does Open Calais reveal?
2. What are the differences in Open Calais' tagging of Scopus and web pages?
3. What terms were uncovered by Open Calais that were unexpected by the subject librarian?

Methodology

Using the faculty lists on the departmental web sites, we compiled a list of tenured and tenure-track faculty, omitting those with instructional or research appointments, resulting in a population of 174 faculty members (Table 2). We performed an author search for each individual in Scopus and selected the result with the highest document count affiliated with the University of Tennessee or, in some cases, Oak Ridge National Laboratory. Selecting the most recent 20 documents associated with that author, we exported the document title, abstract, author keywords, and index keywords as plain text. Each set of 20 records was saved into an individual Google document. We chose to limit the Scopus export to 20 documents because the demo will time out with larger files. We then copied and pasted the text from each individual Google document within the HTML body tags into the Refinitiv Intelligent Tagging Demo and ran the demo selecting the "research document" and "free results" options for each document (Figure 1). The technology and social tags that were generated were then entered into an Excel spreadsheet and imported into SPSS Statistics Version 26 (Figure 2).

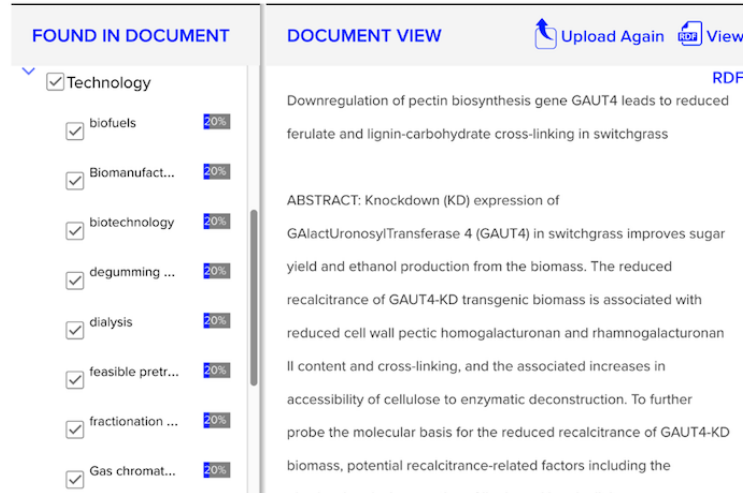Figure 1. Screenshot of Open Calais demo

Figure 2. Example Open Calais results screen

We then captured the content from faculty web pages on the seven engineering departments' websites. Most of them followed a template with the following elements: a brief biography, a brief list of research areas, education, professional service, awards and recognitions, and a list of selected publications and patents. The text was copied and pasted into the tagging demo using the same parameters (research document type and showing free results) and importing the technology and social tags into SPSS.

After the text was processed in Open Calais, we coded the terms associated with two or more researchers for familiarity. Terms were coded as unfamiliar if the librarian either did not know what they meant or if she could not remember examples of research in the area. For example, actinides were coded as unfamiliar even though the librarian knew what the term meant, but she did not remember examples of research on actinides. The unfamiliar terms represent terms the librarian will investigate further to see who is conducting research on these topics. While we could have coded terms generated for only one researcher, we thought tags generated for two or more researchers would be more representative of the college's research areas.

Table 2. Number of faculty members by department

Department	Number of Faculty
Chemical and Biomolecular Engineering (CBE)	19
Civil and Environmental Engineering (CEE)	22
Electrical Engineering and Computer Science (EECS)	42
Industrial and Systems Engineering (ISE)	11
Materials Science and Engineering (MSE)	43
Mechanical, Aerospace and Biomedical Engineering (MABE)	22
Nuclear Engineering (NE)	15

Our analyses included count statistics such as generating crosstabs of the distribution of tags across the seven engineering departments. We also counted how many terms in the top 20

technology and social tags were shared among the Scopus records and the departmental web pages. While the present study reports just 20 terms for each category, the Open Calais procedure yielded many more terms that were useful to the engineering librarian in learning about faculty research interests in the Tickle College of Engineering.

Results

Open Calais revealed a number of interdisciplinary areas of research in the College. Table 3 shows the 20 most frequently occurring technology terms generated by Scopus for each department. All 20 of these terms represented highly interdisciplinary areas, since the terms were generated in multiple departments.

Table 3. Top 20 technology tags for Scopus by department *

	CBE	CEE	EECS	ISE	MABE	MSE	NE	Total
Simulation	12	19	35	11	34	15	13	139
Spectroscopy	12	2	6	0	13	19	12	64
X-ray	11	5	3	0	11	18	6	54
Radiation	6	6	5	2	9	13	12	53
Laser	6	4	3	1	13	9	2	38
Machine learning	0	3	16	5	5	1	1	31
Thermodynamics	8	1	2	0	8	4	4	27
Fluid dynamics	0	4	0	2	16	1	2	25
3-d	2	3	3	1	11	0	3	23
Artificial intelligence	0	2	12	4	4	0	0	22
Tomography	2	2	2	0	7	4	5	22
Crystallization	6	1	2	0	3	8	1	21
Dielectric	5	1	5	0	2	4	3	20
Neural network	0	2	11	2	3	1	1	20
Semiconductors	2	0	7	0	3	7	1	20
Image processing	0	2	6	0	10	0	0	18
Alpha	1	4	2	0	0	4	6	17
Fuel cells	4	0	0	1	5	5	2	17
Heat transfer	0	0	0	0	11	2	4	17
Lasers	2	1	1	0	6	5	2	17

*CBE = Chemical & Biomolecular Engineering; CEE = Civil & Environmental Engineering; EECS = Electrical Engineering & Computer Science; ISE = Industrial & Systems Engineering; MSE = Materials Science & Engineering; MABE = Mechanical, Aerospace & Biomedical Engineering; NE = Nuclear Engineering

While some of the terms were general research techniques (e.g., tomography, simulation) or data mining techniques employed in a wide range of disciplines (e.g., machine learning, neural

networks, artificial intelligence), other terms identified research areas that were surprisingly interdisciplinary. For example, although the subject librarian knew radiation was an important area of study for nuclear engineering, she was unaware that it was being studied by all the engineering departments. In addition, while she knew fuel cells were an important research area for the College, she did not know the extent of interdisciplinary activity for this field. Other terms confirming important interdisciplinary research areas were semiconductors, heat transfer, fluid dynamics, image processing, dielectric, and crystallization. While it may be that some of these terms would be obviously interdisciplinary to a librarian with an engineering background, the breakdown by department was useful to a librarian who had learned about engineering topics on the job.

The social tags generated by Scopus similarly revealed interdisciplinary activity in the College, although with a broader and perhaps less useful level of analysis than the technology terms. Table 4 shows the top 20 most frequently occurring social tags.

Table 4. Top 20 social tags generated from Scopus by department *

	CBE	CEE	EECS	ISE	MABE	MSE	NE	Total
Physical sciences	13	7	6	0	19	21	12	78
Natural sciences	9	2	3	0	10	18	10	52
Chemistry	13	3	2	0	6	13	5	42
Articles	2	2	9	5	6	0	1	25
Academic disciplines	1	2	8	3	5	0	0	19
Materials science	0	2	0	0	2	12	2	18
Branches of biology	5	2	5	0	2	0	0	14
Materials	3	3	1	0	5	1	1	14
Condensed matter physics	2	0	1	0	0	8	2	13
Electromagnetism	0	0	11	0	0	8	2	13
Physics	0	0	2	0	3	3	4	12
Electrical engineering	0	0	11	0	0	0	0	11
Energy	0	0	5	1	3	1	1	11
Building materials	0	6	0	0	3	1	0	10
Emerging technologies	2	0	5	0	1	1	1	10
Fields of mathematics	0	1	1	5	3	0	0	10
Aerodynamics	0	0	0	1	7	0	1	9
Biology	4	2	2	0	1	0	0	9
Fluid dynamics	0	0	0	0	8	0	1	9
Medicine	0	0	2	0	7	0	0	9

*CBE = Chemical & Biomolecular Engineering; CEE = Civil & Environmental Engineering; EECS = Electrical Engineering & Computer Science; ISE = Industrial & Systems Engineering; MSE = Materials Science & Engineering; MABE = Mechanical, Aerospace & Biomedical Engineering; NE = Nuclear Engineering

While some of these terms were not surprising (e.g., "physical sciences") and were represented in all but one department, others generated new information. For example, while the librarian was aware that a few departments were studying biological areas (represented by the terms "biology" and "branches of biology"), she did not realize the breadth of departments conducting research in these areas.

While there was significant overlap when comparing faculty web page tags with Scopus tags, the Scopus records generated more technology and social tags. Tables 5 and 6 show the top 20 most frequently occurring technology and social tags generated from the faculty web pages.

Table 5. Top 20 technology terms from faculty web pages by department *

	CBE	CEE	EECS	ISE	MABE	MSE	NE	Total
Simulation	6	7	2	3	11	4	5	38
Environmental engineering	1	23	0	0	1	0	0	25
Radiation	2	1	1	0	1	7	6	18
Laser	1	1	1	0	10	1	3	17
X-ray	3	2	1	0	2	5	3	16
Spectroscopy	2	1	0	0	5	5	1	14
Fluid dynamics	1	0	0	0	11	0	0	12
Biotechnology	5	2	1	0	1	0	0	9
Alpha	1	0	0	0	1	2	4	8
Flow control	0	0	1	0	6	0	1	8
Fuel cells	3	0	0	0	3	2	0	8
Heat transfer	0	0	0	0	6	0	1	7
Tomography	1	2	1	0	1	0	2	7
Machine learning	0	0	4	0	2	0	0	6
Microwave	1	1	1	0	2	1	0	6
Artificial intelligence	0	0	3	1	0	0	1	5
Fuel cell	2	1	0	0	2	0	0	5
Information technology	0	1	2	1	1	0	0	5
Semiconductor	2	0	1	0	0	0	2	5
Thermodynamics	4	0	0	0	1	0	0	5
*CBE = Chemical & Biomolecular Engineering; CEE = Civil & Environmental Engineering; EECS = Electrical Engineering & Computer Science; ISE = Industrial & Systems Engineering; MSE = Materials Science & Engineering; MABE = Mechanical, Aerospace & Biomedical Engineering; NE = Nuclear Engineering								

Although the total counts were lower, over two-thirds of the top 20 technology terms created by the Scopus records duplicated those generated by the departmental web pages. Although we have provided the top 20 most frequently occurring terms generated by Open Calais, the application produced 167 technology tags from the web pages and 711 technology tags from the Scopus records.

Table 6. Top 20 social tags from faculty web pages by department *

	CBE	CEE	EECS	ISE	MABE	MSE	NE	Total
Physical sciences	8	3	0	0	8	14	7	40
Academic disciplines	3	2	9	8	7	1	3	33
Engineering	2	6	8	9	7	0	1	33
Articles	5	1	2	8	8	0	5	29
Natural sciences	5	2	0	0	6	10	6	29
Chemistry	7	2	0	0	4	4	1	18
Materials science	2	0	1	0	3	10	1	17
Physics	0	1	1	0	3	7	4	16
American Institute of Aeronautics and Astronautics	0	0	1	0	11	0	0	12
EECS	0	0	12	0	0	0	0	12
Computer science and engineering	0	0	11	0	0	0	0	11
Electromagnetism	1	0	9	0	0	1	0	11
Nuclear engineering	0	0	0	0	0	1	9	10
Industrial engineering	0	0	0	9	0	0	0	9
Mechanical engineering	0	0	1	0	8	0	0	9
American Society of Civil Engineers	0	8	0	0	0	0	0	8
Computer science	0	0	8	0	0	0	0	8
Electrical engineering	0	0	8	0	0	0	0	8
Materials	1	2	1	0	2	2	0	8
Neutron	0	1	0	0	0	3	4	8

*CBE = Chemical & Biomolecular Engineering; CEE = Civil & Environmental Engineering; EECS = Electrical Engineering & Computer Science; ISE = Industrial & Systems Engineering; MSE = Materials Science & Engineering; MABE = Mechanical, Aerospace & Biomedical Engineering; NE = Nuclear Engineering

The most frequent social tags derived from the faculty web pages on the departmental web sites are similar to those generated by the Scopus results, with an overlap of 50 percent. The tags again reflected broad categories, such as "engineering" and "academic disciplines," and were associated with multiple departments. However, Open Calais also picked up organization names such as "American Institute of Aeronautics and Astronautics" and "EECS" (the acronym for the Department of Electrical Engineering and Computer Science). These names most likely derived from the publication lists and biographical information found on the web sites. There was less interdisciplinary distribution among the majority of organizational tags. There were 1163 social tags generated by the Scopus text, whereas there were 687 generated by the faculty web pages.

Overall, considerably more social tags than technology tags were produced in the departmental web pages. The average number of social tags and technology tags per researcher from the Scopus records were roughly equal (12.3 social tags and 11.4 technology tags), but we saw a large difference between the number of social tags and technology tags per researcher in the

faculty web pages (7.5 social tags and 2.5 technology tags). The content of faculty web pages was much briefer than their Scopus records and could account for the difference in the number of tags generated by Open Calais.

The technology and social tags that were associated with more than one researcher were examined for familiarity by the subject librarian. As mentioned above, familiarity did not refer to whether the concept was known to the librarian, but whether she previously knew the concept was being studied in the College. Although this familiarity coding was a completely personal, non-objective measure, it served our goal to ascertain Open Calais' usefulness for our purpose. Results will differ for other institutions depending on the knowledge of individual subject librarians and the institutional data sets. Of the 205 Scopus technology tags examined, 65 terms (31.7%) were coded as unfamiliar, whereas only 10 of the 55 (18.2%) technology web page tags were coded as unfamiliar. In examining the social tags, only 30 of the 339 (8.8%) terms generated from Scopus were unfamiliar, and 28 of the189 (14.8%) departmental web page social tags were unfamiliar. Thus, the social tags appeared to be more informative in the departmental web pages, and the technical terms appeared to be more informative in the Scopus results.

Discussion

Open Calais yielded evidence of interdisciplinary activity in the College that both confirmed the importance of known research strengths, as well as making the librarian aware of unfamiliar areas. Overall, Open Calais generated 123 tags relevant to research areas in the department previously unknown by the librarian. While this was a small percentage of the total number of tags, it would have been hard for the librarian to learn these new areas through reading large amounts of Scopus abstracts since the librarian did not have an engineering background and could have quickly been overwhelmed by information overload. In addition, the process allowed the librarian to create a database of faculty and departmental research interests to be used in future personalized outreach.

The research also shed light on how departmental web pages and Scopus records differed in terms of Open Calais tagging results and their usefulness. Although faculty web pages are often used as a primary way of learning about faculty research because they provide a brief and easy-to-read overview, we found that they rarely included the most recently published journal articles. More recently hired faculty members tended to have more current information on their web pages. Despite this lack of currency, there was still about a 70 percent overlap in technology terms between Scopus and the web pages and a 50 percent overlap in social tags.

It was interesting to discover that the percentage of useful social tags and technology tags differed between the two document types. Although some social tags contained terms that were too broad to be useful (e.g., "articles" and "academic disciplines"), the departmental social tags provided more unfamiliar information than those extracted from the Scopus records. We found that 47.2% of the total number of unfamiliar terms were in the social tags, indicating that they were somewhat informative despite the breadth of some terms. Alternatively, there were more unfamiliar technical terms by percentage in the Scopus records than the departmental web pages. This result may in part have been due to the fact that the departmental web pages are written for the non-expert and, thus, have fewer technical terms than the Scopus records.

While we found some benefits in using Open Calais, several disadvantages were noted as well. The process of copying and pasting 174 sets of Scopus records and 174 departmental web pages

into the Open Calais demo was tedious and time-consuming. We could have used a sample set of researchers, but we were interested in seeing all the research areas in the College. Other problems noted with Open Calais are that it often generates both single and plural variants of terms, acronyms are ambiguous, and some terms are entirely too broad (e.g., "articles") and of questionable value. Finally, a very troubling flaw is that the demo version accepts only a small amount of text (about 20 Scopus records) before timing out.

Conclusion

This exploratory study examined the usefulness of Open Calais as a tool for learning about the research interests of faculty in the College of Engineering. Overall, we viewed it as a useful tool. The tags revealed known interdisciplinary research areas that were being conducted in the College, but we also discovered new information about some departments. For example, while we knew that fuel cells are an important research area, we found it interesting that this technology tag was associated with multiple researchers in multiple departments. Some tags, such as "biological sciences," also revealed the degree of interdisciplinarity for research areas spanning across other departments on campus. Subject librarians are often assigned to specific academic departments, but this approach is an artificial division of labor. The multidisciplinary nature of research underscores the need for subject librarians to take a holistic and team-based approach in serving the needs of the campus.

We found different percentages of unfamiliar social tags and technology tags in the two document types, suggesting that the Scopus records and departmental web pages provide complementary information. Results indicate that the departmental web pages generated more unfamiliar social tags, and the Scopus records generated more unfamiliar technology tags. While it is unclear why this difference exists, one possible explanation could be the targeted audience of the documents. Departmental web pages are usually written for an audience who may not have expertise in that research area, whereas, the content in Scopus records is usually aimed at other researchers in the field. Thus, the web pages may be more likely to generate social tags because the research is described in broader terms.

While the technology tags were more specific and narrower than the social tags, they often contained highly technical terms and acronyms that would need further investigation to understand their context or to disambiguate them. For example, the context and meaning of the terms "flow control" and "FDM" were not evident since they can be used in several ways. The social tags, on the other hand, were sometimes too broad and, being based on the Wikipedia folksonomy, might not represent research areas in the same level of detail that the technology tags did. Even though many social tags appeared to be too broad to be useful, almost half of the unfamiliar research areas were identified from the social tags.

Was this work worthwhile in the end? Open Calais uncovered 123 terms representing unfamiliar areas of research and enabled us to create a database of terms linked to departments and individual researchers. While our familiarity measure was subjective and specific to our situation, the concept could be used by librarians at other institutions to test other text mining tools. We used the 123 terms representing unfamiliar research areas to examine the strength of the library's collection in those areas and purchased books to bolster the collection. Some of the 123 terms, like "flow control" needed to be examined in their context since they had multiple meanings in engineering. Some acronyms like FDM needed to be looked up since Open Calais did not spell them out. Some unfamiliar research areas, such as neurosurgery, already had

corresponding books in the library collection purchased by librarians in other subject areas, emphasizing the need to collaborate with colleagues in developing collections for interdisciplinary areas. We also plan to use the information to target outreach efforts.

The analysis did make us much more aware of research areas in the College as well as showing us that we had more to learn. The chief benefit of using a text mining tool was that it helped a librarian with no engineering background to assimilate a large amount of text in order to learn about the research areas of a large college.

References

Díaz, J.O. & Mandernach, M.A. 2017. Relationship building one step at a time: Case studies of successful faculty-librarian partnerships. *Portal: Libraries and the Academy*. 17(2):273–282. DOI: 10.1353/pla.2017.0016.

Feldman, R. & Sanger, J. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge (UK): Cambridge University Press.

Gao, W. 2017. Text analysis of communication faculty publications to identify research trends and interest. *Behavioral & Social Sciences Librarian*. 36(1):36–47. DOI: 10.1080/01639269.2017.1507223.

Gao, W. & Wallace, L. 2017. Data mining, visualizing, and analyzing faculty thematic relationships for research support and collection analysis. *Association of College & Research Libraries Conference*. ALA; Baltimore, MD. Available from http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2017/DataMiningVisualizingandAnalyzingFacultyThematicRelationships.pdf.

Hendrigan, H. 2019. Mixing digital humanities and applied science librarianship: Using Voyant Tools to reveal word patterns in faculty research. *Issues in Science and Technology Librarianship*. 91. DOI: 10.29173/istl3.

Miller, M. 2019. Curriculum, departmental, and faculty mapping in the visual arts department. *Art Documentation: Journal of the Art Libraries Society of North America*. 38(1):159–173. DOI: 10.1086/702919.

Powell, J., Mane, K., Collins, L.M., C., Martinez, M.L.B. & McMahon, T. 2010. The geographic awareness tool: Techniques for geo-encoding digital library content. *Library Hi Tech News*. 27(9/10):5–9. DOI: 10.1108/07419051011110586.

Refinitiv. Intelligent tagging [Internet]. New York: Refinitiv; date unknown; [cited 2020 March 7]. Available from https://www.refinitiv.com/en/products/intelligent-tagging-text-analytics.

Scalfani, V.F. 2017. Text analysis of chemistry thesis and dissertation titles. *Issues in Science and Technology Librarianship*. 86. DOI: 10.5062/F4TD9VBX.

Vine, R. 2018. Realigning liaison with university priorities: Observations from ARL Liaison Institutes 2015-18. *College & Research Libraries News*. 79(8):420–456. DOI: 10.5860/crln.79.8.420.

Whatley, K.M. 2009. New roles of liaison librarians: A liaison's perspective. *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC.* (265):29–32. DOI: [10.29242/rli.265.6](https://doi.org/10.29242/rli.265.6).

Wood, N.B. & Griffin, M. 2016. Liaison librarians in the know: Methods for discovering faculty research and teaching needs. *Proceedings of the Charleston Library Conference.* Charleston, SC. DOI: [10.5703/1288284316466](https://doi.org/10.5703/1288284316466).

Zeng, M.L., Gracy, K.F. & Žumer, M. 2014. Using a semantic analysis tool to generate subject access points: A study using Panofsky's theory and two research samples. *Knowledge Organization.* 41(6):440-451. Available from https://oaks.kent.edu/slispubs/65.