ISTL Issues in Science and Technology Librarianship

Mixing Digital Humanities and Applied Science Librarianship: Using Voyant Tools to Reveal Word Patterns in Faculty Research

Holly Hendrigan
Liaison Librarian
Faculty of Applied Sciences
Simon Fraser University
Surrey, BC, Canada
holly_hendrigan@sfu.ca

Abstract

Awareness of faculty research interests is an important aspect of a subject librarian's responsibilities. This paper illustrates the potential of Voyant Tools, an application in wide use among digital humanities researchers, to reveal word patterns in the research output of applied science faculty. A corpus of recent article citations from Web of Science from two engineering departments was obtained, and the articles' title field was extracted and uploaded to the application. The exercise indicated that articles on fuel cells dominates the research output of one department, and articles on optical coherence tomography dominates the other. Both the corpus of citations and its visualizations in Voyant Tools contribute to librarians' knowledge of their departments and historical spending patterns on specialized resources. This knowledge can be used in professional practice, including collection development and instruction. As academic subject areas become increasingly complex and multidisciplinary, this paper encourages librarians to engage with Voyant Tools to better understand the specialized language and concepts of these evolving fields.

Introduction

Becoming a competent, confident applied sciences librarian is a challenge. The position requires librarians to be familiar with the research output and curriculum areas in their departments, but this information comes in a dizzying array of sources, formats, and platforms, including department and faculty web pages, course outlines, social media feeds, and course materials such as textbooks and online readings (Drewry 2016). Engineering librarians are further challenged by the limited and occasionally dated information provided in department and individual faculty member web sites (Gao and Wallace 2017). The propensity for engineers to consult departmental colleagues and search engines for their information needs (Zhang 2015; Wellings & Casselden 2017) leaves librarians detached from most engineering faculty's research programs. Furthermore, the vast majority of librarians hold undergraduate degrees in the humanities or social sciences (Oliver & Prosser 2017), and therefore lack fluency in applied science subjects and vocabulary. Academic librarians of all disciplines face a learning curve in understanding their liaison areas, but this curve is especially steep in science, technology, engineering and medicine (STEM) fields, due to the highly technical nature of the curriculum and research literature.

Various approaches have been suggested for librarians to attain basic knowledge in their disciplines. These methods include completing subject-specific coursework in library and information (LIS) programs, work experience in discipline-specific libraries, participating in professional associations, and keeping up with the literature on STEM librarianship (Maness 2016). Further suggestions include attending department meetings and social events, studying faculty web pages, running literature searches on faculty members, and reading the subject literature (Cataldo et al. 2006). Jensen (2009) gained insight into STEM faculty collection interests through an online survey; Fritzler (2013) completed undergraduate science courses and utilized STEM reference sources. Whether librarians choose all or some of the above strategies, learning an unfamiliar subject and building new relationships with faculty members requires a sizeable investment of time and professional development (Fritzler 2013).

In 2016, I had served for seven years as a social sciences and humanities librarian, when the academic library I work at underwent a substantial re-organization of its liaison program, and I was assigned to the Faculty of Applied Sciences liaison portfolio. The three schools within this portfolio are Engineering Science[1], Mechatronics Systems Engineering (hereafter referred to as "Mechatronics"), and Computing Science. Mapping out the scope of the departments was a challenge: the departments' web pages were not particularly useful for the level of understanding required for my work, and face-to-face meetings with the faculty representatives to the Library were relatively short in duration due to time pressures. Conducting literature searches on faculty members was similarly of limited use due to information overload: combined, the two engineering departments employ 45 faculty positions (Simon Fraser University 2017 a,b), and each faculty member had several publication credits per year. Even if time permitted reading abstracts or close readings of their recent journal articles, full comprehension would be beyond the capacity of an information specialist with a humanities degree.

In reflecting on the literature on STEM librarianship onboarding, I realized that an additional heuristic has been overlooked: a "big data" approach used by digital humanities scholars. Academic libraries are increasingly supporting students and faculty in these methodologies (Bryson et al. 2011; Smiley & Rodriguez 2017), but as Cleary et al. noted (2017), most academic librarians have not adopted these approaches in their own practice. STEM departmental activities fit the definition of big data in terms of its complexity and the volume of its output: as Graham et al. claim, "big data is simply more data that you could conceivably read yourself in a reasonable amount of time -- or, even more inclusively -- information that requires computational intervention to make new sense of it." (2015, 3). As confirmed by Paul Campbell in his narrative of onboarding as a new subject librarian (2018), attempting to memorize the specifics of each faculty members' research interests is exceedingly difficult. An alternative approach could involve text mining the research output.

The purpose of this research is to explore how a text mining and data visualization application, Voyant Tools (https://voyant-tools.org/), is useful for STEM librarians, and why adopting digital humanities methods to STEM librarianship praxis is appropriate. To investigate these questions, this article will proceed in four parts. The literature review will report on some recent cases on digital scholarship applied to academic librarianship praxis. The next section describes the method for developing the corpus for analysis in Voyant Tools. The third section, Results, documents three attributes of Voyant Tools that shed light on the departments' research output. The fourth section discusses why STEM research citations are appropriate for computational analysis, and applies Franco Moretti's theory of "distant reading" to the process (2013). "Distant reading" explains the epistemological significance of text mining and data visualizations of STEM faculty research output. Ultimately, this paper confirms Gao and Wallace's conclusion (2017) that digital scholarship[2] is an exciting and effective means of gaining insight into faculty publishing patterns.

Literature Review

Although the peer reviewed literature on academic librarians adopting digital humanities methods in their praxis is scant, a growing body of conference presentations indicates an upward trend in such approaches. Taylor Hixson at NYU Abu Dhabi used Voyant Tools to analyze the responses of a self-assessment survey administered to GIS students (Hixson 2018). Lydia Bello et al. (2016) of Claremont College analyzed monographic acquisition patterns on the topic of terrorism using a variety of DH applications including Voyant Tools, TimelineJS, and Tableau. Bello's project provided an opportunity in experiential learning to complement the new digital humanities programs their university had initiated. Gao and Wallace (2017) were frustrated by dated faculty web pages and uneven communication patterns between librarians and departments, and mitigated these deficits by mapping historical patterns in their university's research output. They compiled ten years of faculty citations from Scopus, isolated the title field, and processed the corpus through Mallet, a topic modeling software (http://mallet.cs.umass.edu/topics.php). Mallet provided a list of high frequency co-occuring terms, which they ran through Tableau (https://public.tableau.com/en-us/s/), allowing them to see changing trends in their institution's research focus.

Methods

This investigation involved collecting a corpus of citations, extracting the title field, and running this corpus through a text mining and data visualization application. This method is similar to Gao and Wallace's, but is less labour intensive. I used Web of Science's (WoS) Core Collection database records (https://clarivate.com/products/web-of-science/web-science-form/web-science-core-collection/) as my data source and Voyant Tools as the text mining and data visualization application. I chose Voyant Tools because it is a free, open-source, web based application with numerous text mining and data visualization features. These features include word collocates (which approximates Mallet's topic modeling capabilities) in addition to word frequencies. Voyant Tools requires no further programming support, unlike Mallet. Indeed, none of the applications used in this research (Notepad, MS Excel, MS Word, or Voyant Tools) require any special expertise.

I extracted citation records from faculty members of SFU's Mechatronics and Engineering departments by using the WoS's address field, which captures the institutional and departmental affiliation of primary and secondary authors listed. I accounted for variations on school names and locations as well, due to some fluidity in the Mechatronics' name and campus locations. I limited the result set to the date range of 2015-2018 in order to capture only relatively recent research output. I downloaded the records as delimited text files, opened in a text editor, then copied to an Excel spreadsheet.

From the spreadsheet, the title column was copied into a Word document and normalized. Normalization involved changing all characters in the title to lower case, as well as some cleaning of the data. Some terms were normalized as plurals: in the Mechatronics corpus "cell" and "membrane" became "cells" and "membranes," because an early test of the data revealed an even split between the singular and plural forms of the term. The phrase 3-D was also changed to "three dimensional." Graham et al. (2015, 7) call the resulting corpus a "[bag] of words," ready for analysis in text mining software.

I chose the title column for analysis for several reasons. First, titles of articles in academic journals in STEM disciplines are composed in a straightforward manner and "include many substantive words" (Hagler 1991, 34), such as nouns and nominals, which makes them suitable for subject analysis. Furthermore, words from the keyword field occasionally repeat terms in the title. For example, the author-supplied keywords for the article "Fatigue properties of catalyst coated membranes for fuel cells: ex-situ measurements supported by numerical simulations" are: Fuel cell; Membrane; Catalyst layer; CCM; Fatigue. To include both title and author keyword field in the corpus would falsely double their significance. However, to choose only the keyword field would result in a lack of context and nuance regarding the articles in the corpus. In addition, Web of Science's KeyWords Plus headings are generated from the article's cited references, which provides a set of terms that can be quite different from the terms used in the original article. In "Fatigue properties of catalyst coated membranes...", these terms were: Proton-Exchange Membrane; Polymer Electrolyte Membrane; Gas-Diffusion Layers; Mechanical-Properties; PFSA Membranes; Hygrothermal Fatigue; Humidity Cycles; Finite-Element; Degradation; Model. In this case, uploading the article title, author keyword, and Web of Science keyword fields would obfuscate the patterns. The title field is sufficient to provide a corpus of significant terms.

From the Voyant Tools home page, I uploaded the Word document of article titles and clicked the "Reveal" button. An array of windows appeared, displaying a variety of word frequencies and data visualizations "revealed" from the corpus. (See Tables 1-3 and Figure 2 Results section.) I removed the words "based" and "using" to Voyant Tools' list of 484 stopwords, since they were frequently used yet nondescriptive terms.

Due to the multidisciplinary nature of engineering research, these steps should be repeated to compile a separate corpus of citations of research groups within the departments. This process involves compiling lists of members of research groups, such as Intelligent Systems and Control, or Biomechanics (typically found on department web pages), and inputting the names of the individual researchers in Web of Science. This additional corpus provides insight into more specialized topics within departments.

The author acknowledges, however, some limitations of the dataset and the inferences one can draw from it. To state the obvious, the terms at the top of the corpus' frequency list will be terms used by researchers who have the highest publication rates. In applied sciences, some types of research generate a higher rate of citations than others. Fabrications, for instance, take longer than simulations. This technique thus provides one view of the research output, but further analysis by individual research groups or less frequent terms is essential. The dataset itself consists of citations published in Web of Science-indexed journals and conferences only. While

Web of Science is publisher neutral with comprehensive coverage of STEM disciplines, some gaps in research output doubtlessly still remain. Finally, some have raised questions about multiauthored works in the sciences (Tscharntke 2007): lengthy lists of secondary authors can raise questions about each credited author's contribution: does a credit necessarily indicate a strong research interest? For the purposes of this research, that answer was "yes," but I acknowledge that the correlation between credit and interest might occasionally be weak.

Results

This research confirms Gao and Wallace's conclusion that text mining tools can provide inroads into understanding faculty research output. In this case, I was working with 254 records from Mechatronics and 331 from Engineering Science. While Voyant Tools provides dozens of different "skins" or tools which configure data in various visualizations, I focus on three attributes in this paper: the Terms view, the Collocates graphs, and the Context tool. Additionally, I examine only the most frequently used terms, keeping in mind that the terms found farther down the lists provide insight into the more multidisciplinary aspects of departmental research activity.

"Terms" View

The Terms view is "a table view of term frequencies in the entire corpus." (Voyant Tools n.d.) Table 1 provides the terms view of the Mechatronics and Engineering corpus.

Table 1: Highest frequency terms in the Mechatronics and Engineering Science departments' article title corpus, January 2015-August 2018

Mechatronics			Engineering Science		
	Term	Count		Term	Count
1	cells	48	1	optical	24
2	fuel	42	2	coherence	21
3	membrane	30	3	adaptive	20
4	power	23	4	tomography	20
5	control	22	5	analysis	18
6	thermal	21	6	detection	17
7	applications	19	7	high	17
8	energy	17	8	network	17
9	optimization	17	9	control	15
10	dimensional	16	10	video	15
11	mechanical	15	11	imaging	14
12	modeling	15	12	model	14
13	heat	14	13	applications	12
14	low	14	14	design	12
15	method	14	15	fall	12

The terms view is a straightforward list of the most frequent terms in the corpus with the exact number of occurrences. The terms "fuel," "cells", and "membrane" dominate research output in Mechatronics; "optical," "coherence," and "adaptive," and "tomography" stand out in Engineering Science. These terms, in isolation, provide an important, but limited degree of information regarding the faculty research activity.

Collocates Graph

While counts of individual words is important, research topics of modern, multidisciplinary departments are more likely to be expressed as phrases. Voyant Tools recognizes the need to provide analysis beyond individual word counts and includes visuals for networks of terms that occur frequently. The collocates graph represented in the Links view "represents keywords and terms that occur in close proximity as a force directed network graph." ([Voyant Tools n.d.](#)) Figure 1 reveals the collocates of the Mechatronics and Engineering Science corpuses, providing for additional context and meaning to the frequency charts.

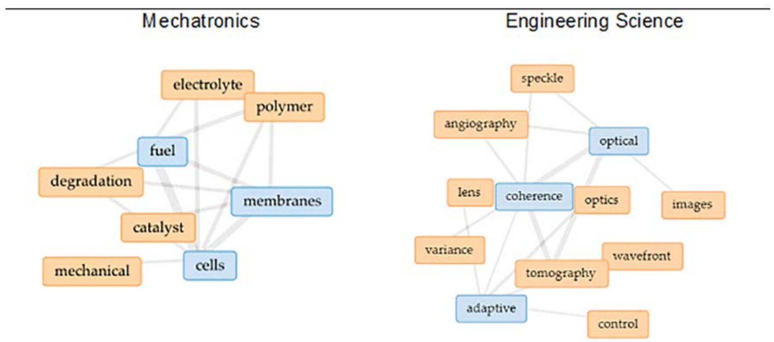

Figure 1: Terms most frequently found in close proximity in the Mechatronics and Engineering Science departments' article title corpus, January 2015-August 2018

The collocates graph displays blue keywords that are linked to orange collocates. We see that eight terms are frequently in the same network of terms, which provides a better sense of the topics of the research articles: for Mechatronics, membranes of polymer electrolyte fuel cells; for Engineering Science, optical coherence tomography. On the application itself, clicking on a term boldens the link between its most frequent collocates.

Contexts Tool

The Contexts tool functions as a concordance, and shows each occurrence of a term with its surrounding text. This feature is useful for studying how terms are used in different contexts. Table 2 presents the term "cells" and its adjacencies in the Mechatronics corpus; Table 3 presents the term "optical."

Table 2: Adjacencies of the term "cells" from Mechatronics corpus, arranged alphabetically by left adjacent term

altered mechanical properties in cancer	cells	as measured by an optical
isolation of rare prostate cancer	cells	from blood improving sensitivity of
two-mode approximation microfluidic electrochemical	cells	array in series: effect of
density of aqueous electrochemical flow	cells	with in operando deposition an
of microfluidic co-laminar flow	cells	a virtual space vector modulation
ionomer membrane constrained by fuel	cells	electrodes mechanical degradation of fuel
model for heavy duty fuel	cells	systems power density optimization of
flow-through microporous electrodes fuel	cells	durability enhancement with cerium oxide
membrane for polymer electrolyte fuel	cells	dynamic model of oxygen starved
fatigue in polymer electrolyte fuel	cells	multi-objective optimization of the
performance of polymer electrolyte fuel	cells	three dimensional printed flow field
gas diffusion layers for fuel	cells	performance of finned tubes used
micro-porous layers for fuel	cells	by nano-scale x-ray
catalyst coated membrane for fuel	cells	: ex-situ measurements supported by
cycle metrics for hydrogen fuel	cells	vehicles evaluation of low-pressure

Table 3: Adjacencies of the term "optical" in the Engineering Science corpus. Arranged alphabetically by right adjacent term

Left	Term	Right ↑
images derived by speckle variance	optical	coherence tomography with histology [illegible]
with real-time speckle variance	optical	coherence tomography a framework for
grade adaptive optics swept source	optical	coherence tomography [illegible] compressed
fovea centralis using speckle variance	optical	coherence tomography automatic annotation of
density of prototype and commercial	optical	coherence tomography angiography [illegible]
developing wavefront sensorless adaptive optics	optical	coherence tomography for in vivo
scanning laser ophthalmoscopy combined with	optical	coherence tomography for simultaneous in
retinal surfaces from 3-d	optical	coherence tomography images intelligent deflection
analysis and classification of retinal	optical	coherence tomography images using the
field multispectral scanning laser ophthalmoscopy-	optical	coherence tomography mouse retinal imager
living correlation between polarization sensitive	optical	coherence tomography and [illegible] microscopy
based registration of serially acquired	optical	coherence tomography angiography exploration of
of the perivascular circulation using	optical	coherence tomography angiography power and
in graphene quantitative comparisons between	optical	coherence tomography angiography and matched
mixture of gaussian distribution quantitative	optical	coherence tomography angiography of radial

While the collocates graph provide insight into word clusters, the contexts tool allows users to view direct adjacencies of any term, in the context of the article title where the term is used. Admittedly, this view provides puzzling results when the term examined is the first or final word in the article title, because its adjacent term is the first word of the preceding or following article in the corpus. Overall, however, the contexts tool greatly enhances the ability to view patterns in short phrases.

Discussion

Voyant Tools rearranged the titles of the journal articles into decontextualized and quantifiable components: words, word networks, and adjacencies. Where the Mechatronics web page listed the research category "energy systems," Voyant Tools more specifically revealed that this research predominantly involves fuel cells. The collocates graph more specifically indicated that the research was focused on the membranes of polymer electrolyte fuel cells. Seven SFU faculty publish in this area (about half of the department), confirming that, indeed, there is widespread interest and not overrepresentation by one faculty member. Voyant Tools also reveals that the School of Engineering Science's research output is dominated by optical coherence tomography, an imaging technique used in medical applications, and, again, multiple research groups are represented in those results. While this article focuses on the most frequently used terms in the corpus, the application allows users to search or click on any term and view its context and adjacencies. In sum, Voyant Tools is a useful application for revealing lexical patterns. Once revealed, these patterns enable a novice engineering librarian to discover the vocabulary used by her constituencies.

On a deeper level, Voyant Tools facilitates a process which Italian literary theorist Franco Moretti recently coined "distant reading" (Oberhelman 2015, 59). Since the mid-twentieth century, American literary scholars favoured textual analysis via close readings: "the critic should give a detailed, almost microscopic analysis...to find its meaning" (Oberhelman 2015, p 57). Moretti was less interested in individual texts of a particular nation; he was more interested in studying literature on a broader historical and national scale. With world literature, "we are talking of hundreds of languages and literatures here. Reading 'more' seems hardly to be the solution." (Moretti 2013, 45). He argued that text mining digitized for common themes and motifs results in a more quantifiable and less selective literary analysis. Where close reading provides a microscopic view of one text, distant reading facilitates a macroscopic view of the bigger picture.

Closely reading the entire research output of faculty members is beyond the scope of most LIS professionals' job duties (Drewry 2016); LIS workers are trained in assigning and working with records consisting of metadata. Still, this Voyant Tools project, which includes in its corpus the titles of journal articles, is an adapted form of distant reading. Understanding trends and themes of 600 journal article titles is overwhelming, whether the reader has a background in STEM or not. However, strategic explorations into entire departments' research output, as well as subdisciplinary research groups and even some individual scholars, will provide a rich variety of words and phrases that assist in revealing the "aboutness" of the department. The output of text mining, as Eric Lease Morgan says, is "akin to a book's table of contents and back-of-the-book index. They outline, enumerate, and summarize.... It is a form of analysis and a way to deal with information overload" (Morgan 2011). In this case, Voyant Tools reveals patterns in the research output of faculty members, as well as frequently used terms that librarians should learn in order to be more comfortable with their STEM collections.

I used VT's visualizations to familiarize myself with phrases such as "optical coherence tomography," but I found it helpful in other facets of my professional practice. The prominence of the term "optical" in the Engineering Science skins made me understand historical decisions of previous applied science librarians, including the choice in 2003 to subscribe to the digital library of SPIE, the Society of Photo-Optical Instrumentation Engineers, and the decisions since 2014 to purchase their e-book collections as well. Similarly, I had inherited an approval plan that had been developed years before I assumed the position; Voyant Tools allows me to map frequently used natural language terms to my profiles' broader subject categories. I use the corpus on a regular basis to make title-by-title collections decisions, searching for the presence (or absence) of terms and phrases in titles available for purchase. Voyant Tools has also helped me make relevant search examples during instruction: for instance, I asked Mechatronics graduate students to look up "fuel cells" in the Frost and Sullivan market research reports. Both the corpus of citations and its visualizations in Voyant Tools has the potential to enhance librarians' basic knowledge about their departments that they can then apply in their professional practice.

Still, librarians may have reservations about Voyant Tools' text mining and visualization outputs. Among other visualization features, it generates word clouds, which are often presented as an exhibit or piece of evidence without further explanation of their significance. In isolation, word frequency lists do not answer the question, "so what?" Jacob Harris, former New York Times interactive news developer, has harsh criticism for word clouds in journalism; he decries "reporters sidestepping their limited knowledge of the subject material by peering for patterns in a word cloud -- like reading tea leaves at the bottom of a cup" (Harris 2011). Harris is correct: word frequencies are not a substitute for knowledge or analysis. This paper argues that text mining a corpus can be a heuristic in learning the lexicon of the corpus, but acknowledges the limitations of this learning object. The Voyant Tools output for Mechatronics Systems Engineering, for instance, does not reveal whether the faculty's articles on fuel cells resulted in any major breakthroughs in this field. Word frequencies do not assist outsiders in constructing an accurate narrative of any academic department. To gain this level of knowledge and understanding, a librarian must consult other data sources and build meaningful relationships with department members.

Librarians might also mistrust natural language word frequencies, even when the corpus is derived from titles of scientific papers. Creating and using controlled vocabulary in organizing and retrieving information is a foundational pillar of LIS practice. As Ronald Hagler reports, this practice led to librarians "shun[ning] natural-language access points for subject retrieval for almost a century" (Hagler 1991, 180). As builders and users of structured metadata schemes that account for lexical factors such as variant spellings and synonyms, librarians are loath to rely on word frequencies rather than controlled vocabularies. Since the practice of librarians moving between controlled and uncontrolled vocabularies is so ingrained, we could use Voyant Tools as an opportunity to reveal the predominant natural language terms in order to find their place in our controlled vocabulary schema.

Conclusion

This paper provides a case where Voyant Tools is a viable heuristic for better understanding two departments' research output. Traditional methods of acquiring knowledge into departmental activities are still relevant: building relationships with faculty, staff and students remains vitally important, as is engaging with networks of STEM librarians. However, because the research activities of faculty in applied science departments are increasingly complex and multidisciplinary, computational analysis is a valid methodological approach. Voyant Tools is an ideal text mining and data visualization application for novices in digital humanities methods. Where reading even the journal article titles of a department or research group is overwhelming for a non-specialist, Voyant Tools illuminates the most frequently used terms and their collocates, shedding light on the departments' research output. Voyant Tools uses fewer steps and requires no programming support in contrast to applications such as Mallet.

Future research with Voyant Tools could be used to analyze a wide range of documents. This research examined research output, but STEM librarians in academic settings could also experiment with teaching and learning documents, such as syllabi and course outlines. One could also return to the Web of Science search results and examine journal titles or the Keywords Plus headings for a different perspective, or add the author keywords and abstracts to the corpus as well. Compiling and comparing corpora from other databases, such as IEEE Xplore, Medline, or graduate theses from institutional repositories, would also be worth investigating. Librarians can gather citations at particular time intervals to examine and compare historical usage patterns. Whatever instance one chooses, this paper demonstrates that librarians should consider adding text mining in Voyant Tools to their traditional methods of learning about their constituents' activities.

Acknowledgements

The author wishes to thank colleagues from Simon Fraser University's Faculty of Engineering Science and members of the Vancouver Science and Engineering Librarians network who provided feedback on earlier drafts of this paper, namely: Steve Whitmore, Mike Sjoerdsma, Deirdre Grace, Julie Jones, Greg Krewski, and Suzanne McBeath. Two anonymous peer reviewers also provided constructive criticism that improved the manuscript.

Notes

[1] The Engineering Science program at SFU is not a traditional engineering school but focuses primarily on the high technology industries, including electrical, communications, computer, microelectronics, and biomedical engineering, as well as engineering physics

[2] In this paper, "digital scholarship" and "digital humanities" are used interchangeably. A full exploration of the definition of the digital humanities is beyond the scope of this paper, but this work--corpus analysis via text mining--falls under Josh Honn's category of scholarship which is "enabled by digital methods and tools." See: Honn, J. 2013 Oct 16. Never Neutral: Critical Approaches to Digital Tools & Culture in the Humanities. Available from https://arch.library.northwestern.edu/concern/generic_works/q524jn78d

References

Bello, L., Clements N., Dickerson M. & Hogarth M. 2016. Critical Collection Analysis: Using DH Tools to Contextualize Historical Collecting Patterns within a Political Framework. [Internet]. [Cited 2018 May 1]. Paper presented at: Charleston, South Carolina. Available from http://scholarship.claremont.edu/cgi/viewcontent.cgi?article=1052&context=library_staff

Bryson, T., Posner, M., St. Pierre, A. & Varner, S. 2011. SPEC Kit 326, Digital Humanities. [Internet]. Washington, DC: Association of Research Libraries [Cited 2018 May 2]. Available from https://publications.arl.org/Digital-Humanities-SPEC-Kit-326/

Campbell, P.C. 2018. Narrowing the scope: transitioning from general librarianship to specialized subject liaisons. [Internet]. [Cited 2018 July 11]. Paper presented at: CAPAL Conference 2018. University of Regina: Canadian Association of Professional Academic Librarians. Available from https://capalibrarians.org/wp/wp-content/uploads/2018/07/6A_Campbell_paper.pdf

Cataldo, T.T., Tennant, M.R., Sherwill-Navarro, P. & Jesano, R. 2006. Subject specialization in a liaison librarian program. *Journal of the Medical Library Association* 94(4):446-448. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1629419/pdf/i1536-5050-094-04-0446.pdf

Cleary, P., Garlock, K., Novak, D., Pullman, E. & Mann, S. 2017. Text mining 101: what you should know. *The Serials Librarian* 72(1-4):156-159. doi: 10.1080/0361526X.2017.1320876

Drewry, J.M. 2016. Changing fields within academic and research libraries. In: Sobel K, editor. *Mastering Subject Specialties: Practical Advice from the Field.* ABC-CLIO. p. 143-154.

Fritzler, P. 2013. Made from scratch: the creation and development of a sciences librarian. In: Kreitz P.A. & DeVries J., editors. *Recruiting, Training, and Retention of Science and Technology Librarians.* London: Routledge. p. 99-112.

Gao, W. & Wallace, L. 2017. Data mining, visualizing, and analyzing faculty thematic relationships for research support and collection analysis. In: *ACRL 2017 Conference Proceedings.* Baltimore: ALA. p. 171-178.

Graham, S., Milligan, I. & Weingart, S. 2015. *Exploring Big Historical Data: The Historian's Macroscope.* London: Imperial College Press.

Hagler, R. 1991. *The Bibliographic Record and Information Technology*. Second ed. Chicago: ALA.

Harris, J. 2011. Word clouds considered harmful. Nieman Lab. [Internet]. [Cited 2018 May 1]. Available from http://www.niemanlab.org/2011/10/word-clouds-considered-harmful/

Hixson, T. 2018. Assessing Spatial Literacy: Results from a one-shot instruction session with undergraduate students [Internet]. [Cited 2018 May 1]. Available from https://www.slideshare.net/TaylorHixson1/assessing-spatial-literacy-results-from-a-oneshot-instruction-session-with-undergraduate-students

Jensen, K. 2009. Engaging faculty through collection development utilizing online survey tools. *Collection Building* 28(3):117-121. doi: 10.1108/01604950910971143

Maness, J.M. 2016. Engineering and applied science librarianship. In: Sobel K, editor. *Mastering Subject Specialties: Practical Advice from the Field*. Santa Barbara, CA: ABC-CLIO. p. 31-38.

Moretti, F. 2013. *Distant Reading.* London: Verso.

Morgan, E.L. 2011. Next-generation library catalogs, or 'Are we there yet?' infomotions mini-mMusings. [Internet]. [Cited 2018 Mar 22]. Available from http://infomotions.com/blog/2011/06/next-generation-library-catalogs-or-are-we-there-yet/

Oberhelman, D.D. 2015. The critical theory of digital humanities for literature subject librarians. In: Hartsell-Gundy, A., Braunstein, L., & Golomb, L., editors. *Digital Humanities in the Library: Challenges and Opportunities for Subject Specialists.* Chicago: American Library Association. p. 53-68.

Oliver, A. & Prosser, E. 2017. Choosing academic librarianship: an examination of characteristics and selection criteria. *The Journal of Academic Librarianship* 43(6):526-531. DOI: 10.1016/j.acalib.2017.08.009

Simon Fraser University, Institutional Research and Planning. 2017a. School of Engineering Science Academic Information Report. [Internet]. [Cited 2018 Dec 12]. Available from http://www.sfu.ca/content/dam/sfu/irp/departments/ensc_summary.pdf

Simon Fraser University, Institutional Research and Planning. 2017b. Mechatronic Systems Engineering Academic Information Report. [Internet]. [Cited 2018 Dec 12]. Available from http://www.sfu.ca/content/dam/sfu/irp/departments/mse_summary.pdf

Smiley, B. & Rodriguez, M. 2017. Digital humanities and academic libraries. In: McDonald J.D. & Levine-Clark, M., editors. *Encyclopedia of Library and Information Sciences.* Fourth edition. Boca Raton, FL.: CRC Press. p. 1298-1306.

Tscharntke, T., Hochberg M.E., Rand T.A., Resh V.H. & Krauss J. 2007. Author sequence and credit for contributions in multiauthored publications. *PLOS Biology* 5(1):13-14. doi: 10.1371/journal.pbio.0050018

Voyant Tools. [Internet]. Available from http://www.voyant-tools.org

Wellings, S. & Casselden, B. 2017. An exploration into the information-seeking behaviours of engineers and scientists. *Journal of Librarianship and Information Science* doi: 10.1177/0961000617742466

Zhang, L. 2015. Use of library services by engineering faculty at Mississippi State University, a large land Grant institution. *Science & Technology Libraries* 34(3):272286. doi: 10.1080/0194262X.2015.1090941