Informing Collection-Development through Citation Examination of the Professional Civil Engineering Research Literature

Abstract

Librarians must scrutinize collection development decisions to ensure that patron use of library materials meets expectations based upon institutional cost and effort to acquire, organize, and provide access to these materials. Some librarians have studied reference citation patterns within engineering as a tool for collection development. Because current commercial citation indexing tools focus on analysis by author and subject, gathering different types of citation information of interest to librarians can prove time-consuming. This paper presents a study that builds upon field-spanning and subfield-specific journal citation studies by Musser and Conkling[8] and Musser[9], a study of theses and dissertations by Eckel[5], and in particular a study of civil engineering theses and dissertation citations by Kirkwood[7]. Kirkwood analyzed citations by format of the cited material, finding that 40% of referenced sources in these theses and dissertations in civil engineering were grey literature, and noted that master's students cited grey literature almost twice as frequently as doctoral students at her institution. This paper seeks to establish whether format use within citations in the professional civil engineering literature differs appreciably from the patterns reported in earlier studies. In addition, the current study proposes a flexible data organization method that should allow for relatively straightforward re-use of the data in future, as yet undetermined, analyses.

Introduction

In engineering librarianship, the economics of scholarly communication place definite limits on any library's ability to build comprehensive collections. Librarians must scrutinize collection development decisions to ensure that patron use of the library materials meets expectations based upon the institutional cost and effort to acquire, organize, and provide access to these materials.

Citation analysis represents one of the most highly developed quantitative tools for obtaining an indicator of information use by communities of practice. By studying the references in journal articles, a library researcher gleans some measure of the previously extant resources used by the researcher to produce the current results. By performing citation analysis over a significant sample size, trends in information use may be measured.

While much of the published research on journal citation analysis has focused on topics such as identifying and mapping social networks among researchers, assessments of journal and subfield impacts, and descriptive studies of the growth of new and interdisciplinary fields, some engineering librarians have studied reference citation patterns as a tool for collection development.

This paper presents a study that builds upon the methods developed by these librarians, creating and analyzing a large sample of citation data from the professional research literature in civil engineering. These data are compared with the data of previous studies, where appropriate, in the hopes of building additional understanding of information use within the discipline, and encouraging others to investigate these questions using quantitatively-driven research.

Literature Review

The current understanding of citation analysis owes its existence to the work of Eugene Garfield, who founded the Institute for Scientific Information and first published the *Science Citation Index*. Garfield himself was principally influenced by concepts expressed in H.G. Wells's 1938 collection of writings *World Brain*, while others have thought of his work in the spirit of Vannevar Bush's 1945 description of the Memex in "As We May Think."[1,4,12] Garfield's thoughts on visualizing the growth of scientific knowledge led to the creation of the first systematic citation indexes of wide subject scope within the sciences. Garfield has written a short article that summarizes his early conception of the citation index system.[3]

Citation analysis provides the researcher with a quantitative measurement tool for the study of information use and communication. Information scientists and other researchers have created a number of different metrics based upon citation data for the purposes of assessing and comparing the importance of works, authors, and journals. However, as has been reported through research such as the historical critique by Brian D. Cameron[2], the development of these "impact factors" has lead to the scrutiny of citation patterns by librarians and information scientists to determine whether researchers and journals engage in market-based practices (e.g. self-citation or increased citation from other articles within the journal) to game these factors for non-academic purposes. Because of the documented existence of such issues with citation analysis, studies done using citation analysis should emphasize that this type of study provides a limited insight into one facet

of a complex system of scholarly communication. In the case of both professional and academically-based engineers, Carol Tenopir and Donald W. King's book provides a thorough review of communications studies for this field.[10]

With that significant caveat, librarians have found citation analysis to be a useful tool to have in the arsenal of resources applied to crafting a collection development strategy. While much research in the information science community (and among science/engineering practitioners) has focused upon impact factors, the social network of research groups, and other studies directed at examining the authors and the journal titles, engineering librarians have analyzed citation patterns to answer questions about other issues, such as the format of sources used and the age of those sources. In their 1996 paper, Linda R. Musser and Thomas W. Conkling provided a benchmark citation analysis study of the entire field of engineering, examining 4,780 citations from 212 articles in 16 different journals. This study was the first in 20 years to look at journals spanning this wide a subject scope in engineering. They found that journal articles made up 53% of all citations, with 19% of citations coming from conference papers, 12% from monographs, and 9% from technical reports. The age of cited sources ranged from those published within the same year to a source 121 years old, with more than 50% of resources 8 years old or less.[8]

Ten years later, Musser applied the same analysis tools to a study of citation patterns in mining engineering, looking at 862 citations from 81 articles in 13 publications. The results were similar for format, with 41% of citations from journal articles, 22% from conference papers, 18% from monographs, and 12% from technical reports. However, mining engineering journal articles had a longer research "tail" than engineering journal articles in general – 90% of mining engineering journal citations were 30 years old or less, while for engineering journals 90% of citations were 21 years old or less.[9]

The implications of Musser's studies include that engineering librarians should collect in a wide variety of formats, and that older resources retain current research interest for engineers.

Engineering librarians have also productively applied the tools of citation analysis to the examination of Master's theses and Ph.D. dissertations from their home institutions. By their nature, theses and dissertations contain extensive and thorough literature reviews of their subject areas. By examining these citations, librarians have a good indicator of what types of resources graduate students have applied to their research. The analysis tools used and measurements reported are very much the same in these theses and dissertation citations studies as those found in Musser's studies of journal article citations.

At the 2009 ASEE Conference and Exposition, Patricia Kirkwood presented the results of a citation study of theses and dissertations in civil engineering at the University of Arkansas, Fayetteville, for the years 2003-2004.[7] A total of 838 citations were studied in both Masters theses and Ph.D. dissertations, with analysis done of the format of the citation. Kirkwood also found differences in citation pattern when considering the Masters' theses separately from the Ph.D. dissertations, with Masters' candidates using more "grey literature," or literature not generally accessed through the main collections of libraries, like technical reports, industry and government standards, government documents and web-based resources.[11] Kirkwood's findings

showed that roughly 40% of the resources cited in all theses and dissertations in civil engineering in the period 2003-2004 were grey literature. Kirkwood noted that this result correlated well with a finding of 41% grey literature citations in a study of engineering graduate student Masters' theses at Mississippi State University reported by Virginia K. Williams and Christine Lea Fletcher.[13] She then checked the citations against library holdings as a way of informing future collection development decisions.

In a similar vein, Edward J. Eckel reported on a citation study of Masters' theses and Ph.D. dissertations completed at Western Michigan University's College of Engineering and Applied Sciences for the years 2002-2006. Eckel analyzed 2,903 citations from 96 Masters' theses and 2,886 citations from 24 Ph.D. dissertations with the primary goal of learning whether one could demonstrate an improvement in graduate student research competence by consideration of choice of resources cited. Eckel found that Masters' candidates cited grey literature more frequently than Ph.D. candidates, and that "Grey literature was more highly cited in the civil engineering theses than in all other disciplines." Eckel's analysis contains classifications for several formats that other researchers have grouped under grey literature, as well as a grey literature classification (e.g. separate classifications for government documents, patents, technical reports, and standards). If these findings were combined, Eckel's data would indicate that, for all Masters' theses in all majors in the College of Engineering, citation of grey literature was roughly 25-30% of all citations in the theses. Eckel was also able to show that Ph.D. dissertations contained more journal articles, fewer web sites, and on average had older resources over a longer span of time than the Masters' theses, perhaps reflecting a more thorough research and literature review.[5]

Research Questions

Citation analysis takes time and expertise, either brought to the study or developed by training, in order to readily identify, interpret and classify different types of citations from a variety of publication standards. For this reason alone, few citations studies look at literature beyond a small set of works. For example, Musser and Conkling's 1996 study reported a citation study on all of engineering by selecting one issue from one resource for each of 16 different discipline areas.[8] The research trade-off for this wide-spectrum approach was a small sample size from each discipline area, and the accompanying inherent risk of bias in results drawn about any individual discipline. In fact, Musser acknowledged in her 2007 study of mining engineering publications that, based upon this more discipline-specific study, the journal chosen for use in 1996 was not representative of the mining engineering field.[9] These results point to a need for more discipline-specific citation studies to understand if there are important differences in resource use for engineering librarians to factor into collection decisions. However, the literature of mining was shown to be similar to the literature of the rest of engineering as determined by the earlier study in terms of format and age of citations.

Kirkwood and Eckel, in their studies, used theses and dissertations produced at their academic institutions, thereby having a closed sample set of materials and an (appropriate and justifiable) institutional and academic focus rather than a professional-discipline scope. This makes their studies valuable for their particular institutions, but raises questions as to whether results found there are applicable to academic literature on the discipline more generally.[5,7]

Engineering students and faculty want full-text, electronic access to the resources they require for research. However, citation practices have been slow to adapt and provide adequate citation formats to indicate that resources were accessed via the web. Is it possible to study citations and determine the adoption rate of online resource access?

This research paper examines the following questions, through a study of citations in the academic literature of civil engineering:

1. Are the relative percentages of materials cited in different formats and the age of citations for civil engineering different from the results in Musser's 1996 study of all engineering fields and Musser's 2007 study of mining engineering?
2. Do the findings of high grey literature use by civil engineers in university Masters' theses hold true for the wider academic literature of the subject?
3. How many citations reflect that the materials were accessed via the web?

The organization and collection of data for this citation study utilized a system that can hopefully be utilized to gather similar data from other engineering disciplines, for comparative analysis purposes.

Methodology

In order to create a representative sample of civil engineering journal articles, six journals were selected using data from *Journal Citation Reports* (*JCR*).[6] The journals selected ranked in the top 15 journals for the subject as determined by impact factor and number of citations in *JCR*. While acknowledging that impact factor provides an imperfect metric, journals with high impact factor do tend to have many articles cited by researchers in the field, thereby giving some level of assurance that the journals and articles sampled for this study were representative of the research practices in civil engineering. In addition, choosing journals with a large number of citations should ensure a relatively large sample size of citations for the study. Journal issues were chosen from the May 2008 time period, selecting the most recently available issue in that time period, given the journals' publication schedules.

Journal Name	ISSN
Building and Environment	0360-1323
Coastal Engineering	0378-3839
Energy and Buildings	0378-7788
Journal of Hazardous Materials	0304-3894
Journal of Hydraulic Engineering - ASCE	0733-9429
Journal of Hydrology	0022-1694
Transportation Research Part B - Methodological	0191-2615

Each citation from a research article in the May 2008 issues of these journals received a code identifier in the spreadsheet used for data analysis. This code identifier links the journal to the article in the issue and the citation within the article, facilitating verification or validation of the

data. This coding format should also assist in sharing data with other researchers and enable the data gathered to be used in future studies.

The citation formats were classified mostly in accordance with the categories established in Musser and Conkling, with some elaboration.[8] The following interpretations of the classification terms were used:

Monograph – book resources, including textbooks and reference books

Journal – any citation for a regularly-scheduled serial publication

Conference – any citation for a conference proceedings, workshop, seminar, etc.

Standard – any citation for a standard, regulation, or recommended guideline from a recognized organizing body (government and non-governmental)

Technical Report – any citation from a report, briefing, working paper, research notes, etc. obtained from either a governmental or non-governmental organizing body

Patent – any citation for a government-issued patent instrument

Theses/Dissertations – PhD Dissertation or Master's Thesis

Software / Software Manual

Product Literature – any citation for promotional product literature (company white papers were classified under Technical Report)

Unpublished – usually personal correspondence or lab notes

Unknown

In addition to these exclusive categories, a non-exclusive category identified any citation that indicated a web presence, or the possibility of electronic/web-based retrieval of the reference.

The age of each citation from the date of publication of the article was calculated and recorded.

Results and Analysis

From the six journals, 150 articles were studied, yielding 4,172 citations. The articles in the current study therefore averaged 27.81 citations per article, as compared to the 22.55 citations per article averaged in the Musser and Conkling study of 1994 articles (14 years difference in publication date). The number of citations per article varies widely, from a low of 3 citations to a high of 62 citations. This finding suggests that civil engineering articles now contain more citations on average than engineering articles at the time of Musser and Conkling's study.

Further study of the citation data, looking for patterns and differences when comparing citations at the article level, is warranted.

The following table presents results for citations by format type. The first column represents a percentage based upon all citations. The second column is normalized for journal title, meaning that the proportions of citations by format from each journal are weighed equally. Because of differences in the number of articles in each journal sampled, this normalization was explored in hopes of presenting a better picture of the proportional distribution of format use across the discipline. However, the results do not change dramatically under the normalization process.

	Civil Engineering Journals	Civil Engineering Journals - Normalized	Kirkwood - Civil (2009)	Musser - Mining (2007)	Musser and Conkling (1996)	Eckel (2009)
Journal Article	70.7%	64.9%	38.2%	41%	53%	36%
Monograph	10.4%	11.9%	14.6%	18%	12%	21%
Conference Paper	6.2%	9.0%	15.4%	22%	19%	13%
Technical Report	5.9%	7.5%	10.1%	12%	9%	2%
Standard	1.8%	1.4%	2.4%			
Theses/Dissertations	1.2%	2.4%		3%	3%	2%
Software/Manual	0.9%	0.9%				
Product Literature	0.8%	0.7%				
Unknown	0.7%	0.7%				
Unpublished	0.3%	0.6%				
Patent	0.05%	0.03%				

Notes:

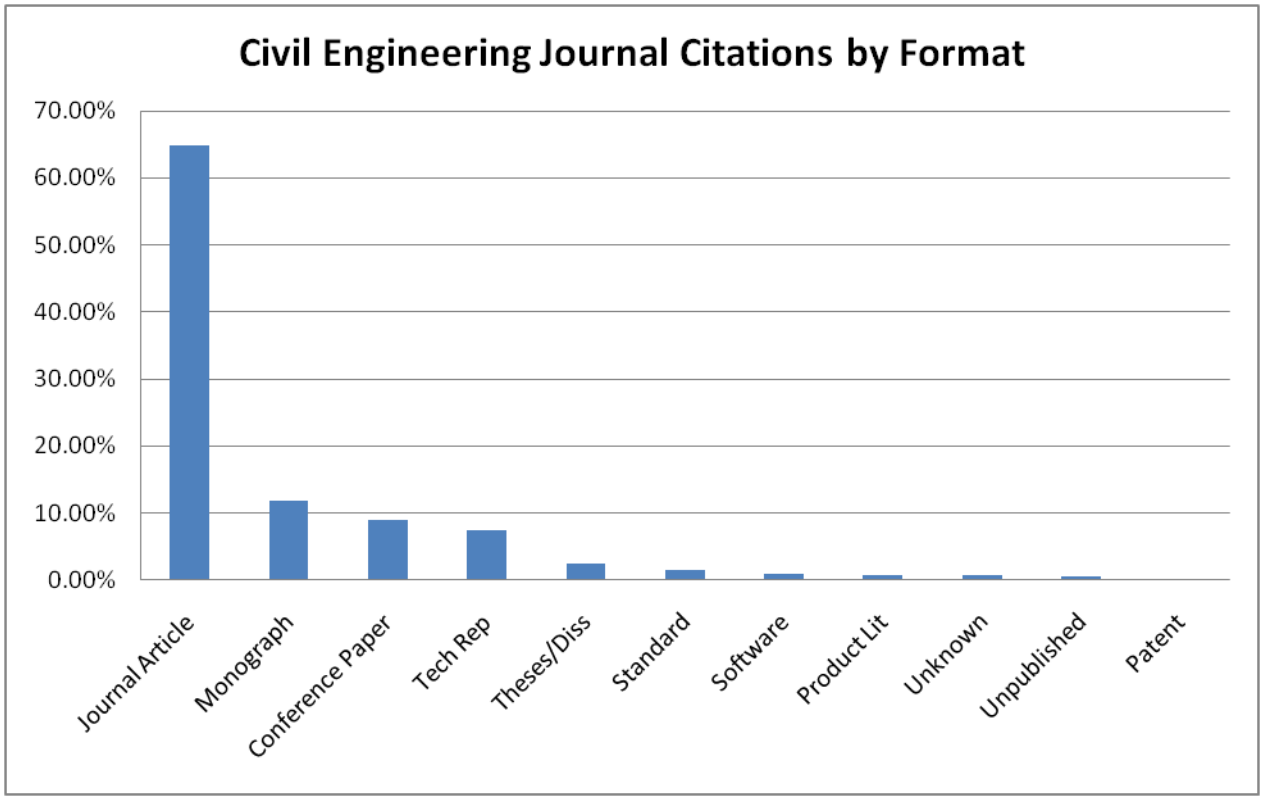
- Publication year given in table
- The data from Kirkwood are for both theses and dissertations[7]
- Only data from corresponding categories presented; not every column adds to 100%
- The value expressed for "Journal Articles" from Eckel is a summation of the data gathered in his categories of "Scholarly Journal," "Trade Magazine," and "Popular Magazine."[5]

Citations appearing in civil engineering journals differ appreciably in format type from the citations in the mining engineering journal citation study by Musser.[8] Mining engineering authors cited more monograph and conference paper resources, while civil engineering authors cited a much higher percentage of journal articles.

In comparison with Kirwkood's study of civil engineering theses and dissertations, the civil engineering journal article authors cite other journal articles more frequently than do the students writing theses and dissertations, while citing fewer conference papers and somewhat fewer technical reports. However, Kirkwood included a category for Government Documents in her format category analysis.[7] Those documents were classified as technical reports in the present study. If those government documents counted separately by Kirkwood were factored into the

comparison, citation of technical reports in civil engineering would appear to be much more prevalent in theses and dissertations than in the professional literature.

The results for category format from the second column of the above table are presented visually in the graph below:

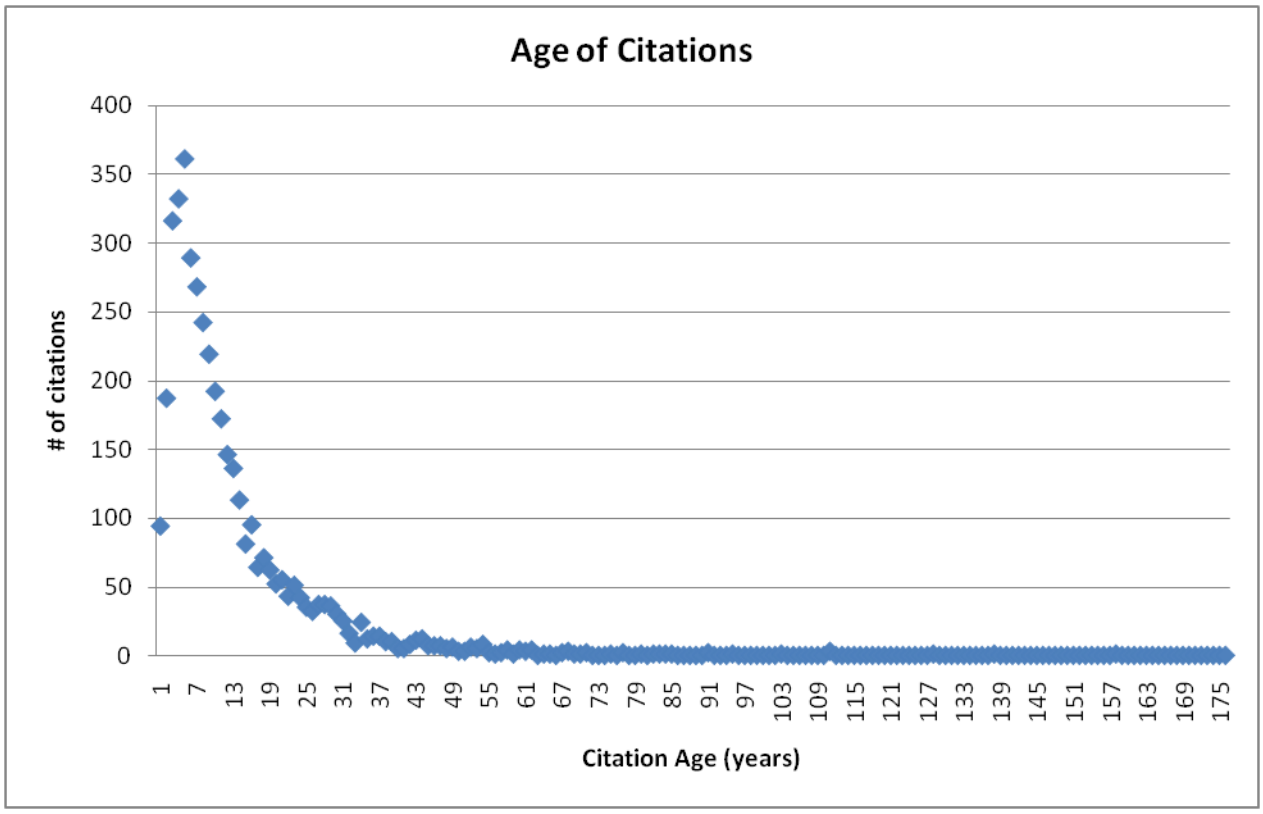


By examining the rate of citation of different format types for each journal, it became clear that there exist significant differences in citation of resources within sub-disciplines of civil engineering. For example, all of the citations involving standards came from three journals – *Building and Environment*, *Energy and Buildings*, and *Journal of Hazardous Materials*. These journals showed a marked difference from each other as well in their citation format types.

	Civil Engineering Journals	Building and Environment	Energy and Buildings	Journal of Hazardous Materials
Journal Article	70.7%	47.1%	55.3%	85.5%
Monograph	10.4%	17.5%	8.8%	7.5%
Conference Paper	6.2%	11.9%	10.8%	1.5%
Technical Report	5.9%	9.4%	11.9%	1.8%
Standard	1.8%	3.2%	4.8%	1.6%

The authors in *Journal of Hazardous Materials* cited conference papers and technical reports much less frequently than authors in the other civil engineering journals sampled, while citing the journal literature much more frequently than the authors in the other journals sampled. Also, authors in the two buildings-related journals cited technical reports and conference papers at roughly twice the frequency seen when considering the entire population of civil engineering citations.

The age of citations in the civil engineering journal can be graphed into the following curve:



The "long tail" of citation age in the sample extended to an oldest citation of 157 years (i.e. an 1851 publication date). The peak of the graph, the most frequent age of cited material, was 4 years.

The following table shows the rates of aging (measured by the age where more than *x* percent of citations are that age or newer)

Civil Engineering Journals	50% of All Citations	75% of All Citations	90% of All Citations
Journal Articles	7 years	13 years	23 years
Monographs	15 years	25 years	33 years
Conference Paper	8 years	14 years	22 years
Technical Reports	7 years	18 years	29 years
All Formats	7 years	14 years	25 years

These results are very similar to the results for all fields of engineering from Musser and Conkling's 1996 study.[8]

Musser and Conkling (1996)	50% of All Citations	75% of All Citations	90% of All Citations
Journal Articles	8 years	16 years	25 years
Monographs	11 years	19 years	30 years
Conference Paper	5 years	9 years	16 years
Technical Reports	7 years	16 years	32 years
All Formats	7 years	15 years	25 years

Also, the aging rates for mining engineering citations in Musser's 2007 study (50% 8 years or less, 75% 16 years or less, 90% 28 years or less) correlate well with the findings for civil engineering journals.[9]

Together, these results show that while engineers cite older resources in their work at a significant rate (10-12% being 25 years old or more), the majority of citations in articles are from materials less than 10 years old in civil and mining engineering. These field-specific findings are supported by the earlier Musser and Conkling study that examined engineering literature citations generally.

In Kirkwood's paper she reported that 40% of the citations from the Master's Theses and Ph.D. dissertations of civil engineering graduate students could be classified as grey literature.[7] For the current study, grey literature included the following format type categories: Technical Report, Standard, Theses/Dissertations, Patents, Software/Software Manual, Product Literature, Unpublished, and Unknown. These format types accounted for 13% of all the citations in the study. This proportion appears to be in closer agreement with the proportion that would be calculated for grey literature from Musser and Conkling's study of journal literature than the proportion found in Kirkwood's study of theses and dissertations. Based upon citation choice, this comparison demonstrates that civil engineering graduate students cite a different mix of information resources in conducting and communicating their academic research than authors writing for the civil engineering journals cite in their research reporting.

During the collection of data, any citation that contained a URL or indicated that the resource was obtained via the web was noted. Unfortunately, citation in 2008 had not completely caught up, in terms of standardization and in practice, with the widespread use of materials accessed via the internet. Of the 4,172 citations in the study, only 107 (2.5%) were cited as from an e-resource. This researcher believes that this number must represent a severe under-reporting. It is hoped that, in future years, a citation study will be able to quantify the extent of the use of electronic resources in engineering research.

Conclusions

This study sought to provide a quantitative analysis comparing citations from recent civil engineering journals with results presented in earlier papers about citations in engineering journals and theses and dissertations. Several key results were found:

- The average number of citations per article for civil engineering journals from 2008 was higher than that reported for citations for engineering journals from 1994 by Musser and Conkling by 5 citations per article, on average.
- Authors of civil engineering journal articles cited different proportions of journals, monographs, and conference proceedings than did the authors of mining engineering journal articles reported by Musser in 2007.
- Within the study, journals related to specific sub-fields (e.g. buildings and hazardous materials) showed markedly different citation patterns in terms of format type than the pattern for the civil engineering journals considered as a whole.
- The age of materials cited by authors in engineering journals has shown little change in the last 14 years.
- Authors of papers in civil engineering journals cite grey literature significantly less frequently than did the civil engineering graduate students in Kirkwood's and Eckel's studies of theses and dissertations.
- As of the date sampled for this study, citation practice for e-resources was not standardized and well-adopted. This may be a prime factor in the low percentage of citations recognized as being accessed through the web.

Future research directions based upon the findings of this study include examining another engineering field as a further comparison with the current data, conducting an article-level analysis to determine if characteristic styles of information use can be inferred from citation study, and examining citation practices for web resources over time.

Bibliography

1 Bush, Vannevar. (1945). As we may think. *The Atlantic Monthly*, July 1945. Accessed at http://www.theatlantic.com/magazine/archive/1969/12/as-we-may-think/3881/

2 Cameron, Brian D. (2005). Trends in the usage of ISI bibliometric data: Uses, abuses, and implications. *portal: Libraries and the Academy* 5 (1), 105-125.

3 Garfield, Eugene. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science* 122 (3159), July 15, 1955, 108-111. Accessed at http://www.garfield.library.upenn.edu/essays/v6p468y1983.pdf

4 Garfield, Eugene. (1999). From the world brain to the Informatorium. *Information Services & Use* 19 (2), 99-105.

5 Eckel, Edward J. (2009). The emerging engineering scholar: A citation analysis of theses and dissertations at Western Michigan University. *Issues in Science and Technology Librarianship*, 56 (Winter 2009). Accessed at http://www.istl.org/09-winter/index.html

6 *Journal Citation Reports*. (2008). Thomson Reuters.

7 Kirkwood, Patricia. (2009). Using engineering theses and dissertations to inform collection-development decisions, especially in civil engineering. *American Society of Engineering Education 2009 Annual Conference and Exposition, paper AC 2009-140*, Austin, TX, June 14-17.

8 Musser, Linda R. and Conkling, Thomas W. (1996). Characteristics of engineering citations. *Science & Technology Libraries*, 15 (4), 41-49.

9 Musser, Linda. (2007). A study of references in mining engineering publications. *Issues in Science and Technology Librarianship*, 49 (Winter 2007). Accessed at http://www.istl.org/07-winter/index.html

10 Tenopir, Carol, and Donald W. King. (2004). *Communication patterns of engineers*. New York: John Wiley & Sons, Incorporated.

11 Thomspon, Larry A. (2001). Grey literature in engineering. *Science & Technology Libraries*, 19 (3), 57-73.

12 Wells, H.G. (1938). *World brain*. Garden City, NY: Doubleday, Doran, & Co.

13 Williams, Virginia K. and Christine Lea Fletcher. Materials used by Masters' students in engineering and implications for collection development: A citation analysis. *Issues in Science and Technology Librarianship*, 45 (Winter 2006). Accessed at http://www.istl.org/06-winter/refereed1.html