

Use of Information Resources in Writing up Dissertation: A Citation Analysis

Szu-chia Lo Associate Professor
Graduate Institute of Library and Information Science
National Chung-Hsing University (Taiwan)
losu@dragon.nchu.edu.tw

Abstract - Referencing relevant works in one's own work to conceptualize ideas, enrich work and get support for the arguments in the scientific writings is commonly seen in scholarly communication. Citation analysis is a method applied often to reveal the information behavior, construct the possible network among scholars and identify the research trends to enhance the information services. In this study, the author examined the dissertations written by the PhD students from Life Science related departments of National Chung-Hsing University in Taiwan and the references cited in those dissertations by taking the Citation Analysis approach to reveal the trends of student's research environment.

Index Terms - Citation analysis, Information resources, Dissertations, Biotechnology research.

INTRODUCTION

Referencing relevant works in one's own work to conceptualize ideas, enrich work and get support for the arguments in the scientific writings is common practice in scholarly community. Citation analysis has a history going back to the 1920s of using as a tool for library collection management, and measuring usage of resources by libraries [1]-[3]. Recently, there were attempts to use citation analysis for various purposes. Citation analysis of students' works presents as a way to quantify students' information behavior involved with library collection and resources [4]. The cited references in students' work not only present the availability of information resources, but also show the origins of students' research and the ability of selecting and retrieving relevant sources. Citation analysis is a method applied often to reveal the information behavior, construct the possible network among scholars and identify the research trends to enhance the information services.

Reviewing the studies done in the past, it is found that total number of citations and the media types of cited materials were often examined in the studies [5]-[6]. Other issues, such as currency of the citations and quality of the citation were also common in relative studies [7]. In this study, the author examined the dissertations written by the graduate students from Life Science related departments of National Chung-Hsing University in Taiwan and the references cited in those dissertations by taking the Citation Analysis approach. Dissertation is the physical manifestation of the learning results of PhD students and the references cited in the dissertations not only present the origins of students' researches but also show students' information

literacy in retrieving, accessing and selecting information resources. The author took this approach to reveal the characteristics of citing behavior of the PhD students, including what information formally used when writing the dissertations and what the characteristics of the information were, such as age of cited references, language of the materials, types of the information media (monographs, journals, technical reports, etc.) and geographic distribution of cited references. A total of 1,791 cited references from 13 dissertations of 2005 and 2006 were examined in this study. The author sought to investigate any trends in the PhD students' research environment, such as material types, publication dates and languages of cited works, the distribution of institutions and geographic area of cited authors were examined for the highly cited materials. The availability of cited references in university library collection was further inspected to show the efficiency of library services.

METHOD AND DATA

The author took citation analysis approach in this study. Citation count was the core method used. The results were further examined by the material types, languages, authors and sources of the cited references. The citation age, the duration from the time when the cited materials published to the time they were cited, was calculated to show the length of influential period of the research works. The dissertations related to the biotechnology research and issued in 2005 and 2006 were chosen for this study. The bibliographic information of the dissertations was downloaded from National Chung-Hsing University - Electronic Theses and Dissertations Services (NCHU-ETD, www.lib.nchu.edu.tw/lib_main/etds/etds_index.htm). A total of 14 dissertations were retrieved via the system and one dissertation obtained through the system did not include the necessary information for analysis, which was omitted from the study. 13 dissertations and 1,791 cited references were examined in this study.

RESULTS

13 dissertations, 7 dissertations were written by the students from College of Life Sciences and 6 were by the ones from the College of Agriculture and Natural Resources, and 1,791 cited references, the unique citations were 1,689, were examined in this study. The subjects covered in the dissertation included cell transformation, preparation of protein, enzyme and peptide, treatment of disease, and study

on bacterial. DNA recombinant, introducing gene materials were the techniques used in the research. Examining the cited references, it was found that the shortest list of references contained 27 citations and the longest one contained 278 citations, with the median being 123, only few references were co-cited. With various material types covered in the cited references, most of the cited references were journal articles. There were two students cited journal articles only. The ratio of citing journal articles was over 90%. Table 1 shows the basic information of each dissertation and the results of citation count.

TABLE I
RESULTS OF CITATION COUNT

College	Department	S*	Year	No. Ref.	Unique Ref.	No. J. Paper	% of J. Paper
Life Sciences	Life Sciences	T	2006	73	72	72	98.63
	Life Sciences	E	2005	50	33	48	96.00
	Life Sciences	C	2005	278	278	252	90.65
	Life Sciences	T	2005	219	219	212	96.80
	Life Sciences	C	2005	101	101	83	82.18
	Biochemistry	E	2005	212	210	190	89.62
	Biochemistry	P1	2005	42	42	39	92.86
Agriculture and Natural Resources	Biotechnology	P2, V	2006	233	219	215	92.27
	Biotechnology	B	2006	249	191	216	86.75
	Biotechnology	P2, V	2005	43	35	35	81.40
	Biotechnology	P2, V	2005	123	123	110	89.43
	Biotechnology	P2, R	2005	27	27	25	92.59
	Biotechnology	P2, R	2005	n/a	n/a	n/a	n/a
	Biotechnology	T	2005	141	139	133	94.33

* Bacterial (B), Enzyme (E), Peptide (P1), Protein (P2), Cell transformation (C), Treatment of disease (T); DNA Recombinant (R), Introducing Gene Materials via Vector (V)

English is the major language of the cited references, 99% of the literatures were written in English. The average citation age was 8.33, the citation ages of 52.76% of the cited references were under 6 (include), 70.35% were under 10 (include) and 92.57% were under 20. With the average citation age 8.33, there was 1.9% of the cited references were published over 30 years when they were cited by the dissertations. The earliest issued literature cited was a book, "The Vegetable Proteins", which was published in 1924. This book was cited by the author whose dissertation was about introducing genetic materials into rice. Besides citing earlier works, 8.21% of cited worked were published at the same year of the dissertation did. Comparing the citation ages of cited materials of dissertations by students from different colleges, Life Science and Agriculture, the average citation age of dissertations of the former discipline was 7.96 and 8.77 for the latter discipline. The median citation ages for both disciplines were 6. Besides the students in the

discipline of Life Science cited more works that were published at same year of they finished their dissertations, there was not significant difference of distribution of citation ages among the dissertations from these two disciplines. Figure 1 shows the distribution of the citation ages.

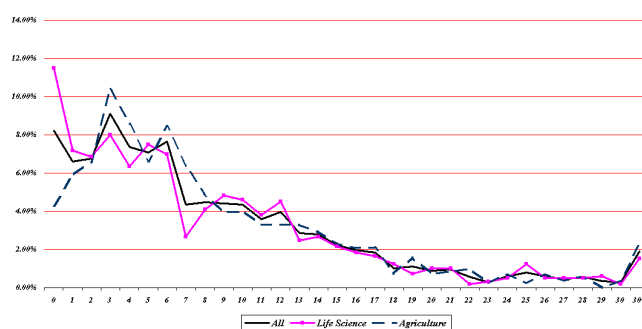


FIGURE 1
DISTRIBUTION OF CITATION AGES.

I. Composition of material types of cited references

Journal article was the most cited material type found in this study although books, lab manual, journal articles and theses were also seen in the cited references. 90.95% (1,629) of cited references were journal articles. Only one author cited significant amount of literatures other than journal articles. 17.82% of his cited references were technical reports, thesis, books and proceeding articles. The only two Internet resources found cited by the dissertations covered in this study were also cited by this particular author mentioned above. Although significant amount of electronic journals were available during the period of writing up the dissertations, there was no evidence to distinguish if the journal articles cited were in printed format or electronic format since all the bibliographic data of the citations were in the format for printed journal articles, unless the journals were electronic only in library collection (assuming the students obtained copies of reference materials from library collection). The author checked the journal titles against the library collections and found library supported both formats for most of titles. It was highly possible that the electronic journals were the used resources for the easy accessibility. Figure 2 shows the composition of the material types of cited references.

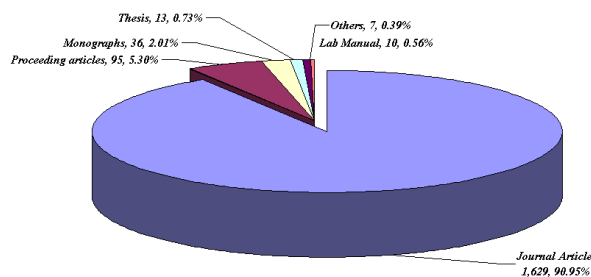


FIGURE 2
COMPOSITION OF MATERIAL TYPES OF CITED REFERENCES.

II. Highly cited journals

Further examined the 1,629 journal articles cited, 414 journals were identified. Among them, *Journal of Biological Chemistry*, *Journal of Bacteriology*, *Plant Physiology*, *Biochemistry*, *Science*, *Nature*, *Cell*, *Journal of Virology* and *Molecular and Cellular Biology* were the 10 most cited journals. Comparing the highly cited journals of two disciplines, *Journal of Biological Chemistry*, *Biochemistry*, *Molecular and Cellular Biology*, *Science*, *Nature*, *Cell* *Applied Environmental and Microbiology* *Journal of American Chemical Society* were the influential journals in dissertations done by the graduates of College of Life Sciences. For the ones from College of Agriculture, *Plant Physiology*, *Journal of Bacteriology*, *Journal of Virology*, *Plant Cell*, *Molecular Microbiology*, *Plant Journal*, *Arthritis Rheum*, *Science*, *Molecular Plant-Microbe Interactions* and *Nature*. Table 2 lists the top 10 cited journals and the times cited.

TABLE II
HIGHLY CITED JOURNALS

All Journal Citations			Life Sciences			Agriculture		
Journal	Cited	%	Journal	Cited	%	Journal	Cited	%
J Biol Chem	92	5.65	J Biol Chem	78	8.71	Plant Physiol	55	7.50
J Bacteriol	62	3.81	Biochemistry	47	5.25	J Bacteriol	49	6.68
Plant Physiol	57	3.50	Mol Cell Biol	32	3.57	J Virol	36	4.91
Biochemistry	49	3.01	Nature	31	3.46	Plant Cell	28	3.82
Science	48	2.95	Science	31	3.46	Mol Micro	23	3.14
Nature	46	2.82	Cell	30	3.35	Plant J	22	3.00
Cell	39	2.39	Appl Env Micro	22	2.46	Arthritis Rheum	21	2.86
J Virol	37	2.27	J Am Chem Soc	16	1.79	Science	17	2.32
Mol Cell Biol	34	2.09	EMBO J	15	1.67	MPMI	16	2.18
Plant Cell	28	1.72	J Mol Biol	15	1.67	Nature	15	2.05

Except a few of highly cited journals with multi-disciplines features were cited heavily by the works done in both subject area, such as *Science* and *Nature*, it is clear that there were two sets of highly cited journals used by the PhD students from two colleges. Some titles were even cited by sole citing dissertation, such as *Arthritis Rheum*, *Journal Bacteriology* and *Journal of Virology*, all cited by particular dissertation only due to the differentiations and diversities of the research topics. Besides the journals, one of the works categorized as proceeding in this study, *Proceedings of the National Academy of Science*, was also highly cited. This title could be seen as a journal title in a broader sense. Comparing to other highly cited journal titles, *Proceedings of the National Academy of Science* was heavily used for its multi-disciplinary characteristics. It was cited 94 times, only 3 dissertations did not cite articles issued in this title.

III. Highly cited authors

Examined the first authors of the highly cited works, 350 authors were identified from the 425 works. Most of the authors were cited with limited times, over 98% of the authors were cited more than three times. LS Kappen was the author who was cited most times, 7 times. Her works dated from 1979 to 1989 that were related to the molecular mechanisms of drug action (antitumor) were cited 7 times by the work relate to the research in treatment of cancer. Y.X. Xiao was the second most cited author. His works done with Hutcheson related to bacterial protein had impact on the research of "Pseudomonas syringae" done by one of the PhD students from College of Agriculture and Natural Resources. The works done by Deng, Ikemura and Povirk who were listed as first authors were cited 4 times and ranked at the third place of most cited authors. Deng's works about "Pseudomonas syringae" also presented similar impact on the work done by the same student, Ikemura's works mainly related of production of subtilisin and the works of DNA recombinant done by Povirk influenced the research related to the cancer treatment. Except Deng, Li and Lin, the highly cited authors were all based in the institutions in US when the articles were published; even some of the authors originally were from other countries, Kappen was from India, Xiao was from China, Chin and Tsai were both from Taiwan.

TABLE III
HIGHLY CITED AUTHORS

Author	Institution	Department	*Self cited	Times Cited
Kappen, LS	Harvard Medical School	Pharmacology		7
Xiao, YX	Univ. of Maryland	Botany		6
Deng, WL	Chung-Hsing Univ. (Taiwan)	Agricultural Biotechnology	✓	4
Ikemura, H	Robert Wood Johnson Medical School	Biochemistry		4
Povirk, L	Harvard Medical School	Pharmacology		4
Cambronne, ED	UCLA School of Medicine	Micro. Immune. & Mole. Genetics		3
Charkowski, AO	Cornell Univ.	Plant Pathology		3
Chin, DH	Harvard Medical School	Biol. Chem. & Mol. Pharm.	⊙	3
Dedon, PC	Harvard Medical School	Biol. Chem. & Mol. Pharm.		3
Gao, XL	Glaxo Research Institute	Structural and Biophysical Chem.		3
Huang, HC	Cornell Univ.	Plant Pathology	⊙	3
Li, YI	Chung-Hsing Univ. (Taiwan)	Agricultural Biotechnology	✓	3
Lin, LJ	Chung-Hsing Univ. (Taiwan)	Agricultural Biotechnology	✓	3
Tsai, CH	Oregon State Univ.	Genetics	⊙	3

* Authors and the PhD students were at the same institution when the articles were published.

⊙ The authors are currently working for the same institution.

Examining the co-authors of the works done by the authors listed in table 3, the record showed the team efforts of the research works. Irving Goldberg, the chairman of Department of Pharmacology of Harvard Medical School from 1972 to 1986 was co-author of several cited references by Kappen, Chin and Dedon. SW Hutcheson from University of Maryland was co-author with works done by Xiao and HC Huang was teamed up with WL Deng.

III. Availability of cited works

The availability of the highly cited titles in library collection was also done in this study. The ISSNs and journal titles were used for searching against the library online catalog to see if the library has subscriptions to those titles. The holdings and format of the journals were also checked. 21 journals, which were cited at least 15 times, were examined in the study. It was found that all the titles could be found in the library current subscriptions. 15 titles of them were both in printed and electronic formats; the users could either access the titles via campus network or obtain hard copies at the library. 3 titles are in printed format only and 2 are in electronic format only in the library collection.

OBSERVATIONS AND CONCLUSION

I. Impact of Journals on the research

From the heavily cited journals articles, over 90% citations were journal articles, it was obvious that journal had significant influence during the progress of researches and writing up the dissertations.

II. Core journals in research topics

Reviewing the topics of the dissertations and the materials cited in their works, especially the journal articles, several core journals related to various research topics could be identified, such as *Journal of Bacteriology*, *Journal of Virology*, and *Plant Physiology*, besides the journals which cover multi-disciplines that had impact on the research, such as *Proceedings of the National Academy of Science*, *Journal of Biological Chemistry*, *Science* and *Nature*.

III. Share references were rare

Comparing the materials cited in the dissertations included in this study, few shared references were found. Even with the dissertations, which focus on the researches in the similar area, such DNA recombinant, introducing genetic materials and preparation of protein, share references were very uncommon.

IV. Information sources

Not only the individual literatures were rare co-cited, only a few journal titles were co-cited found in this study. The ones with multi-disciplines coverage had better chances to be cited by multiple dissertations.

V. Link to the early works done in the same institution

With limited citing local works, it could still link the researches to the prior works done locally. Examining the authors of the cited references against the advisors and committee members, it was found that several dissertations showed the results of continuing the works done by the members from the same institution especially the works done by the advisors.

VI. Availability of literature in the library collection

The highly cited works had very high accessibilities in local library collection, either in printed, electronic or both formats. The high availability was not necessary presenting the good performance of library on providing needed information since the students might choose the materials based on the availabilities of the materials locally rather than the properness. The library is still worth the credits for high availability and accessibility of the cited materials.

This study with focus on the PhD students with interests in biotechnology research tried to reveal any possible trends in the students' research environment. Journal articles, English materials, journals with special research interest were the primary sources that influenced the progress of research. Traces of the prior art done by the faculty members in the same institution could be found in the cited references. By doing more examining of the citing behavior, the better understanding of the information behavior could be shown. Further investigations for the theses and dissertations in other disciplines are worth to take to enclose more insights of students' information behavior.

REFERENCES

- [1] Reba Leiding, "Using citation checking of undergraduate honors thesis bibliographies to evaluate library collections," *College & Research Library*, Vol. 66, No 5, pp. 417-429, 2005.
- [2] Linda C. Smith, "Citation analysis," *Library Trends*, Vol. 30, No. 1, pp. 83-109, 1981.
- [3] Erin T. Smith, "Assessing collection usefulness: an investigation of library ownership of resources graduate student use," *College & Research Libraries*, Vol. 64, No. 5, pp. 344-355, 2003.
- [4] Anne Middleton, "An attempt to quantify the quality of student bibliographies," *Performance Measurement and Metrics*, Vol. 6, No. 1, pp. 7-18, 2005.
- [5] Phillip M. Davis and Suzanna A. Cohen, "The effect of the web on undergraduate citation behavior, 1996-1999," *Journal of American Society for Information and Technology*, Vol. 52, No. 4, pp. 309-314, 2001.
- [6] Holly Heller-Ross, "Assessing outcomes with nursing research assignments and citation analysis of student bibliographies," *The Reference Librarian*, No. 77, pp. 121-40, 2002.
- [7] Virginia E. Young, and Linda G. Ackerson, "Evaluation of student research paper bibliographies: refining evaluation criteria," *Research Strategies*, Vol. 13, No. 2, pp. 80-93, 1995.