Evidence Based Library and Information Practice

Evidence Summary

Thematic Categorization and Analysis of Peer Reviewed Articles in the LISA Database, 2004-2005

A Review of:
Gonzalez-Alcaide, Gregorio, Lourdes Castello-Cogolles, Carolina Navarro-Molina, et al. "Library and Information Science Research Areas: Analysis of Journal Articles in LISA." Journal of the American Society for Information Science and Technology 59.1 (2008): 150-4.

Reviewed by:
Carol Perryman
TRLN Doctoral Fellow, School of Information & Library Science, University of North Carolina at Chapel Hill
Chapel Hill, North Carolina, United States of America
E-mail: cp1757@gmail.com

Received: 05 December 2008 **Accepted:** 10 February 2009

Objective – To provide an updated categorization of Library and Information Science (LIS) publications and to identify trends in LIS research.

Design – Bibliometric study.

Setting – The Library and Information Science Abstracts (LISA) database via the CSA Illumina interface.

Subjects – 11,273 item records published from 2004-2005 and indexed in LISA.

Methods – First, a search was set up to retrieve all records from 2004-2005, limited to peer review items (called "arbitrated works" by the authors (150)) and excluding book reviews. Second, thematic descriptor terms used for the records were identified. Frequency counts for descriptor term occurrence were compiled using Microsoft Access and Pajek software programs. From the results of this search, the top terms were analyzed using the Kamada-Kawai algorithm in order to eliminate descriptor term co-occurrence frequencies under 30. A cluster analysis was used to depict thematic foci for the remaining records, providing a co-word network that visually identified topic areas of most frequent publication. Conclusions were drawn from these

findings, and recommendations for further research were provided.

Main Results – The authors identified 18 "thematic research core fields" (152) clustered around three large categories, "World Wide Web", "Education", and "Libraries", plus 12 additional peripheral categories, and provided a schematic of field interrelationships.

Conclusion – Domains of greatest focus for research "continue to be of practical and applied nature," (153) but include increased emphasis on the World Wide Web and communications technologies, as well as on user studies. A table of the most frequently occurring areas of research along with their top three descriptor terms is provided (Table 1, 152) (e.g., "World Wide Web" as the top area of research, with "online information retrieval" (268 occurrences), "searching" (132 occurrences), and "web sites" (115 occurrences)).

Commentary

In this brief research report, the researchers used co-occurrence and cluster analysis methods to find emergent foci for research in the LIS literature. The authors appear to have begun their research by making the assumption that all the articles indexed in LISA concern LIS literature, a concept quickly disproved. Among the article titles that comprise the dataset, this reviewer found the following:

> Schoech, D., J.D. Fluke, R. Basham, D.J. Baumann, and G. Cochran. "Visualizing multilevel agency data using OLAP technology: An illustration and lessons learned." Journal of Technology in Human Services 22.4 (2004): 93-111.
>
> Engin, M., O. Cidam, and E.Z. Engin. "Wavelet transformation based watermarking technique for human electrocardiogram (ECG)." Journal of Medical Systems 29.6 (2005): 589-594.
>
> Isik, H. "Design and construction of thermoelectric footwear heating system for illness feet." Journal of Medical Systems 29.6 (2005): 627-631.

While it is possible that the authors restricted their analysis to LIS areas of concern, it was not made explicit whether this was done. The query used in retrieval is documented as PY (Publication Year) = 2004 OR PY = 2005, restricted to peer reviewed works (called "arbitrated works" (150)), and only book summaries (book reviews) were excluded by design. This reviewer cannot assess the percentage of materials that are not obviously related to LIS research, but it seems likely that the findings are affected.

Additionally, although the authors describe their frequency analysis of 'descriptors,' in the connected table (Table 1,151), the terms are called 'keywords' ("choose the keyword option to search the title, abstract, and descriptors simultaneously")[1]. In the LISA database, key words are terms found in the title, abstract, descriptor, or identifier fields, while 'descriptors' are limited to controlled vocabularies (in this case the LISA Thesaurus).

[1]LISA CSA Illumina Help and Support: Keyword Search:
Descriptors, DE = This field contains indexing terms taken from a thesaurus or controlled vocabulary.

Keyword, KW = This search strategy simultaneously searches the Title (TI), Abstract (AB), Descriptor (DE), and Identifier (ID) fields. [CSA Illumina Help & Support: Fielded Search

In order to validate co-occurrence analysis, it would be helpful to consider how descriptor terms were assigned, in order to determine whether the terms are indicative of the research focus for the articles indexed. For example, a descriptor term such as 'meta-analysis' might equally be used to discuss the *process* for meta-analysis or to describe the research *method* used to conduct the indexed paper. Here, examination of a random sample of articles would have improved the quality of this paper. There is also no description of the methods used to assign thematic areas, which should include a definition of the area as well as a description of the exclusion and inclusion criteria. The ability to replicate this search is made more problematic by the use of thematic areas assigned without explanation.

Although the authors' uses of cluster analysis and co-occurrence frequency appear to be appropriate methods for examining a corpus of literature, the results are questionable due to the absence of consideration for database content coverage in the LISA database. Future research using these replicable methods and with a more carefully selected initial dataset would constitute a substantial contribution to our understanding of LIS research areas.

This work was published as a brief communication, so severe space constraints likely barred inclusion of further discussion or detail. However, the authors failed to provide a literature review, only mentioning other research in discussing the findings from their analyses. Some mention of the rational for methodologies used, especially in support of term occurrence as a justification for topicality (for example, Zipf's law of inverse proportion of term usage to topicality), as well as of the numerous previous attempts to characterize LIS output and to evaluate the LISA database would have better informed their approach to this important topic.

Critical evaluation of this brief research report was done using the University of Salford School of Nursing HCPRDU Evaluation Tool for Quantitative Analysis.

Works Cited

University of Salford School of Nursing. "Evaluation Tool for Qualitative Studies." HCPRDU. University of Salford. 30 November 2008. <http://www.fhsc.salford.ac.uk/hcprdu/tools/quantitative.htm>.

Zipf, George K. Human Behavior and the Principle of Least-Effort. Cambridge, MA: Addison-Wesley, 1949.