

An Introduction to Research and the CS/IS Professional Literature for Undergraduates

Sally Jo Cunningham
Department of Computer Science
University of Waikato
Hamilton, New Zealand
email: sallyjo@waikato.ac.nz

Abstract

One of the more difficult aspects of computing to incorporate into the undergraduate curriculum is an authentic research experience. This paper describes a successful third year project in which computing students designed, conducted, and wrote up bibliometric experiments. This project gives students a flavor of the scientific method, and has the added benefit of encouraging familiarity with the scientific publishing process and with the computing and information systems professional literature.

Introduction

Exposure to research, and particularly active participation in research, is believed to be an effective way to encourage undergraduates to continue education in the sciences and engineering [EN93]. Computing is an experimental science, like physics or chemistry; students benefit by learning the scientific method and coming to an understanding of how it applies to computer science. And, of course, research can allow students to experience the excitement of discovery and the satisfaction of adding to the body of knowledge in a field.

This paper describes a bibliometrics project assigned to a third year database course. Bibliometrics (described in more detail below) is the study of how scholars in a field use and disseminate knowledge both formally (through journals, books, and conference proceedings) and informally (through letters, conversations, and now over the Internet). Students designed

and conducted original bibliometrics experiments, the best of which are being revised for submission to journals or conferences. The goals of this project were:

- to teach students about the scientific method—formulating a hypothesis, designing an experiment to test the hypothesis, and effectively presenting results—in the context of an authentic research experience
- to provide an opportunity for students to gain an understanding of the academic publication process and the dissemination of scientific information
- to encourage students to gain familiarity with the community of scholarship for a field or topic of interest
- to teach students to effectively use the various abstracting and indexing publications in computer science and information systems

Previous work

Formal instruction in research methods generally occurs at the graduate level—for example, with a credit course or series of non-credit seminars on research skills [WI93]. Undergraduates may have the opportunity to participate in a research-oriented senior project. The course associated with this project sometimes provides an introduction to the scientific publishing process by requiring submission of results to a formal departmental conference or journal ([JA90], [CO92], [HA94]).

Aside from formal coursework, one-to-one or small group mentoring schemes are supported by the National Science Foundation through several initiatives, including the Research Experience for Undergraduates (REU) program. This program provides funding to support summer research for students in the sciences and engineering, and institutions report success in gearing their individual programs to increasing participation in research by women and minorities ([EN93], [LO94]).

Providing an authentic research project for undergraduates can be difficult, however; a summer or semester may be too short to bring an undergraduate up to speed in both an application area and in general research methods. In addition, the above programs are time-intensive for staff, and are generally available only for a small subset of the computer science student population.

An alternative approach is to provide instruction in research methods to large numbers of students through projects in a regular course, by assigning coursework that requires students to design and conduct small-scale mock "experiments" to explore the subject matter. Previous reported assignments include using this technique to teach experimental algorithm analysis [BE94], to explore program behaviour in first year CS labs [BA92], and to investigate the relative effectiveness of AI game-playing techniques [CU94]. The student experiments are conducted on an existing workbench, as it is impractical to require students to develop all the software as well as to design and execute the experiments.

Finally, familiarization with the computing literature and important literature searching tools is a common teaching objective. The usual approaches are to require a large "term paper", smaller essays summarizing one or more journal articles, or summary articles from popular periodicals (see, for example, [JA90], [HO88], [HA94]). The project described in this paper attempts to get a bit off this well-beaten track, and to encourage exploration of the production and dissemination, as well as the content, of CS/IS research documents.

Bibliometrics

Bibliometrics is a statistical or quantitative study of the characteristics of a body of literature ([BO90], [NI78]). Physical document collections or computerized bibliographic databases are studied not for the *content* of individual documents, but to determine trends or features of an academic discipline. Typical bibliometric research questions include:

- What is the rate of growth/decline of publications in a discipline?
- What are the qualifications, geographic distribution, etc. of scholars that compose a research community?
- What information about inter-relations between research groups or disciplines is reflected in citations between the groups/disciplines?
- How "important" is a particular researcher, research group, journal, etc.?
- How long does a paper/book stay relevant? When is it likely to become obsolete?

Research questions are tested by performing the appropriate queries on text or online bibliographic databases (eg, Compumath Citation Index, ERIC, Science Citation Index, etc), or by directly examining a set of documents (eg, performing obsolescence studies by examining the publication dates of references in a volume of a selected journal). Setting up a bibliometric experiment requires an awareness of the limitations of the chosen resources; for example, a trend analysis of the rate of growth for publications about "expert systems" can be skewed by the choice of bibliographic databases, yearly fluctuations in the coverage of these databases, database indexing schemes, etc. And, of course, analysis of the results requires that these limitations be carefully factored in, to determine an answer for the research question.

The project assignment

Students in a third year database course were asked to design, perform, and write up a bibliometric experiment. Students could work singly or in groups of two or three.

Class lectures included an introduction to bibliometric methods, the scientific publishing process, and how to use the common abstracting and indexing resources for CS/IS. To familiarize the students with the paper and CDROM indexes in the university library, the class was given a library "treasure hunt" exercise. A set of 12 bibliometrics journal articles were placed on desk copy for the students to use as examples of possible research topics, and 3 articles were set as required reading and discussed at length in class.

Finally, the scientific methodology—setting a research hypothesis, experiment design, and evaluation—was discussed in the context of bibliometric-type experimentation. A list of sample topics was distributed and discussed, and then the class was turned loose to design their projects. Students were required to submit a research proposal before beginning their work, so that I could help them fine-tune overly ambitious, inappropriate, or poorly thought out projects.

Student bibliometric studies

The 59 students in the course completed 26 bibliometric studies. Of these, only one student completely missed the point of the assignment and handed in a conventional summary-of-articles essay. Research topics tackled by the students included:

- determining the rate of obsolescence for the management communication, communications network, and information processing literature
- examining the impact of the introduction of 4GLs to the rate of references to 3GLs
- measuring the percentage of publications by men and women in management communications and accounting
- measuring the rate of increase in the number of journals published over the Internet (electronic journals)
- comparing the research performance of four US computer science departments

Most of the projects were closely based on existing bibliometric techniques as presented in course lectures: co-citation studies, subject growth rates, obsolescence measures, etc. A standard project took one of these techniques and applied it directly to a CS/IS subtopic of personal interest. Several students came up with interesting twists on the papers presented in class: for example, the analysis of gender ratios in management publications listed above.

Some of the studies were quite comprehensive, requiring hand analysis of thousands of bibliographic references. Others could be performed relatively quickly through broad queries on CDROMs. Some students learned that sheer effort isn't enough—that major undertakings can be ill-focused and unenlightening, whereas a well-constructed but simple investigation can yield interesting results.

Conclusions

The main objective of this course project was to provide students with an introduction to research methods. Several key factors influenced the choice of bibliometrics as a focus for the project:

- *opportunity for participation in authentic, cutting edge research*
Few bibliometric studies have addressed problems in computer science and information systems, so the students were able to directly reach the forefront of research.
- *quick start-up in research*
Bibliometric research does not rely on software or large systems; instead, it requires a basic understanding of scientific publishing and a few simple techniques. Most other research programs for undergraduates described in the literature either force the student to begin by constructing a sizeable piece of software, or are supported by a large, previously existing workbench or testbed. Use of a workbench may limit the types of research questions that can be formulated.
- *practice in formulating research questions*
The students were free to tackle problems of their own choosing, and were not constrained with artificial problems or by the need to first construct large software systems.

- *applicability of the topic to the course*
Nearly all the students in the database course were pursuing computing or information systems degrees. While bibliometrics is an unusual topic for these programs of study, it is by no means off-track; the lectures emphasized the advantages and disadvantages of different information retrieval interfaces, methods for measuring the effectiveness of retrieval systems, types of document-based information systems, etc.
- *familiarization with the scientific literature and publishing*
The project requires students to explore the literature in their chosen field, and provides a natural introduction to the scientific publishing process.

How well did the students do? Most performed reasonable assignments, and a few did sterling jobs. Four projects are being augmented or re-written for submission for publication. A final, unforeseen benefit of this topic is that there is a wide range of possible venues for the students' work, including library/information science conferences and journals (eg, *Scientometrics*, *Journal of Information Science*, *ASIS Conference*) as well as CS/IS subject conferences and journals. As noted above, the sciences in general and computing in particular have not been well-studied by the bibliometrics community; it adds to the excitement of the project to realize that real, ground-breaking, potentially publishable research is being created.

References

- [Ba92] Baldwin, D., and Koomen, J.A.G.M. "Using scientific experiments in early computer science laboratories." SIGCSE Bulletin 24, 1 (1992), pp. 102-106.
- [BE94] Berque, D., Bogda, J., Fisher, B., Harrison, T., and Rahn, N. "The KLYDE workbench for studying experimental algorithm analysis." SIGCSE Bulletin 26, 1 (1994), pp. 83-87
- [BO90] Borgman, C.L. (ed.). *Scholarly communication and bibliometrics*. Sage Publications, 1990.
- [CO92] Cote, V., and Custeau, G. "An integrating pedagogical tool based on writing articles." SIGCSE Bulletin 24,1 (1992), pp. 38-41.
- [CU94] Cunningham, S.J. "Using a computer conferencing system to support writing and research skill development." SIGCSE Bulletin (1994).
- [EN93] Engel, G.L., Dershem, H.L., McFall, R., Lopez, A.M., and Wiltz, S. "Research experience for undergraduates." SIGCSE Bulletin 25,1 (1993), p. 303.
- [HA94] Hafen, M. "Developing writing skills in computer science students". SIGCSE Bulletin 26,1 (1994), pp. 268-270.
- [HO88] Hovis, R.A. "Using current literature in two courses." SIGCSE Bulletin 20,4 (1988), pp. 62-64.
- [JA90] Jackowitz, P.M., Plishka, R.M., and Sidbury, J.R. "Teaching writing and research skills in the computer science curriculum." SIGCSE Bulletin 22,1 (1990), pp. 212-215.
- [LO94] Lopez, A.M., and Messa, K.C. "An undergraduate research program in multi-paradigm software design." SIGCSE Bulletin 26,1 (1994), pp. 271-275.
- [NI78] Nicholas, D., and Ritchie, M. *Literature and Bibliometrics*. London: Clive Bingley, 1978.
- [WI93] Witten, I.H., and Bell, T.C. "Getting research students started; A tale of two courses." SIGCSE Bulletin 25, 1 (1993), pp. 165-169.