Harried Researcher Invents In-House Computer Literature Search System

By Douglas A. Wilcox

A major problem for research scientists in any organization is keeping current with the literature. Even when access problems can be solved by personal journal subscriptions, Current Contents, computer literature searches, reprint requests, and Inter-Library Loan, another major problem still exists – keeping track of personal literature files. When faced with 7-8 linear feet of reprints shelved in my bookcase in 8-10 subject categories and a number of journal manuscripts to prepare, I decided that some kind of action was necessary. The logical solution appeared to be an in-house computer literature search system.

The solution to my problem cast a larger shadow of apprehension than the work in all the manuscripts combined. I decided that help was needed beyond the bio-tech and clerical time available. Fortunately, I had the assistance of Gail Gorka, an experienced VIP who had been working on data collection for me for two years. To acquire more assistance, I wrote a VIP position description and prepared a news release for several local newspapers, advertising for volunteers trained in biological sciences or with computer and clerical skills. Within two weeks I had 13 prospective volunteers to select from. My choices included Terry Albrecht for clerical work and Linda Lewis, Sandy Townsend, and Fran Reigle for reviewing papers.

After some orientation, a discussion of alternative approaches to the problem, and a couple of trial runs, we devised a system that has worked well. Initially, I established a 5-digit, alphanumeric labeling system for the reprints in my collection. The first two digits are letters that designate a subject category for the papers; the last three digits are numbers that allow a capacity of 999 papers in each category. I then formatted a standard FORTRAN coding sheet to provide columns for the labeling code and five 14-letter keywords on each line. Data entry would thus be facilitated once the papers were categorized and keywords assigned. Many of the categories had been long-established in my files, some were beginning to form substantially thick folders and needed to be split further, and some new categories were dictated by new research projects. At present, I am up to 49 subject categories.

To begin the keywording process, I would review 6-8 papers in each category and formulate as many keywords as were necessary. The keywords could range from simple terms (Succession, Zonation), to genus names *(Sphagnum, Scirpus)*, to more complex abbreviated terms (WaterLevFluct, FenHydrology). Usually after reviewing 5-6 papers in each category, I would not need to add very many additional keywords. All keywords were kept on a master list to avoid creation of synonymous keywords and to keep spellings and abbreviations consistent. The remainder of the papers in a selected category were then given to Gail, Linda, Sandy, or Fran. After reviewing the papers and keywords I had worked on to get a feel for the category, the VIP would finish the category, adding new keywords if needed. The labeling codes and citations were listed for each paper and the keywords were recorded on the data sheets following the codes. I then briefly reviewed each paper and changed or added any keywords I felt necessary.

The final steps in the process were handled by Terry, who entered the data into an IBM-PC using the PFS File software. The citations were typed onto 3x5 cards along with label codes and placed alphabetically by subject category into card files.

If I wish to search my literature for papers dealing with a certain subject matter, the appropriate keywords can be selected from the alphabetized master list printed out by the computer. When the keywords are entered back into the computer, the result is a printout of the label codes for any reprints to which those keywords have been assigned. The reprints can then be easily extracted from my files. An additional refinement in the system, which has not been undertaken yet, is to enter all of the citations into the computer.

Adaptations or revisions of the system I use may be more useful to other researchers. The advantages I see in my system relate to ease of data entry and ready access to the appropriate reprints. By sorting and storing the reprints by subject categories, I can readily take an entire file home when working on a research proposal or manuscript dealing with the subject. Assigning of keywords is also greatly facilitated by dealing with groups of similar papers. I like having ready access to the citations in my card file, but a computer listing would greatly facilitate manuscript preparation. I found the VIP mechanism to be a fruitful means of getting a lot done in a fairly short period of time (3-4 months). Much of the credit goes to Gail for continually keeping things organized and doing the lion's share of keywording.

Despite my overall satisfaction with the system, it is not without frustrations. There are certain subject categories that only I can assign keywords for either because the papers are too general or too technical. In both cases, I'm the only one who knows why I'm interested in the paper. Finding time to work on these files can be difficult, as was trying to keep ahead of the VIPs during our initial big effort. The incredible number of keywords now in the system (well over 250) can create confusion also, as can papers that don't fit into any subject category.

By far, the biggest frustration is trying to keep up with the incoming literature. Reprints come filtering in at a fairly regular rate, but if any outside computer literature search is done, I can suddenly be deluged by dozens of papers from Inter-Library Loan. Since my VIP team has been dismantled, biotech Michelle Mueggler and I must tackle the incoming literature ourselves, and I am looking at a forbidding stack of unfinished work even as I write these words. I must say, however, that there is a sense of security in having a good handle on the tools that are part of how I do my job. When I think of how frightening a pile of 1615 unorganized reprints would be, I can only recommend that everyone work out an in-house system similar to mine.

Wilcox is a research aquatic ecologist at Indiana Dunes National Lakeshore.