CORRECTED COPY

ONLINE PERSONAL BIBLIOGRAPHIES, BIBLIOGRAPHY SEARCHING, AND KNOWLEDGE BROADENING

Charles V. Phelps
Jet Propulsion Laboratory, California Institute of Technology

Abstract

This paper discusses how annotated online bibliographies compiled by researchers could be of great value to the research community in the process of knowledge broadening. It then proposes a scheme analogous to citation searching that would greatly enhance the value of this practice.

Introduction

Knowledge broadening is an inexact science. The term is used in this paper to denote the process of acquiring known information about a topic, mainly published information. It generally does not refer to searching for an answer to a specific question, although there is no definite line of demarcation. (This latter process often does stimulate a curious mind into the process of knowledge broadening, and it is hoped that the scheme proposed in this paper will encourage this. Knowledge broadening is also often a prerequisite to finding and understanding the answer to a specific question.) This paper describes how the practice of compiling online publicly-available annotated bibliographies by academics and other researchers could play a useful role in knowledge broadening, and then goes on to propose a scheme analogous to citation searching that would make this practice much more effective.

Knowledge broadening is engaged in by many types of persons, including graduate students exploring potential dissertation topics, university faculty changing direction in their research or preparing to teach a course with whose subject they are not familiar, and R & D personnel embarking on a new project. There are certain typical approaches to knowledge broadening, usually used in combination. One of these is the standard literature search, which includes using indexing and abstracting databases to look up articles, searching for books in library catalogs, looking up articles in reference works, etc. Another approach is to ask someone who knows something about the subject for information. This information could include references to specific papers or books, general tips on where to look ("George Smith at X University has done work in that area"), an oral discourse on the main issues, referrals to colleagues, and many more variations. Once some works have been procured, by whatever means, citations listed in these works can be looked up. Still another approach, which has become quite popular in the last few years, is to "surf the web".

Citation Searching

Citation searching refers to the process of finding works that cite a given (or "seed") work. Citation indexes date back to the nineteenth century, when Shepard's Citations was created as a tool to research precedents in legal decisions. Beginning in the 1950's, Eugene Garfield began work on citation indexes for scientific literature that resulted in the creation of Science Citation Index, and later Social Sciences Citation Index and Arts and Humanities Citation Index.[1] Studies have

concluded that citation searches usually find relevant references that searches of traditional indexes do not, and that the use of both methods together is more effective, in terms of number of relevant documents discovered, than the use of either alone.[2][3][4]

Citation searching, in addition to being used as a way to find documents relevant to an information need, is also frequently employed as a way of judging the impact of a scholar on his or her field, for purposes of promotion, tenure, hiring, and performance evaluation.

Personal Bibliographies

Every researcher collects copies of and references to relevant literature, and employs some technique of accessing that literature. These personal techniques vary a great deal in their orderliness and effectiveness. For some, the technique may be to keep photocopies scattered about at home and at the office, and simply doing a random search when a particular paper is sought. At the other end of the spectrum, there are those who use commercially available bibliographic software to create their own database of relevant literature. Bibliographies compiled by individuals for use in their own research, whether or not commercial software is used, shall be called personal bibliographies. Research groups also sometimes compile lists or databases of works of interest to the members of the group; for simplicity these shall also fall under the label of personal bibliographies.

Many individuals and research groups have begun the practice of listing their own works on publicly available web sites. This is one way of organizing one's (or a group's) own literature, but usually does not extend to the organizing of all works of interest. Usually, this is done as a way to expose one's work to a wider audience, and not with the specific intent of organizing them for the use of the individual or group that created the site.

Online Personal Bibliographies

Imagine that a university faculty member or other researcher were to put an annotated bibliography on the World Wide Web that included all the works that were of interest to this researcher, as well as his or her comments about them. This would not necessarily be just a single alphabetical or chronological list, but might be divided into possibly hierarchical categories of the researcher's own choosing and design. Individual categories might have their own introduction and commentary. Such a presentation would quickly provide an interested individual with that researcher's view of, or "take" on, the subject.

If it were to become standard practice among researchers to publicly organize their work in this way, it would be a boon to knowledge broadening. It would be an extension on a grand scale of the concept of asking colleagues for information about a subject. It would also be a boon to the process of individual organization.

If scholarly works were generally available freely on the World Wide Web (a not-unlikely scenario for the near future, at least in the case of papers published after such a practice became common), then they could be accessed from the online personal bibliographies with just a click,

thus making online personal bibliographies an even more potent tool in knowledge broadening.

What would it take to persuade scholars to create personal online bibliographies? It would probably require an inexpensive or free software tool for creating and managing online bibliographies that would have the following capabilities (among others):

1. Ease of use: This is probably the most important factor. Most potential users would probably not be patient with having to type in all the bibliographic fields manually more often than occasionally. Ideally, if appropriate standards were established and scholarly works were generally available online, bibliography entries could be created with a click or two in most cases.
2. Ability to create "occult" entries: A scholar might want to put an entry into a bibliography for personal use, but, for any number of reasons, not want others to see it. The software would need to be capable of showing all entries to the user, without displaying on the web site entries that are specially marked as "occult". Entire sections of the bibliography could also be marked as occult.
3. Flexibility of presentation: The user would be able to create sections, which could be hierarchical or not, with one's own choice of heading. Entries could appear in more than one section. Sections as well as individual entries could be annotated, and annotations could be of any length. The presentation could be formatted to the user's liking.

A "critical mass" would probably also be necessary. If enough researchers were to make it a practice to do this, the culture of research would be such that most researchers would want to or be expected to participate. This would be an opportunity to put one's own "spin" on the field, and given enough participation, others would not want to be left out. (Remember that one of the most important features of this scheme is the ability to annotate.) The establishment of the features described in the next section would also create interest in this scheme and would be a general inducement to jump on the bandwagon.

Bibliography Searching: An Analog to Citation Searching

Recall that in citation searching, one finds articles that have cited a given article. This same idea can be applied to online bibliographies. If standards were established and adhered to, a system could be created that would allow one to find bibliographies that included a given article (or a work of any type). (To be most effective, standards would have to be adopted and used not only for bibliographies, but for the identity of works in general, such as the Digital Object Identifier [5], NISO's Serial Item and Contribution Identifier (SICI) [6], and Internet Engineering Task Force's Uniform Resource Names (URNs)[7]. However, even without the general use of such latter standards, a system could be created that would be able to identify works unambiguously in the great majority of cases.)

Using such a system, one could, after identifying a pertinent work, quickly identify those individuals who are interested in the same area and the writings that they consider important, and get their general view of the field. This could be done in an iterative process: after finding the bibliographies that include the original work, the user could select works listed in those bibliographies and do an additional search, etc.

An exciting additional capability of the system would be to automatically collect and present all the annotations to a given work. Authors would be able to get quick feedback on their publications from the research community; it would constitute a form of democratic peer review.

It would be a short step to implement another type of search using this system, wherein works would be found and ranked by their relevance in relation to a given work based upon how many times they appeared together in bibliographies with the given work. This is somewhat analogous to the "related records" feature in Institute for Scientific Information's online and CD-ROM citation index products.

It seems safe to conclude that such a system would be interesting to use and would stimulate thinking. Creative thinking is often triggered by serendipity, that is, by chancing upon interesting items that were not specifically sought and are not necessarily directly related to the immediate subject of one's concern. The proposed system would no doubt foster serendipity, as users explore the world views of those with common interests.

Conclusion

This paper has presented a vision of a new form of knowledge broadening that combines do-it-yourself, decentralized knowledge management (personal online bibliographies) with a simple yet powerful way of searching. Many details would have to be worked out and agreed upon to make the scheme as effective as it could be. In principle, however, there is nothing that would prevent an experimental and prototype system from being currently built and tested.

References

[1] Garfield, E. Citation Indexing—Its Theory and Application in Science, Technology, and Humanities. New York: John Wiley & Sons, 1979.

[2] McCain, K.W. "Descriptor and Citation Retrieval in the Medical Behavioral Sciences Literature: Retrieval Overlaps and Novelty Distribution." Journal of the American Society for Information Science, v.40, n.2, pp.110-114, March 1989.

[3] Pao, M.L. "Term and Citation Retrieval: A Field Study." Information Processing and Management, v.29, n.1, pp.95-112, Jan.-Feb. 1993.

[4] Shaw, W.M. Jr. "Subject and Citation Indexing. Part II: The Optimal, Cluster-Based Retrieval Performance of Composite Representations." Journal of the American Society for Information Science, v.42, n.9, pp.676-684, Oct. 1991.

[5] DOI: The Digital Object Identifier System. http://www.doi.org/.

[6] National Information Standards Organization. Serial Item and Contribution Identifier (SICI): An American National Standard. Bethesda, MD: NISO Press, 1997.

[7] [Drafts of documents on Uniform Resource Names.] http://www.ietf.org/ids.by.wg/urn.html.

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.