
Introduction

MARK E. RORVIG

NOT QUITE A GENERATION HAS PASSED since Patrick Wilson (1968) published his slender volume entitled *Two Kinds of Power: An Essay on Bibliographic Control*. In this work Wilson described two simple teleologies guiding bibliography—the power to list and the power to find. Though this *Library Trends* issue is not concerned with the control of the documents used as examples by Wilson, but rather with graphical objects—i.e., pictures, line drawings, and images of pages—nothing much has happened to change the basis of his philosophical exposition. To be sure, the technology of bibliography has changed (and there is much to say about technology in this issue), but the underlying sense of things remains the same.

For example, the article by Petersen, “Developing a New Thesaurus for Art and Architecture,” does not deal much with technology, but addresses the more homely issue of linguistic meaning (or descriptive unraveling) and the management processes used to control the international development of controlled vocabularies. The issue addressed by this volume boils down to only one point: What can be listed cannot always be found, and the relationships among the components of the list enrich or impoverish the finding process.

Indeed, this *Library Trends* issue, taken as a whole, describes the great revolutionary transfer of ideas once confined to documents, to the universe of nonlinguistic knowledge; a great stripping away of the “biblio” portion of bibliography from the graphic component. Moreover, this truly *is* a revolution. Usually when an author uses the word *revolution* it is a misnomer for the real word *evolution*. But intellectual access to graphic records, as opposed to textual records, has returned bibliography to the same complex of concerns

it had twenty-five years ago. We are exactly at the point of inquiry on descriptive processes that was abruptly terminated by advances in computer storage and memory during that period. The IBM 360 of the mid-1960s replaced more than clerical workers in accounting departments; it replaced whole sets of intellectual problems as well.

In the mid-1960s the literary picture of bibliography was one of limitation. Authors addressed the problems of documents as mere objects; things to be retrieved by the assignment of a few meager linguistic phrases to a complicated body of thought. It was much like trying to distinguish individual elephants by regarding a common set of handkerchiefs assigned to a herd of elephants in an orderly way. Quite a few articles were published. A little progress was made. But, in general, everyone seemed quite happy when the character of the problem changed so that the insides of texts, the many words of titles, abstracts, and the text itself, could be turned inside out by application of basic principles of relational database organization (then referred to by the weird phrase "inverted files").

With graphic records though, we are again confronted by these same gross limitations. But this time we will not be rescued by the machine of the gods. A photograph has no way to tell us about itself as a document is able to so inform us (see Howard Besser's "Visual Access to Visual Images"). An engineering drawing cannot be decomposed into words but only into lines and angles (see Bill Beazley's "Impact of CALS on Electronic Publishing Systems"). And as for the beautiful systems that retrieve images of documents as pages (see Frank L. Walker and George Thoma's "Access Techniques for Document Image Databases"), without an index and browsing tools they are as dumb as fishes and pigs.

This issue of *Library Trends* has thus been divided into two sections following the line of Wilson's thought. Section one deals with the power to list while section two addresses the power to find. Within each section the articles are arranged in order from concrete tools and applications to the more arcane theoretical issues and applications. In keeping with the policy of *Library Trends*, the authors have been encouraged to describe their use of specific machines and techniques as fully as possible and to make the reader an insider to the decision paths that were discarded or followed. All of these topics and issues are quite recent. Indeed, most of these articles describe systems and products that are either still in development or that entered into production only in 1989/90. A few articles describe developments that are prototypical only and that may never reach production status. Harold Thiele's "Heraldry and Blazon," article in section one is purely theoretical and has been included only for providing a view of what may be necessary to retrieve graphic records

as directly as we now retrieve documents by words.

Section one directly addresses the power to list graphic records, not only by descriptive rules (though as noted throughout, the formal systems of descriptive cataloging available for monographs must be adapted for application to graphic records) but also by intellectual content. The chief tools for this effort are controlled vocabularies of several types applied by various automated techniques to assist the intellectual task. Broadest in applicability among these vocabularies is the *Art and Architecture Thesaurus* (AAT) described in Toni Petersen's lead article. As in the 1960s work addressing document description, cost elements play the major role in determining the applicability of controlled vocabularies to graphic records. Jeanne Keefe's article, "The Image as Document," describes some of the Yankee ingenuity required to structure system dynamics so that cost effectiveness may be achieved. Gary Seloff's work, "Automated Access to the NASA-JSC Image Archives," details the steps used to create a subset of the *NASA Thesaurus of Technical Descriptors* for application to photographs. Further, Seloff's final production tool—a "visual" thesaurus—incorporates not only relational database technology for the presentation of descriptors, but also the machine linkage of image and text to add greater uniformity to term assignments. In the next two articles, the past is used to recreate the future in two forms. Lois Lunin, in "Descriptive Challenges of Fiber Art," takes us to the world of an art form that dates from the eleventh century Bayeux Tapestry and the ancient art of the Navajo. In this work, the problems of description involved in even a small subset of the art world are described in terms of record users and uses, vocabulary control, and vocabulary development. Finally, Thiele applies the medieval art of heraldry to the problem of reducing trademarks to searchable machine strings, equivalent to string-matching textual systems. The article, "Heraldry and Blazon," may tax readers unfamiliar with algorithmic description, but the application fully formed in his appendix will be a delight.

In section two, working systems that assume an intricate index already exists are presented as examples of the power to find. In this section we are not so much concerned with describing retrieval objects (though descriptive problems are also addressed by these authors) as we are with their presentation as analog and digital displays. This section is intended for those most interested in current technology and its application to the retrieval of graphic records. Significantly, and well in keeping with the cost sensitivity of graphic record intellectual access, many of these systems are personal computer and workstation oriented, making full use of optical disc storage

media. The first of these articles, "ArchivISTA," written by Stone and Sylvain (two systems engineers at the Public Archives of Canada), describes the development path for a cost-effective blend of analog and digital techniques in the presentation of caricature art to the general public. George Thoma's "Access Techniques for Document Image Databases" details the latest product of the National Library of Medicine's longstanding program for optical disc publication of journal articles. His article is the most complete presentation of this large-scale program presently available in the journal literature. Besser's "Visual Access to Visual Images" describes the prototypical work in image collection access carried out over the last five years at the University of California at Berkeley. The Berkeley program is a very ambitious one and will require liberal applications of money for completion to production status. In some respects, this article should be read in parallel with Seloff's in section one since both authors conclude that presentation of text alone in either the descriptive or retrieval processes associated with graphic records will be insufficient for outstanding system performance. Finally, Beazley's report on the development of graphic exchange standards ("Impact of CALS on Electronic Publishing Systems and Users") within government and industry brings the work of the field up to date on the most ambitious technical publication standard to be developed since the inception of the scientific journal.

For any collection of articles such as these, it is the dubious honor of the editor not only to present what was found in the field as the collection of works developed, but also to comment on significant elements that were missing. Most important among these is a total absence of regard for the role that probabilistic indexing (see Maron & Kuhn, 1960, pp. 216-44) might play in systems for graphic record intellectual access. This is a strong departure from the work of the 1960s in which the role of probability assignments to terms applied to documents as retrieval objects was a major source of debate, but one need not be very prescient to see its applicability to graphic records.

A fundamental problem of retrieval is that too much material is retrieved in response to any query. The present thinking is that it is much better to rank the objects to be retrieved in the most likely order of relevance to the query than simply to list them. Conventional methods for creating rankings of retrieval objects usually rely on a large corpus of text terms or a "fine grained" linguistic base. Graphic records, however, are usually not finely grained but coarsely grained in that only a limited number of descriptive elements are present for each record. The cost of indexing is usually too great to do otherwise.

Probabilistic indexing as developed in the 1960s, however, does appear to meet the needs of graphic record ranked-retrieval since it was originally developed for use in document retrieval in a coarse grained, limited description environment. Although, as described earlier, the technique was not widely implemented because advances in hardware (storage and speed) made fine grained retrieval possible from the large source of textual components of the bibliographic record. The technique may be expressed with respect to graphic records by the following formula: $P(I_j, GR_i) \sim P(GR_i, I_j)$ where, $P(I_j, GR_i)$ denotes the conditional probability of relevance that a user will ask for the i th graphic record by the j th index term, $P(GR_i)$ denotes the absolute probability that a system user will want the i th graphic record, and $P(GR_i, I_j)$ denotes the conditional probability that if the user were to be satisfied by graphic record i , the user would request it by index term j .

The outcome of this method is that, with respect to any given graphic record, a user will retrieve or view records ranked by order of the weights assigned to them as modified by their absolute probability of choice. In implementing this technique, it may be assumed that the initial weights will be assigned by an indexer, and absolute probabilities assigned by a behavioral proxy of some type, perhaps that of ordering a print of a particular item in a given search session.

The editor wishes to thank the authors of this issue and the many individuals and institutions that supported the authors in these writings. Revolutions are profound events. I have not seen many in the field before this one.

REFERENCES

- Maron, M. E., & Kuhns, J. L. (1960). On relevance, probabilistic indexing and information retrieval. *Journal of the Association for Computing Machinery*, 7(3), 216-244.
- Wilson, P. (1968). *Two kinds of power: An essay on bibliographic control*. Berkeley, CA: University of California Press.