

The Reference Collection in the Digital Library

Michael Buckland,
School of Information, University of California, Berkeley, USA

Abstract

In a paper-based environment a library's reference collection provides a very useful service for finding answers to factual questions and for finding the context of any topic, person, place or event. Comparable library reference service has not yet matured in the digital library environment. Union indexes, federated search, search term recommender systems, and a faceted approach (What, Where, When and Who) provide useful components.

Introduction

Understanding images, words, or documents depends on knowing their context. For a place one needs to know where it is and what has happened there, its geographical and historical context. An event needs to be seen in the context of its times and related events. Mention of a person raises questions about what that person did, with whom he or she associated, and so on.

In a paper-based environment the reference library provides a very valuable service. The reference works in the reference collection are carefully selected and arranged to enable anyone to find answers to questions for which a concise answer is plausible. (Questions requiring long, discursive answers are likely to need the library's main collections.) One may think of reference libraries as providing a "ready reference" service, the ability to look up or verify isolated factual data such as a bibliographical citation, a date of birth, the location of a place, or the meaning of a word. But a reference library can equally be used to compile a range of different factual data pertaining to a subject, person, or event. In this way, a reference library can be very helpful in enabling one to understand a topic because it allows one to learn about its context quickly and efficiently.

The Internet now provides access to a massive and rapidly-growing range of resources and people search the Internet to find information on all sorts of topics. However, as a resource, the Internet sometimes lacks the effectiveness that arises from the careful selection and high reliability of a library's reference collection. At the same time, although libraries have made substantial progress in moving from a paper to a digital environment, the valuable functionality of a library reference service is still largely absent from the digital library environment.

Recreating reference library service in an Internet environment will require us to make it at least as easy to use and as reliable as in a paper-based reference library. In considering this challenge we have formulated our design objective as the capability of finding the context of any object, any document, any person, or any event: What is related to it in what it is, where it came from, when it originated, and who is associated with it?

In practice it is helpful to use a traditional faceted approach, distinguishing WHAT, WHERE, WHEN, and WHO, because these different facets require different treatment. (See Fig 1).

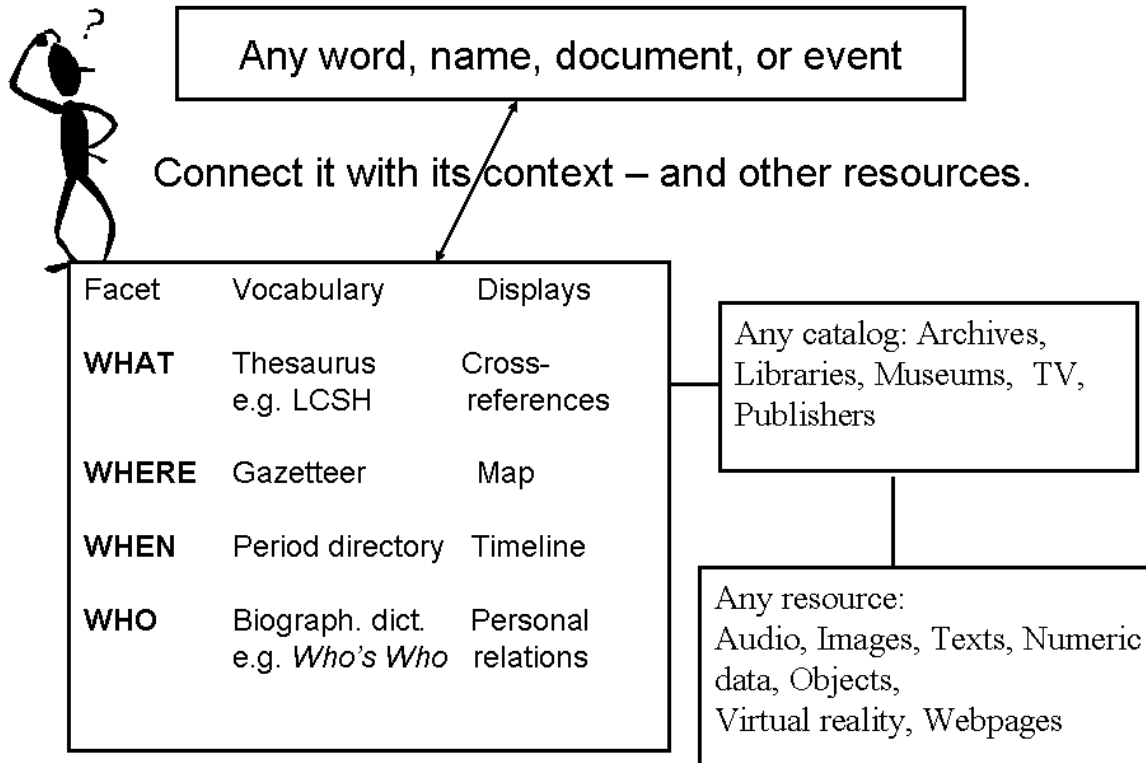


Fig 1. A Structure for Search Support

WHAT

Topic description systems vary greatly with respect to detail, terminology and notation. Even the use of so-called natural language systems are stylized and obsolescent. As one example, if you searched for *Kung fu films* in the Library of Congress Subject Headings, it used to be necessary to look for *Hand-to-hand fighting, oriental, in motion pictures*, but now one should look under *Martial arts films*. When one sees such headings one can understand them, but it is often difficult to guess what heading one should look for and, with alphabetical ordering, topically-related headings may be widely scattered. With foreign languages and artificial notations, as in the Dewey Decimal Classification, the difficulties increase. The standard solution is to build a network of cross-references from the searcher's terms to the indexers' terms. The importance of cross-references between different terms *within* a vocabulary is widely understood. In a network environment, the larger problem of cross-references between different terms in *different* systems' vocabularies becomes important. Creating these mappings by hand is excessively laborious, but algorithmic search term recommender systems can be useful and cost-effective (Buckland Buckland, Chen, Gey & Larson 2006)

WHERE

Place names are commonly multiple, ambiguous, and/or unstable, but a specialized genre of reference work, the place name gazetteer, supports disambiguation and, by providing latitude and longitude, enables places to be located in map displays.

WHEN

Chronologies and time-lines are established genres. Directories of named events are less well developed than place name gazetteers but can be developed along similar lines (Petras, Larson & Buckland 2006).

WHO

Biographical inquiry is aided by two well-established genres: personal name authority files and biographical dictionaries. The problems of disambiguating personal names – distinguishing different persons with the same name and associating different names for the same person – are well-understood and the development of personal name authority files are used to handle both problems. Short biographical descriptions in biographical dictionaries are an important and long-established reference genre. In some cases, biographical records in digital form include links to explanatory resources, as is common in the biographical records in the *Wikipedia*. Two challenges remain, however. One is that standard terminologies for basic familial relationships (e.g. parent, child, sibling, spouse) need to be extended to other common interpersonal relationships such as business partner, class-mate, and mentor. The other is that, useful as internal links within a digital resource are, even better would be links to the best available digital resources anywhere. (See Fig 2.) The Electronic Cultural Atlas Initiative is seeking to address both challenges (*Bringing* 2006).

Eric V of Denmark

From Wikipedia, the free encyclopedia

Eric V "Klipping" (1249– November 22, 1286) was King of Denmark (1259–1286) and son of Christopher I. Until 1264 he ruled under the auspices of his mother, the competent Queen Dowager Margaret Sambiria. In this period he was for some time 1261–1262 a prisoner in Holstein after a military defeat and afterwards he was for some years brought up in Brandenburg. He tried to enforce his power over church and nobility. His conflict with the former was brought to a satisfying result but in 1282 he was forced by the latter to accept an agreement (a "handfastening" – a kind of a Danish Magna Carta) which limited his authority. In the 1270s, Erik Glipping attacked Småland.

Life events metadata

WHAT: Actions

prisoner

WHERE: Places

Holstein

WHEN: Times

1261-1262

WHO: People

Margaret Sambiria

But ideally we need *external* links to the best resources!

Fig 2: Biographical entry from the Wikipedia, annotated.

WHAT, WHERE, WHEN AND WHO

So far we have spoken of What, Where, When and Who separately and noted that separate genres of reference work exist. Each has a distinctive display requirement: for places you need map interfaces, time-lines are used for events, and so on.

In practice these facets need to be considered in combination. Subject headings for What may also contain an auxiliary element for Where and When, as in the Library of Congress Subject Heading (LCSH) *Architecture -- Japan -- Edo period, 1600-1868*. An entry in a place name gazetteer ordinarily includes a geographical "feature type" code indicating what kind of place it is (e.g. castle, lake, town) and, sometimes, when that name was in use. In principle these auxiliary elements for what, where, and when can also be linked to form a rich navigational infrastructure to support searching. For example, the geographic descriptions codes (often called "feature types"), such as "Castle," in a gazetteer represent physical objects not concepts, but mapping them to library subject headings allows smooth transition between literature about, say, castles, and real world examples of actual castles.

The Reference Library Online

The research literature on library reference service especially in the U.S.A. has focused on high-tech support for the call-center role of the reference librarian rather than on the reference

collection or on reference library users. Somehow in moving to the digital library environment the reference collection itself has been neglected, and, even when present, its valued functionality is lacking. But, as working in a digital environment becomes increasingly the norm, library services need to become an integral part of that environment.

An interesting example of a carefully designed reference library online can be found in the Internet Public Library <www.ipl.org>, which has a reference department resembling the reference department of a paper-based library. In it a good selection of reference resources are arranged by topic and genre as in a traditional library. And, as in a paper-based reference library, one selects resource then drills down into it for each search. In effect this is a skillful reconstruction of the affordances and limitations of the technology of the bound book, the codex. However, digital technology does not need to copy the hierarchical structure of the codex. Digital techniques can be used to link directly and horizontally. Three approaches are needed:

1. Unified indexes. Just as a “union catalog” is a catalog of multiple libraries’ collections, a “union index” is an index to multiple books. Think how much of the time, effort and frustration in using a paper-based reference library could be saved if one knew reliably ahead of time exactly which reference works mention whatever is being sought. Union indexes can be found in the print world but they are scarce because of the labor required to create them and to update them. With the cooperation of the publishers, union indexes to digital reference resources should not be difficult to create and to update. A good example is the *World Biographical System Online*, an index to some 10 million entries for over 3 million people in over 8,600 reference works. Publishers’ cooperation would need to extend to allowing union indexes to include all publishers’ reference works.
2. Search interoperability. Procedural interoperability using “federated search” techniques such as the Z39.50 search and retrieve protocol enable a search to originate at any point in the digital environment and to extend it to any other point. Federated search harmonizes the diverse search capabilities and command languages of different resources; and
3. Vocabulary interoperability. Terminology varies between specialties and changes over time. Union indexes do not resolve this difficult problem. Techniques for harmonizing diverse vocabularies, pioneered in Melvil Dewey’s “Relativ” index to the Decimal Classification, are finding a new life today in search term recommender systems.

An inverted relationship

Indexes are created by inverting the relationship between a fragment of a whole and the whole. The science citation index illustrates this process of inversion. The science literature exists as article-filled journals arrayed on shelves. Citations are at the periphery, at the ends and edges of the articles. A citation index is created by inverting this arrangement to create a complex array of citations and when a citation of interest is found, one can follow a link to the text of the article cited.

With an index, especially a union index, instead of selecting a resource and then drilling down into it, one can, in effect, go directly to the mention of the term sought and then consider the currency and probable reliability of any resource in which it is found. Of course, indexes

become much easier to generate from digital resources than from paper and Google's search infrastructure can be considered a huge union index to web pages.

Conclusion

In the paper library, the reference collection plays a very useful role. It is a distilled, concentrated library, with a well-chosen selection of the best and most up-to-date resources organized into a careful arrangement.

In the traditional reference collection one can move easily among encyclopedias, gazetteers, chronologies, biographies, biographical dictionaries, and the rest, like a bee gathering pollen from flowers. In the delegated digital environment much better interfaces to support this movement are still needed.

Vocabulary issues are commonly exacerbated in a digital environment in which the resource is hidden behind a glowing screen and the array of choices is typically less apparent than with a printed index. Appropriate tools are available, notably union indexes, federated search, search term recommender systems, and a faceted approach (What, Where, When and Who), and now need to be deployed.

Acknowledgment: This work was partially supported by two grants from the U.S. federal Institute of Museum and Library Services "Support the learner" (LG-02-04-0041-04) and "Bringing lives to light" (LG-06-06-0037-06). I thank Kimberly Carl, Fredric Gey, Ray R. Larson, and Ryan Shaw for their help.

References

- Bringing Lives to Light* (2007). [Project website] <http://ecai.org/imls2006>
- Buckland, M., A. Chen, F. C. Gey & R. R. Larson. (2006). Search Across Different Media: Numeric Data Sets and Text Files. *Information Technology and Libraries* 25, no 4 (Dec 2006): 181-189.
<http://www.lita.org/ala/lita/litapublications/ital/252006/number4decembera/buckland.pdf>
- Buckland, M., A. Chen, F. C. Gey, R. R. Larson, R. Mostern & V. Petras. (2007). Geographic Search: Catalogs, Gazetteers, and Maps. *College & Research Libraries*. 68, no 5 (Sept 2007). <http://www.ala.org/ala/acrl/acrlpubs/crljournal/collegeresearch.cfm>
- K. G. Saur Verlag. (2007). *World Biographical system online*. [Http://www.saur-wbi.de](http://www.saur-wbi.de) Visited Sept 25, 2007.
- Petras, V., R. R. Larson and M. Buckland. (2006). Time Period Directories: A Metadata Infrastructure for placing Events in Temporal and Geographic Context. In: *Opening Information Horizons; 6th ACM/IEEE-CS Joint Conference on Digital Libraries 2006*. New York: Association for Computing Machinery, 2006. pp.151-160.
<http://portal.acm.org/citation.cfm?id=1141782>