

Preprint of text for poster paper at the American Society for Information Science and Technology Annual Meeting, Vancouver, Nov 9, 2009. Published version in the conference proceedings may differ slightly.

Finding Explanations Online: Easy Queries AND Trustworthy Answers

Michael Buckland

University of California, Berkeley

School of Information, University of California, Berkeley, CA 94720-4600, USA

buckland@ischool.berkeley.edu

Ray Larson

University of California, Berkeley

School of Information, University of California, Berkeley, CA 94720-4600, USA

ray@ischool.berkeley.edu

Ryan Shaw

University of California, Berkeley

School of Information, University of California, Berkeley, CA 94720-4600, USA

ryanshaw@ischool.berkeley.edu

Abstract

Web search engines provide easy access to a huge range of web-accessible resources but the most trustworthy explanations are often in hard-to-search resources in the deep web. The literature on information literacy reflects widespread concern that people tend not to use complex online digital versions of traditional reference works and instead depend heavily on easy-to-use resources of questioned reliability, notably Google and the Wikipedia. The proper response lies in better design to combine ease of search with trustworthy resources. We will report on efforts to demonstrate how that could be done.

Explanations in the Print on Paper Environment

The difference between seeing and understanding lies in knowing the context and relationships of whatever is of interest. To this end genres of explanatory resources evolved over time: dictionaries and encyclopedias; bibliographies and library catalogs; place name gazetteers and maps; time-lines and chronologies; biographical dictionaries; and so on. In a print environment the reference collection of the library provides a carefully constructed environment of auxiliary resources well-designed for providing relatively trustworthy explanations to questions for which a short answer is plausible. In a well-stocked library one can quickly build up an understanding of any topic. In practice, libraries' reference collections are only lightly used and are severely constrained by the limited affordances of the printed codex. Using a print reference collection is often an inefficient, discouraging experience, looking in one volume after another, but finding no mention of what is sought. If only one could murmur a topic and have little green lights appear on the shelves to indicate the volumes that mention it!

Mediated Reference Service and Self-Service

The well-stocked paper reference library is obsolescent now that online services are preferred. In the digital library environment, a reference *librarian* are increasingly available 24/7, which is excellent, but it is not a substitute for empowering users to find explanations for themselves. Tools for user self-service constitute the only scalable strategy for significantly increased service and benefits. The literature on reference and user services has had little to say about new tools to empower users (as distinct from librarians). Two significant concerns are the difficulties of inducing users to use complex existing tools and users' understandable preference for Google and the Wikipedia, which are so easy to use but lack the

structure and selectivity of the resources provided in a library. (For a detailed discussion see Buckland (2008)).

The proper response to these concerns is improved design: tools and techniques that empower *users* and promote the use of *trustworthy* resources. The challenge is make their use competitive with Google and the Wikipedia in ease of use *and* to add the selectivity and trustworthiness.

Technology transfer is commonly a two stage process: First the new technology is used to do the *same thing better*; second, the potential of the new technology is exploited to do *different better things*. The Internet Public Library reference department is a fine example of the first stage; now it is time for the second stage: Doing better things.

Research Report and Prototypes

We will present and explain interfaces developed in the context of three projects: *Support for the Learner* (which clarified the need for precise but extremely easy-to-use search support); *Bringing Lives to Light* (which focuses on the use of mark-up for What, Where, When and Who in biographical narratives); and *Context and Relationships: Ireland and Irish Studies* (which examines linking texts with explanatory resources).

In one initial prototype, personal names and place names in the text being read are identified by the interface (using named entity software) or by the reader and listed by facet, here Where and Who, in a box. Passing the cursor over a name (in the text or in the menu) illuminates it and just two clicks are required to find an explanation: One click opens a menu of recommended reference resources and a second click selects an explanatory resource, induces the interface to formulate and send a query and to display the search result – all automatically. (See Figure). This is very different from merely following a link to a resource's home paper.

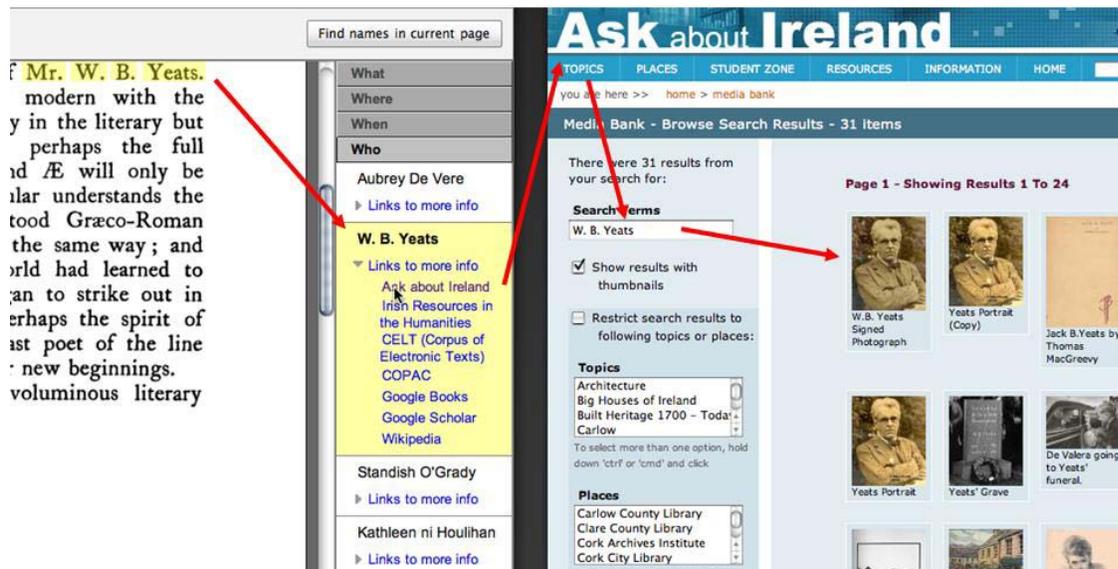


Figure: Screen shots of interface leading from name in text to explanatory resource with two mouse clicks.

The interface described above has one fixed menu of searchable sources for everyone and for each facet. Work is in progress to allow a flexible choice of explanatory sources. Any user should be able to add and arrange their own choices, different for each facet, and to have different menus for different purposes. The user should be able to add and delete sources using a simple online form, but it would be more convenient to click and drag (or “bookmark”) items from any guide to reference works into the

interface menu. Also, like “pathfinder” leaflets, it should be possible to import, export, and share menus crafted for any given topic and level of expertise. Customization is ordinarily seen as tailoring for an individual, but learning is commonly a social event in ways that affect search behavior (Hyldegard 2009).

Finding is not enough. Details of what was found (source, query, what was found) should be easily saved with a complete citation -- without any error prone rekeying! -- into personal notes, into the text being read as XML mark-up ready for the next reader, and/or pasted into an essay or paper being written. (Compare keying URLs letter by letter with a mouse-click “Copy link location”).

Discussion

Ease of search is achieved by delegating to the interface the complex details of identifying the URL of the target search, making the connection, ascertaining the kind(s) of search(es) supported, inserting an acceptable query, sending it, and display of the result. (An automobile analogy: Automatic transmission is just as complex as stick-shift, but the complexity has been moved into the machinery away from the driver.)

The automation of peer-to-peer search and retrieval is achieved through the use of search protocols such as Z39.50 (ISO 23950), OAI-PMH, SRU, and CQL. Convenient “two click” search depends on an interface knowing what search protocol(s) each resource will support. Unfortunately the implementation of search protocols by resources is very uneven and this key detail is still absent from bibliographical descriptions, which ordinarily only provide the URL for (and sometimes a link to) a resource’s homepage.

Web search engines index an astonishing range of web-accessible pages, many outdated, ephemeral, unreliable, and obscure. The emphasis on recall (at the expense of precision or currency) means that the yield is mostly dross not gold. The continuing rapid expansion of the Web and the wave of mass digitization projects can only amplify this effect. The more extensive a Web search engine’s reach becomes, the greater the need for alternative options with a complementary emphasis on the kind of precision, currency, and trust associated with a reference library collection. (Tools are being developed for algorithmic search of the deep web, but even if they were successful this would not address the issue of trustworthiness.)

The dream of a reference library in which little green lights identify the explanatory resources mentioning the topic of current interest becomes feasible in the digital library environment because union indexes (strictly, like Google, more or less edited concordances) can be generated for any set of explanatory resources if the indexing software has access to the full-text of the resources. Guides to reference works, commonly known by their compilers’ names, such as Mudge, Winchell, Malclès, Walford, and now Kieft, expertly describe thousands of resources. Search of these descriptions may be helpful but is no substitute for search of their contents. In a digital environment it become feasible to generate union indexes to the entire contents of customized selections of well-regarded resources. Established Natural Language Processing techniques and XML mark-up can help address problems of term disambiguation and vocabulary control. The creation of such union indexes and the wider adoption of peer-to-peer search and retrieval protocols provide the basis for a transformation of reference service in the digital library environment.

Acknowledgment: Several people assisted the work reported. We are grateful for the support from the Institute of Museum and Library Services through award LG-06-06-0037-06 “Bringing Lives to Light: Biography in Context” and from the Advancing Knowledge grant PK-50027-07 “Context and Relationships: Ireland and Irish Studies,” jointly funded by the National Endowment for the Humanities and the Institute of Museum and Library Services.

References:

Project websites:

Bringing Lives to Light: Biography in Context. [Project website] <http://ecai.org/imls2006/>
Context and Relationships: Ireland and Irish Studies. [Project website] <http://ecai.org/neh2007/>
Support for the learner. [Project website] <http://ecai.org/imls2004/>

Buckland, M. 2006. Access to Digital Heritage Resources using What, Where, When and Who. Presented at the DEN Conference, Rotterdam, December 12-13, 2006.

<http://people.ischool.berkeley.edu/~buckland/bucklandrot.ppt>

Buckland, M. 2007. The Digital Difference in Reference Collections. *Journal of Library Administration* 46: 87-100. [Http://www.ischool.berkeley.edu/~buckland/OKLABuckland.pdf](http://www.ischool.berkeley.edu/~buckland/OKLABuckland.pdf)

Buckland, M. 2008. Reference library service in a digital environment. *Library and Information Science Research* 30, no 2 (2008): 81-85. Preprint at

<http://people.ischool.berkeley.edu/~buckland/libref.pdf>

Hyldegard 2009. Beyond the search process – Exploring group members' information behavior in context. *Information Processing and Management* 45: 142-158.

Larson, R. R. and others. 2008. *Bringing Lives to Light: Biography in Context*.

[Http://metaphor.ischool.berkeley.edu/Friday_2_15_08.ppt](http://metaphor.ischool.berkeley.edu/Friday_2_15_08.ppt)