

June 2004

## XML in Action

Norman Desmarais

*Providence College*, normd@providence.edu

Follow this and additional works at: [http://digitalcommons.providence.edu/facstaff\\_pubs](http://digitalcommons.providence.edu/facstaff_pubs)

---

Desmarais, Norman, "XML in Action" (2004). *Library Faculty and Staff papers*. 27.  
[http://digitalcommons.providence.edu/facstaff\\_pubs/27](http://digitalcommons.providence.edu/facstaff_pubs/27)

This Article is brought to you for free and open access by the Phillips Memorial Library at DigitalCommons@Providence. It has been accepted for inclusion in Library Faculty and Staff papers by an authorized administrator of DigitalCommons@Providence. For more information, please contact mcaprio1@providence.edu, hposey@providence.edu.

## XML in Action

By Norman Desmarais

*This article first appeared in 'Against the Grain', June 2004*

The software developers at xrefer envisioned the potential of XML (eXtensible Markup Language) as early as 1999. When they began to create xreferplus, they tagged the content with an abundance of metadata in XML compatible format. This metadata is fully searchable and interactive with the various components of the product. Because xreferplus uses XML compatible metadata, it is also compliant with SFX and similar link resolvers (see For SFX See Librarian. *Against the Grain* 15:3 (June, 2003) pp. 102-103).

xreferplus contains the full text of 164 books from 36 publishers. Subject areas cover art, biography, business, geography, history, language, literature, law, medicine, music, philosophy & psychology, religion, science, social sciences, and technology. Technology and art, health, and science are the areas that saw the most growth in 2003. The library includes general encyclopedias like the *Crystal Reference Encyclopedia*, *The Columbia Encyclopedia*, *The Hutchinson Encyclopedia*, and *The Macmillan Encyclopedia*. The *Philip's Encyclopedia 2004* was added in the last week of January. The library also includes ten language dictionaries. There are about 76,000 audio pronunciations in MP3 format for the *American Heritage Dictionary*, *the Academic Press Dictionary of Science and Technology*, and *Dorland's Medical Dictionary*. Any images from the published sources are also included.

There is also a conversion engine that will quickly calculate English or metric equivalents for area, distance, energy, speed, temperature, fuel consumption, power, volume, and weight. This engine is fully integrated within xreferplus.

Many of the titles are from British publishers, especially Peter Collin Publishing, Blackwell, Bloomsbury, Routledge, and Thames and Hudson, showing the product's origins; but American titles have been added from publishers like Houghton Mifflin, HarperCollins, and Wiley. Content from American publishers will be the fastest growth area for 2004. Harvard University Press has recently signed up for a couple of music dictionaries. The individual titles are all richly indexed with XML compatible metadata that allows them to be fully cross-referenced. Unlike traditional cross-references which are tied to the framework of a single book, xreferplus's cross-references span the whole digital library to connect relevant information.

One can browse each book in the digital library, as in the print equivalent, to explore its content; but xreferplus's real power lies in the search engine. This engine, which resembles that for an Advanced Google search, permits Boolean searching, thematic searching, stemming, and concept searching. It supports the usual Boolean operators (AND, OR, NOT) but not proximity operators like NEAR or WITHIN X WORDS. One can search by subject, exact phrase, quotations, or words that are spelled or sound like the keyed entry. One can search the full text or only the headings, all the words or any of them, or exclude certain words. One can also search by exact phrase or limit the search to a particular topic or group of topics.

This is impressive enough for a search engine; but it just hints at xreferplus's power. When the results of a search display on the screen, a column of "xreferences" appears on the left. These are related cross references generated by the metadata. They go beyond traditional cross-references, which connect related terms within a single source, by cutting across topics, titles, and publishers as well as text, audio, and image files.

Clicking on an xreference triggers the retrieval of additional relevant material. In other words, it operates like a hyperlink that generates additional xreferences. Whereas traditional cross references are one dimensional, confined to the framework of a single book, xreferences add another dimension to reference works by providing links to navigate a collection of books or sources.

This feature is called the Research Mapper and uses information visualization tools that characterize hyperbolic browsers (see *Innovations Affecting Us – Hyperbolic Browsers: From GUI to KUI. Against the Grain* 15:5 (November, 2003) pp. 95-97). This can be fascinating and intriguing to watch because the researcher can observe the retrieval engine at work as it builds the nodes and changes the distribution layout in real time, depending on what it locates. If the search engine detects a relationship between two terms, it uses xrefer's algorithm to test that relationship. It then groups the results by subject in nodes.

The presentation looks like a complex molecule that grows like an amoeba (see Fig. 1). The Mapper shows not only the entries that match the search term but also how they relate to each other and to other entries as well. The closer the results appear to another, the more they have in common. This guides the researcher to the most relevant information related to his or her research and leads to areas that warrant further exploration.

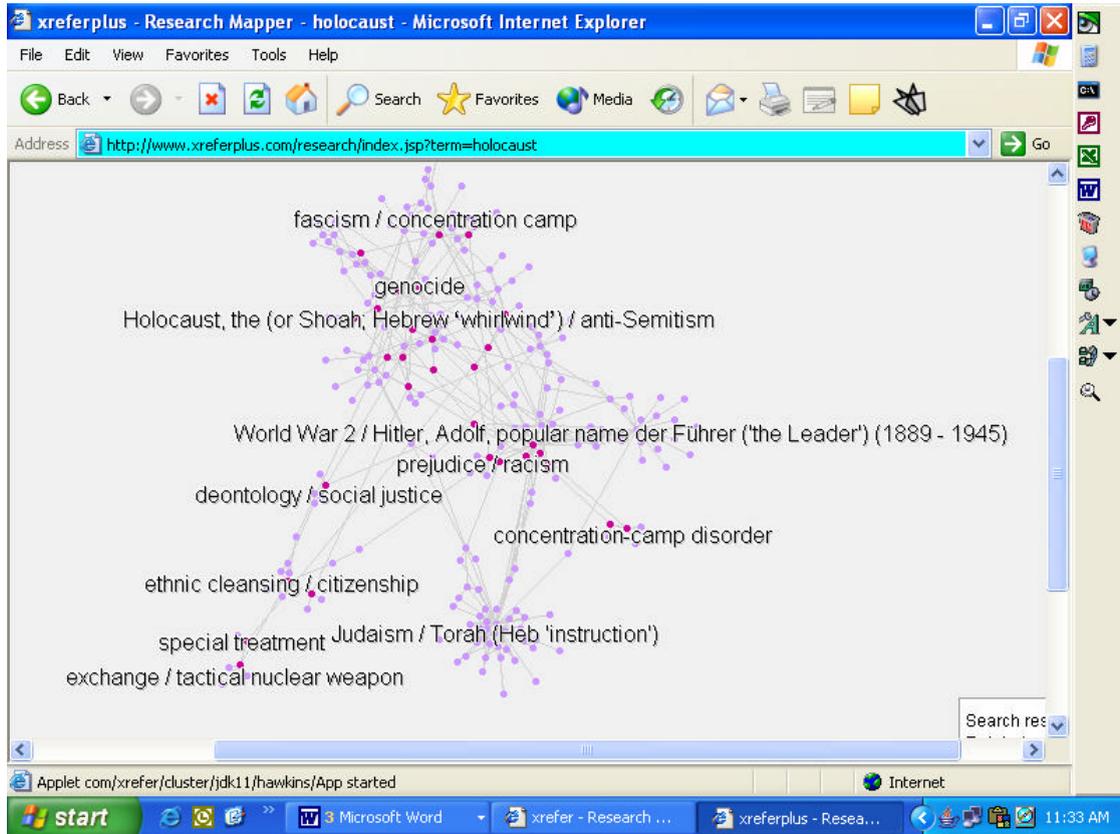


Fig. 1 Xreferplus's hyperbolic browser display for the term HOLOCAUST.

Moving the cursor over a node displays the title or subject content of that node. Pressing the shift key and clicking on a term expands the concept of that term and builds a bigger cluster, generating new and unforeseen results/relationships. This is a great tool to find desired information quickly, to expand one's knowledge of a given area, or to locate information when one doesn't know exactly what to look for. This is particularly good for someone who doesn't know what they don't know and wants to learn. By representing the search results in a visual form, xrefer lets researchers put more context into their research and see more results on a single page.

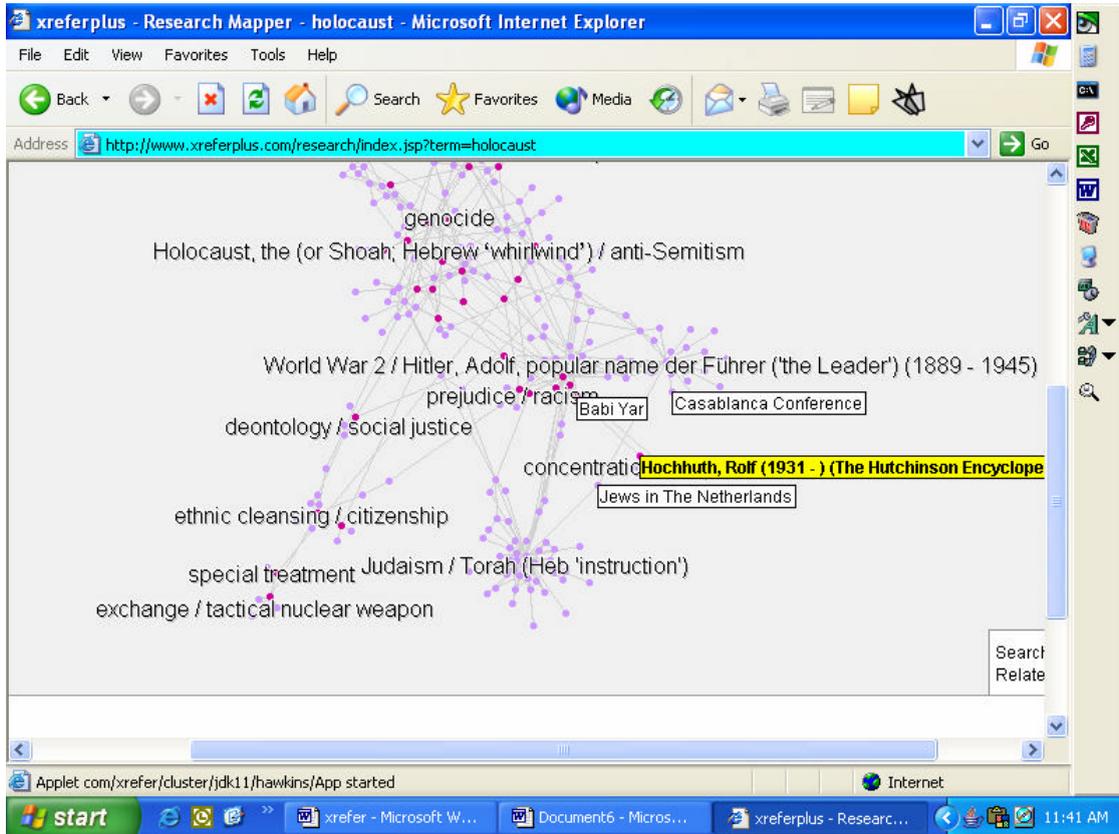


Fig. 2 Moving the cursor over a node highlights the content represented by that node in a yellow box and may display the content of other nodes as well. Clicking the entry in the yellow box will display its entry in a new window.

The Research Mapper provides a visual map of how search terms and topics in xreferplus are interconnected. One can access it either from the set of search results or from the home page. One can also analyze the structure of a map by holding down the Ctrl key while moving the mouse around the map. The relationships change as one moves the cursor from one node to another. The effect is illustrated below.

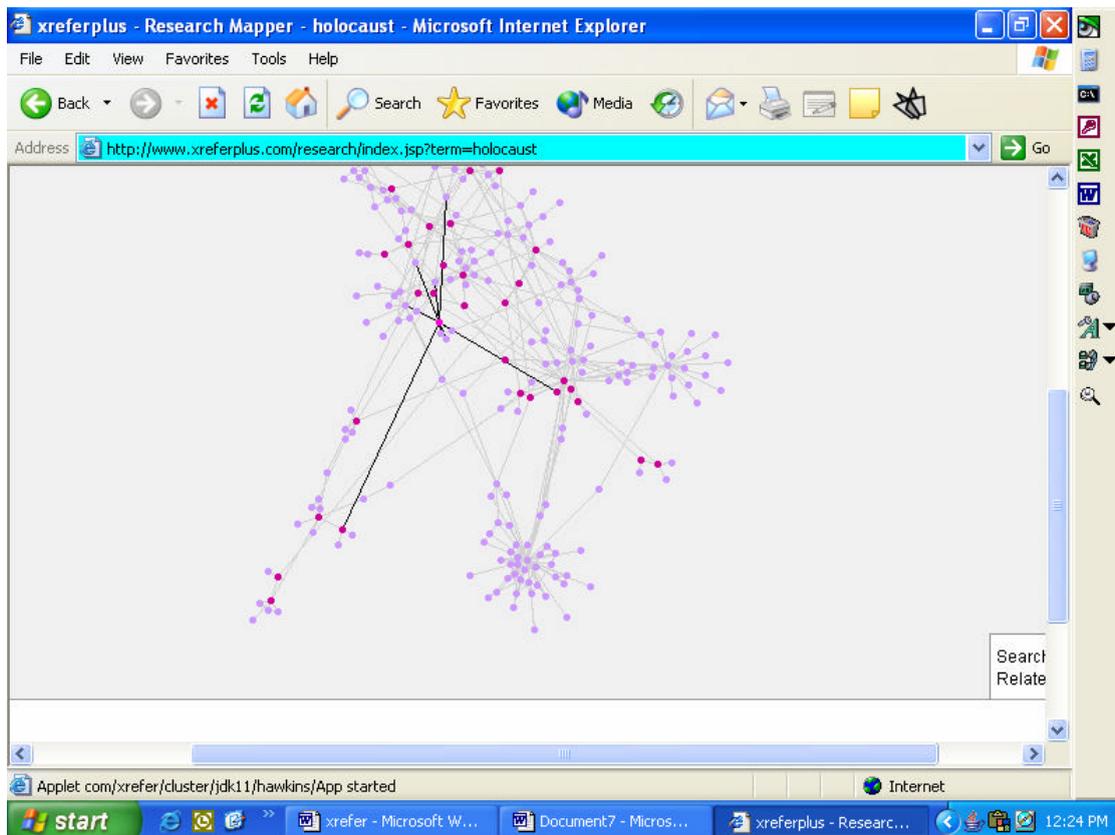


Fig. 3. Researchers can analyze the structure of a map by holding down the Ctrl key while moving the mouse around the map.

The bottom right corner of the Mapper window displays the number of results being mapped. Researchers can opt to have a maximum of 100, 200 (default), 300, or 400 results mapped. The higher the number, the longer it will take to complete the visualization. Research Mapper requires Java.

While xreferplus is SFX compliant because of the XML metadata, subscribers will probably opt not to set it as a target, as most of the entries in the reference works contained in the library have brief entries. Rather, this product would best serve as a starting point for research. Subscribing institutions might choose, instead, to have their link resolver use xreferplus as a starting point to link to other library resources like the public access catalog or specialized databases or more detailed reference sources. Institutions might link to individual titles in xreferplus from a library web page or a departmental portal.

Another nice feature of xreferplus is that it lets the researcher format the results in APA, MLA, or Chicago Manual of Style format when sending a reference to a friend or to oneself. There does not seem to be any option

for downloading to disk other than email, copy and paste, or printing. xreferplus also automatically adds a bibliographic citation to identify the source of the reference. MARC records are also available for download free of charge. Catalogers can select all the records, created by DGI (Donohue Group Inc.), or a subset which they can select by title.

One can subscribe to the complete xreferplus library or choose a "pick and mix" option to subscribe to 100 or 125 titles of their choice. The price of xreferplus is negotiated by site and a 30-day free trial is available. Xrefer does not license xreferplus by simultaneous users; so the product is available to all users of a subscribing institution, including remote users.

Database administrators will appreciate the usage reports that will allow them to monitor title use and change their subscription mix if they opt for the "pick and mix" option. Statistics are collocated monthly and provide the following information: number of sessions; the number of searches performed; the number of entries displayed; the top twenty entries viewed; the top twenty search terms performed; totals for pages including advanced search, help, home, tour, and xreferences; and the number of entries viewed in each of the titles in the subscription. The product can also be integrated with the "Ask-a-librarian" link which appears beneath the institutional name at the top-right of the xreferplus pages. xrefer's American office is located at 31 St. James Ave. Suite 370, Boston, MA 02116 tel. 866-426-2654. Public views of some pre-determined interactive maps can be viewed at <http://www.xreferplus.com/>.