

Online Currents - Vol 19 (8) October 2004

THE WEB LIBRARY: BUILDING A WORLD CLASS PERSONAL LIBRARY WITH FREE WEB RESOURCES

BY NICHOLAS G TOMAIUOLO

reviewed by Belinda Weaver

Anyone who has surfed the Web knows that there is a whole universe of free information out there for the taking. The Web has attracted enormous numbers of people with information to share – from the keen volunteers painstakingly entering books into Project Gutenberg, to organisations, associations, government agencies, universities and other authoritative services who have discovered “that the Internet is simply a cheaper and more effective way of disseminating their information than print on paper.” Author Nicholas Tomaiuolo has that right. He goes on to say: “Many are commercial publishers and Web content producers who are willing to risk giving some of their content away for free on the Web, in the hope that they will be able to entice you into a subscription or at least make you stick round long enough to see some of the ads on their sites.” Despite the free stuff, a lot of people still want to charge.

It may be hard for librarians to convince people that not everything is freely available on the Web when books like this document so much free information. But Tomaiuolo does try to dampen some expectations:

“Among resources that you will still need to pay for, or at the very least, make a trip to your local library to use, are sensitive company information, the newest bestsellers, and personal articles that never make it to the Web.”

He’s right. Information that is costly to produce, such as market research or business data, will always come with a price tag. What is amazing about Tomaiuolo’s compilation though is how much material – even scholarly research – is actually out there – gratis. The job of tracking down and organising this vast store of freebies cannot have been easy, and I for one am grateful that someone else has created the road map for me. The book comes with a companion Web site at <http://www.ccsu.edu/library/tomaiulon/theweblibrary.htm>. This site is updated frequently to compensate for the fact that addresses listed in the book might vanish or change. My only quibble with the site is its appearance – messy layouts, hard-to-read fonts – but that aside, the site provides a much easier way to navigate what’s inside the book than simply using the print.

In addition to reviewing free online sites, Tomaiuolo intersperses the text with interviews with Web experts. These interviews place the material being discussed in context and give the point of view of either the producer of the material or that of the dedicated user.

The first chapter tackles the problem of finding background or research material online. Students, researchers and academics alike are interested in finding magazine or journal articles and Tomaiuolo has developed a nifty strategy for making the most of what’s freely on offer. He separates the job into two parts – the first consists of identifying the articles you might want to read; the second tackles the problem of getting the articles for nothing. Many of the sites that offer free articles, such as FindArticles.com (which includes content sources from Gale Group publications) do not offer very sophisticated search facilities. Tomaiuolo recommends using the free search facilities of sites such as Ingenta or Scirus to identify the works wanted, and then going on to use sites like FindArticles to grab the content. FindArticles provides free access to scholarly journals such as The Lancet, so it is not a destination for popular magazines only. Tomaiuolo documents other free article sites and gives a good description of each.

He also covers the burgeoning growth of preprint and other open research archives where

scholarly research papers or theses can be found. The book was probably issued before the launch of the brilliant cross-archive search tool, OAIster, which to date indexes the content of more than 300 open archives. These services include e-theses, working paper archives, open access journals such as BioMed Central, preprint servers, museum collections, digital image libraries, technical reports and other scholarly materials, and are a boon to anyone who wants to get research on the cheap. However, Tomaiuolo does mention other finding tools such as ARC and provides links to some preprint sites. With OAIster having struck a content sharing deal with Yahoo!, much of this material can now be found in search engine databases as well as through these kinds of specialised tools.

Tomaiuolo deals with online news in chapter 2. Part of his argument in writing the book is to convince librarians to stop paying for material and opt for free models online instead. This argument must be weakest with news, as news sites are still struggling with the issue of how to make money online. If they charge for content, Web surfers may go elsewhere. If they give their news away for free, they risk eroding their print subscription base or losing their TV audience. With issues of compulsory registration, site subscriptions, and archiving still up in the air for many news organisations, cancelling paid subscriptions in favour of free content would seem foolhardy. Perhaps more free archival content is available to US readers, at whom Tomaiuolo’s



book is targeted, but the Australian situation is quite different – to get much Australian news after the current day, libraries have to sign up to services such as Factiva, LexisNexis or NewsText. Tomaiuolo also fails to mention one good source of free archival news – the UK Guardian's archive which stretches back several years. The UK Daily Telegraph also has an extensive free archive that is good for international news. To his credit, though he does have a very American-centric focus, Tomaiuolo does cover the use of news alerts and the growth of Web logs (blogs) as providers of news, so the chapter is still worth reading.

Also of benefit are the two chapters on online look up tools. The Reference on the Web chapter points to the vast range of material already out there for look-up – almanacs, encyclopaedias, dictionaries, health and other specialist subject material, factbooks, directories and more. Again, the focus is American, which limits its usefulness, though many of the tools listed, such as literary, geographical or scientific resources, will have universal relevance. The chapter is mainly pages of lists, arranged by category, of what can be got. The next chapter covers Expert, AskA .. services and digital reference. Several of these expert sites are reviewed, which helps with choosing the right one. Tomaiuolo then goes on to suggest other paths for when you draw a blank, such as USENET newsgroups or posting questions to lists such as STUMPERS-L. He also weighs up the usefulness of Ask a Librarian online reference or chat services.

Free books are covered next, with Project Gutenberg predictably getting a good write-up. In the chapter, Tomaiuolo includes an interview with Michael Hart, the founder of the project, who provides a cost comparison of using Project Gutenberg books as against buying them in print. Other e-text sites are reviewed, and there are further interviews with library professionals about the usefulness of e-texts. No mention is made of Amazon's new facility of peeking inside new books before purchasing them, but there is a comparison availability chart for selected titles to demonstrate the superiority of Web archives over ordinary bookshops, especially since one of the selected titles is out of print. No mention is made of antiquarian or secondhand booksellers, who could reasonably be expected to fill such gaps, presumably because they charge for their services.

Images are the subject of the next chapter and Tomaiuolo deals with a range of issues here – from

finding images to acquiring them, issues of copyright and the plethora of image formats. Tomaiuolo covers some public domain image services as well as providing a few links to paid services such as Getty Images, Art Today and Timepix. I would disagree with Tomaiuolo's assertion that Web surfers should not have problems finding the images they want. I find using search engines such as Google or AltaVista to track down images frustrating; the lack of metadata means you have to page through screen after screen to find what you want. Searching in digital image archives is

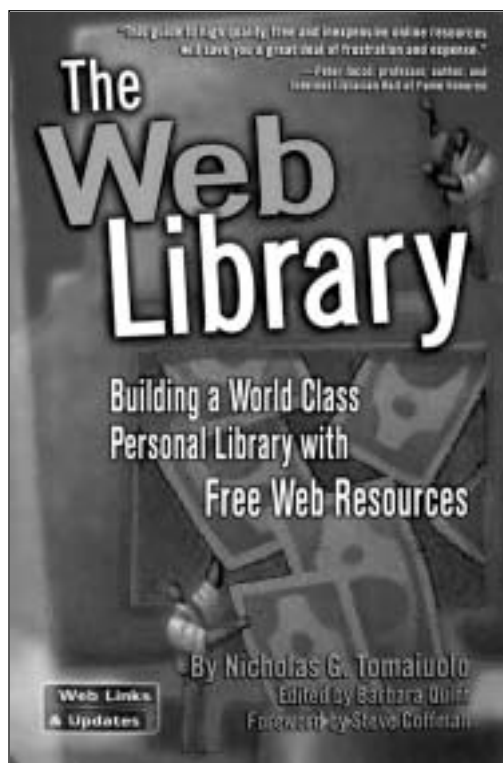
certainly a more satisfying method of targeting queries, since many images have been extensively described. In my opinion, good image searches will not eventuate until more images online are adequately described by metadata.

From images, it is logical to move on to exhibitions and the Web is certainly rich in these with many galleries, museums and libraries mounting online exhibitions, some permanently, some more ephemerally. The chapter reviews a number of gallery sites such as the Museum of Modern Art and then goes on to discuss the use of such sites for research, for example, the use of digitised versions of the Gutenberg Bible by scholars. The Library of Congress's American Memory collection gets big coverage but so too do sites designed more purely for

entertainment. Tomaiuolo also talks about the use of gallery and exhibition sites for trips and travel planning.

The next chapter is a must-read for any Web user as it covers the kinds of software necessary for Web use – the basic plug-in software, such as Adobe Acrobat, without which certain classes of material cannot be used or viewed. Other basics include RealPlayer for audio and video, QuickTime for movies, and Flash and Shockwave for animations. Tomaiuolo also covers downloadable toolbars such as the Google toolbar. Again, the time lag may be operating against the book here – while Google's toolbar is handy, it is nowhere near as versatile as the new deskbar which sits on the Windows desktop rather than bolting on to a Web browser. I use the deskbar several times a day, as I can launch searches and dictionary lookup from within other applications, such as Word, without having the browser window already open or active.

Tomaiuolo has also included a reality check by covering in his final chapter the kinds of materials that may never be in the Web Library. He has also tried to compensate for the lag between publication and new developments by adding material to the Web site that is not covered in



the book, and by adding new material to existing chapters of the book online. The book's appendix lists all the sites mentioned throughout, chapter by chapter, which is a much easier look-up aid than going back over the book itself. For those with Web access, if the idea of typing in all those URLs is a chore, use the companion Web site and let your mouse take the strain. Advice from librarians to evaluate what's trustworthy and what's not is also a feature.

While the book will be an undoubted godsend to those keen to organise and manage the vast shoals of free online information, its American focus limits its usefulness somewhat (though this is not decisive). The book's evangelical argument that librarians should start using these free services and dispense with paid ones failed to convince me. It might be possible to drop Medline and switch users over to PubMed, as the latter seems guaranteed as a public service. But PubScience vanished overnight, despite being a seemingly bullet-proof US government initiative. The same could happen to other currently free services. They could easily close down, or start charging. As developments on the Web show, sites change their *raison d'être* constantly. To cancel a newspaper subscription only to see a free online service start to charge for content

would make many a librarian regretful about the damage done to a hitherto unbroken collection.

In his foreword, Steve Coffman claims that it "simply ludicrous for libraries to continue to purchase things that everybody else can get for free. It is high time we re-evaluated our traditional collections and roles in light of the changing reality of the Web." Alas, the reality is that that many, many Web sites change or disappear. You can certainly use this book to create or add value to an existing virtual reference collection, or to add many marvellous services to your existing catalogue. But to throw out subscription materials in favour of sites that may be gone tomorrow would only be for the foolhardy. This book is a tremendous summing up of the materials many of us will want to link to online. However, it can never be the whole answer. The Internet is not yet ready to replace the traditional library collection.

The Web Library: Building a World Class Personal Library with Free Web Resources by Nicholas G Tomaiuolo, edited by Barbara Quint. Medford, NJ, Information Today, 2004. US \$ 29.95.

Belinda Weaver is Co-ordinator, ePrints@UQ - the UQ ePrint Archive, University of Queensland Library.

NET NOTE

FAST SEARCH & TRANSFER UPDATE

Fast Search & Transfer (<http://www.fastsearch.com>) has recently announced that InfoData has signed an agreement to deploy the FAST Enterprise Search Platform (FAST ESP) across multiple applications throughout its enterprise. InfoData, which was acquired by the BTJ Group in March 2004, is a privately held company and a leading supplier of online databases and direct marketing services in Sweden. It has developed one hundred Swedish databases and one thousand international databases under the name of InfoTorg. InfoData (<http://www.infodata.se>) offers a variety of services such as Electronic Commerce, EDI

handling, contact centre development, and direct marketing services.

Fast Search & Transfer has also announced the release of FAST AdVisor, an enterprise-wide software solution that facilitates the introduction of buyers to suppliers for Internet Yellow Pages (IYP), Business to Consumer (B2C) and Business to Business (B2B) verticals, online classifieds, and traditional publishers. By improving the overall search experience, consumers will be able, more quickly, to discover relevant information, and advertisers, in turn, will connect to these customers more rapidly.

NET NOTE

NETCONTENT LAUNCHES SCOOP

NetContent has launched Scoop (<http://www.scoop.com>) to help sales, marketing and communication executives monitor business news and redistribute important information to their selected audiences. NetContent provides alert services, company research, electronic news feeds, and newsletters to businesses and professional organisations. It secures copyright in order to deliver information from over 15,000 sources including magazines, professional journals, newspapers and online publications.

User benefits include improved selling efficiency, competitive intelligence, and credibility among co-workers, clients and prospects. In addition to news monitoring and company research reports, Scoop provides a publishing service that lets executives deliver key information to their co-workers, clients

or prospects via Web site postings, e-mail or RSS feeds.

It is available by personal subscription or enterprise license; it is a Web-based Software as Service product. It can be personalised by the individual end user and is configurable for specific industries or corporate information needs.

NetContent has recently been selected as an initial provider of premium business news article content for the Find.com business search engine. Find.com (<http://www.find.com>) is a business-focused search engine. It combines free search of business-oriented Web sites with pay-per-view access to premium articles and research.

Available content sources provided by NetContent include articles and company research from Proquest, Thomson-Gale and NewsNow.