

Casting a wide net

The Early English Books Project meets at Northwestern

by Jeffrey Garrett

The history of the computer is widely known to have begun with Pascal's adding machine of 1642 and Thomas Hobbes's reflections on "computation" in 1656. Less known is that the poets of the 16th and 17th centuries imagined what some day would become the Internet and the Web.

In a poem of 1611, John Donne anticipated the way an abstract network could be "thrown upon the heavens" and, like the nets used by fishermen, bring the universe to us:

For of Meridians, and Parallels,
Man hath weav'd out a net, and this
net throwne

Upon the Heavens, and now they
are his owne.

Loth to goe up the hill, or labour
thus

To goe to heaven, we make
heaven come to us.¹

Hyperlinks, image maps, the use of icons as memory devices representing larger meanings, even interactivity between medium and reader (16th-century almanacs performed astronomical calculations for the user and contained blank pages on which readers could add their own observation data), all of these "tools" were anticipated—and many fully developed—in early modern Europe.²

The 21st century is in the process of repaying the debt it owes to the philosophers, mathematicians, and dreamers of the 16th and 17th centuries.

Early English Books Online

When finished, Early English Books Online (EEBO) will bring almost every work published between 1475 and 1700 in England, 125,000 in all, to the computer monitor in your home or office, not only as browsable reproductions of printed pages, to be read like a book on the screen, but also as electronic text, in which every occurrence of a requested word or phrase can be located and collated with every other occurrence of that same word or phrase elsewhere in a vast electronic library of Renaissance and Restoration England.

EEBO traces its origins to an enormous microfilming project begun in 1938, which reproduced on film the entire corpus of early printed books in the British Museum Library. In the 1990s, the creators and owners of this microfilm, University Microfilms (UMI) of Ann Arbor, Michigan, began to transfer this vast archive from film to bits, opening up the prospect of manipulation by computer and access through digital networks. Northwestern University Library and many other U.S. libraries acquired EEBO in 1999.³

About the author

Jeffrey Garrett is acting assistant university librarian for collection management at Northwestern University, e-mail: jgarrett@northwestern.edu

EEBO's Text Creation Partnership

Reading centuries-old texts online is all well and good, especially if the alternative is a trip to a microfilm reader at a university library on a cold winter's night, but what if we want to know how often Shakespeare's name (or Cromwell's or King Charles's) was invoked by writers of the 17th century, and in what context? For this, we need electronically searchable text—and a page image, regardless how accurately it reproduces the original, does not relieve a human reader of the need to scan every page to locate these occurrences. For millions of pages this could take a lifetime and, in the history of scholarship, it often has.

This is where EEBO currently stops, but where another recently established initiative, the EEBO Text Creation Partnership (EEBO-TCP), has received a sweeping mandate.

As of January 2002, 53 universities in the United States, Great Britain, and elsewhere in the world (Northwestern among them) were contributing members of the TCP, working under the leadership of the University of Michigan, Oxford University, the Council on Library and Information Resources (CLIR), and ProQuest Information and Learning (formerly University Microfilms).

The initial goal of the TCP is to create searchable text versions of 25,000 EEBO titles over a five-year period, concentrating on the first editions listed in a standard compilation, the *New Cambridge Bibliography of English Literature* (NCBEL).

This process involves much more than just creating a digital copy of the text of these works. It also involves what is called "tagging," which means labeling parts of a text as "author," "title," part of poem, a foreign word, or occurring on a particular page, in a particular chapter, or in a publication of a particular year. The power of tagging is that it allows readers to limit a search for words to

particular parts of a text or to works published during a given range of years. Therefore, tagging allows for searches of enormous precision or "granularity."

EEBO-TCP Summer Camp

Last July, representatives of six TCP member institutions located in the Midwest met at Northwestern University Library for a two-day summer camp of lab time and intensive consultations to help create the

all-important interface that will stand between users of the texts and (literally) millions of pages of electronic text.

Imagine this interface as a kind of dashboard on your computer, with adjustable controls a driver will need to navigate through this new cyberspace of Renaissance e-text. It is, of course, very important that the drivers get the controls that they will need to get safely and quickly to their destination.

Among the participants of EEBO-TCP Summer Camp were students and faculty members of the English and history departments at Indiana University, Notre Dame, University of Michigan, Michigan

State, Wisconsin, and Northwestern.

Librarians, especially subject specialists in English and history, were paired with their respective faculty, and computing specialists—some also with Ph.D.s in humanities disciplines—noted carefully what the future users of the full-text database said they would need.

Problems and wishes

It is clear that early modern books pose particular problems for digital projects such as this one. How, for example, do you search for occurrences of a particular word if that word can be spelled many different ways? The word "green," for example, could be spelled "grene"—but "grene" was also an alternative spelling for our modern word "grain."



The reading wheel, from Agostino Ramelli's *Le diverse et artificiose machine* (Paris, 1588). Reproduced by permission of the Harry Ransom Humanities Research Center, University of Texas at Austin.⁴

Pagination is often lacking. In that case, what navigational aids can be given to the electronic reader? Scholars warned that modern genre categories, even as general as "fiction" and "nonfiction," cannot be safely applied retrospectively to early modern books. Histories of England, to give just one example, may be presented as richly allegorical epic poems. If not ours, then what genre categories should be used?

It also became clear that social scientists have different needs and expectations of the TCP than humanists do. A trivial example would be the different meaning that the word "act" has for humanists, for whom it is a part of a play, and for legal historians, for whom an "act" is, of course, a law promulgated by Parliament.

Historians present pointed to the importance of capturing electronically the text of the 22,000 so-called "Thomason Tracts," collected between 1640 and 1661 by London publisher and bookseller George Thomason—an extraordinarily valuable record of the turmoil in England during the Civil War. Humanists emphasized, on the other hand, the need to tag typographical information in EEBO texts. This includes ornamental initials, signature information, catchwords, and other information, much of which was used by printers to properly assemble a book after the pages had been printed and cut. These details are also important for distinguishing different editions or even different printings of the same book.

Everyone present was concerned about how a sub-corpus of five, ten, or hundreds of texts might be identified based on content, year, author, or other search criteria, to allow further searches to be restricted to a specific group of works.

Staff from the main TCP office in Ann Arbor returned home with thick quires of notes and suggestions. Though sorting through it all will take a long time and careful analysis, the team charged with creating the interface got what it hoped for. They will now set to work creating a powerful "Renaissance computer," such as the one the likes of John Donne and Thomas Hobbes dreamed of 400 years ago.

Notes

1. An image of the page in Donne's *An Anatomy of the World* (London, 1611) containing this poem is available online from EEBO at <http://www.lib.umi.com/eebo/image/5655/10>.

2. A host of these innovations are detailed in a new book, *The Renaissance Computer: Knowledge Technology in the First Age of Print* (London: Routledge, 2000), edited by Neil Rhodes and Jonathan Sawday.

3. Example of EEBO text are available at the EEBO "Featured Content" page at <http://www.lib.umi.com/eebo/featured>. There you can download, for example, images of every page of the 1600 edition of Shakespeare's *Midsummer Nights Dream*.

4. "Hypertext" and "intertextuality" are new words, but the concepts behind them are hundreds of years old. In this engraving from the late 16th century, a scholar follows "links" that lead him from one text to the next, using an elaborate wheel instead of today's mouse click to navigate intertextual space. ■

(*Modeling . . .* continued from page 96)
students should learn better and retain the skills you want to teach.

Conclusion

Those of us who have been developing library instructional tutorials in different formats have been constantly trying to improve our products. I have argued here that the use of more images is required when creating Web tutorials if we are to take full advantage of the concept of modeling in social learning theory. Because a large part of student learning is visual, we need to use images that stimulate, motivate, and reinforce our tutorial content; using a familiar and appropriate metaphor is a useful way of doing this.

Notes

1. Albert Bandura, *Social learning theory* (Englewood Cliffs, N.J.: Prentice Hall; Toronto: Prentice-Hall of Canada, 1977).

2. Library Research Instruction, Faculty of Applied Health Sciences, Brock University Web page at <http://spartan.ac.brocku.ca/~dsuarez/physeduc/index.html>.

3. Library Research Instruction Department of Community Health Sciences Web page at http://spartan.ac.brocku.ca/~dsuarez/physeduc/chsgen_develop.htm.

4. Exercise Quiz, Department of Recreation and Leisure Web page at http://spartan.ac.brocku.ca/~dsuarez/physeduc/reclgeneric_quiz.htm. ■

NEW RELEASES FROM APA BOOKS



**Rethinking the DSM:
A Psychological Perspective**
Edited by Larry E. Beutler
and Mary L. Malik

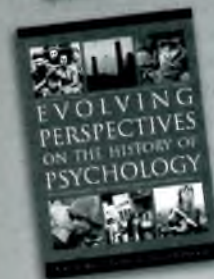
2002. 288 pages. Hardcover.
List: \$39.95 • APA Member/Affiliate: \$34.95
ISBN 1-55798-841-2 • Item # 431668A



**A Beginner's Guide
to the MCMI-III**

Dan Jankowski
Foreword by Theodore Millon

2002. 221 pages. Hardcover.
List: \$39.95 • APA Member/Affiliate: \$34.95
ISBN 1-55798-843-9 • Item # 431670A



**Evolving Perspectives
on the History of
Psychology**

Edited by Wade E. Pickren
and Donald A. Dewsbury

2002. 608 pages. Softcover.
List: \$39.95
APA Member/Affiliate: \$34.95
ISBN 1-55798-882-X • Item # 4318998



**Self and Motivation:
Emerging Psychological
Perspectives**

Edited by Abraham Tesser,
Diederik A. Stapel, and Joanne V. Wood

2002. About 327 pages. Hardcover.
List: \$39.95 • APA Member/Affiliate: \$34.95
ISBN 1-55798-883-8 • Item # 4318999

**On the Consequences
of Meaning Selection:
Perspectives on Resolving
Lexical Ambiguity**
Edited by David S. Gorfein

2002. About 360 pages. Hardcover.
List: \$39.95 • APA Member/Affiliate: \$34.95
ISBN 1-55798-879-X • Item # 4318997



**The Psychodynamics of
Gender and Gender Role**

Edited by Robert F. Bornstein
and Joseph M. Masling

2002. 368 pages. Hardcover.
List: \$49.95 • APA Member/Affiliate: \$39.95
ISBN 1-55798-894-3 • Item # 431674A



**Self-Injurious Behavior:
Gene-Brain-Behavior
Relationships**

Stephen R. Schroeder,
Mary Lou Oster-Granite,
and Travis Thompson

2002. About 312 pages. Hardcover.
List: \$39.95 • APA Member/Affiliate: \$34.95
ISBN 1-55798-885-4 • Item # 431673A



1-800-374-2721
WWW.APA.ORG/BOOKS

In Washington, DC
Call: 202-336-5510
TDD/TTY: 202-336-6123
Fax: 202-336-5502

From Europe, Africa,
or the Middle East:
Call: 44-20-7240-0856
Fax: 44-20-7379-0609

