

## Homer to Home-Page: Designing Digital Books

*William J. Mitchell*

"Tell me, O Muse, of the man of many devices, who wandered full many ways . . ." Are we about to hear of a cybnaut surfing the Net? Actually, as the dwindling band of the classically educated will recognize, this is a popular translation of the opening line of the *Odyssey*.

But it's also an excellent starting point for thinking about the character and uses of text in an online world since, in the days of Homer, words had no material embodiment; they floated freely in the air, and faded away as the itinerant poet ceased to speak. In the thousands of years since, humankind has figured out innumerable ways to bind words permanently to matter, to carve them into clay and stone, to print them on paper, to form them out of unlikely things like neon tubes, and furtively to spray them onto walls. Now, in some ways, we're back where we started. If I want to consult the text of the *Odyssey*, I no longer bother to seek out the tattered volume that's somewhere on my shelves; I just call up Net Search, type in some keywords and click a couple of times, and the bits that I want come flowing down the line to my laptop computer. The ancient text has finally been freed from its long enslavement to materiality; it inscribes itself briefly on my screen, then disappears when I click to dismiss it.

Don't get me wrong. I still love the feel of that old clothbound volume in my hands. I cherish the memories it evokes. I do feel a little guilty about leaving it to gather dust. But the attractions of the newcomer are just too seductive to ignore. Without having to carry a weighty package of paper around with me, I can get to the digital version at any time, from anywhere in the world. It doesn't cost me anything. It's never unavailable because it's been borrowed by someone else. I need not fear losing it by accidentally leaving it somewhere. Since it doesn't have a limited number of physical copies, it cannot go out of print. I can instantly copy quotations (without worrying about transcription errors), and paste them into texts, like this

very one, that I am constructing myself. I can click on hot-linked words to discover where they show up in other ancient Greek texts. And (if I were scholar enough to find these capabilities useful) I could go back to the original Greek at any point and click on words to find dictionary entries, run morphological analyses, and even analyze frequencies of occurrence in different contexts. Finally, I can even make a hard copy whenever I need that for some reason. The digital text has new pleasures. Does this make the printed text obsolete? Will printers, binders, bookshops and libraries soon be things of the past? I don't think so. But the online digital text does take over some of the traditional functions of ink on paper, and it does enable some strikingly new ways of producing, transforming, and using literary material. Its emergence requires writers to reconsider their craft, it forces designers to rethink the task of making language visible, and it leaves publishers anxiously scrambling to find new business models.

### The Case of City of Bits

In 1995 I had a chance to explore these questions in a practical context when, with the MIT Press, I published my book *City of Bits*. Since it dealt with the digital revolution and the new relationships that were being created between the material and virtual worlds, we decided that it should be self-exemplifying—that it should appear simultaneously as a hardback and in a full-text World Wide Web version. As far as I

know, it was the first book to be published simultaneously in print and on the Web. (At the very least, it could not have had many predecessors.)

We made the marketing people happy by providing a link to an online order form from the opening screen of the Web site; enter your name and address, include your credit card number (in a secure transaction), click to transmit your order, and a copy gets sent to you immediately. Conversely, we published the URL (the address in cyberspace) of the Web version on the dustjacket of the print version. So a reader of either one could always conveniently obtain the other.

We provided free access to the Web version. (As the Web develops, convenient mechanisms for charging for access to online material are being put in place, and these will obviously be crucial to the development of an online publishing industry. But these were not highly developed when we put *City of Bits* online, and attempting to charge just didn't seem worth the trouble at that point. There was some risk in this, of course; why would anyone buy a copy when the online version was right there at no cost? Perhaps we would lose sales. But we guessed that the additional sales generated by the Web site would outweigh such losses, and there is some good evidence that we were right; in the first two printings, about 2% of the total sales were directly through the online order form, and it is likely that the Web site also stimulated bookstore and mail-order sales. Why should this be so? The answer is that the hardback and online versions added value to the text in different and complementary fashions. (The dimensions of that complementarity will be explored in the discussion that follows.) So readers of the Web version are not necessarily potential customers for the hardback. And lots of people decided that they wanted both, to use in different ways.

#### Hardback, Paperback, and No-back

Of course, publishing a book in different versions is not a new idea; it has long been a common strategy to put out both hardbacks and paperbacks. The hardback is more expensive and more robust, and it is aimed at libraries and at buyers who want to keep it permanently on their bookshelves, while the paperback is cheaper is not designed to have such a long life. Depending on the content and the marketing strategy for a particular book, it may appear in hardback only, in paperback only, in paperback with a small number of hardbacks for sale to libraries, or in hardback followed by a less expensive paperback at a later point.

With the Web, the online no-back emerges as a third option at the inexpensive and ephemeral end of the spectrum. It can be used, even by very small publishers, to achieve instant world-wide distribution; certainly we found, with *City of Bits*, that it was quite widely read and even reviewed in some countries long before copies of the hardback were available there. But, since publishers generally have not begun to guarantee the permanent existence of Web sites, you still need a hardback copy if you want to be sure of continued access in the future.

You may also want a well designed, well produced print version for ease of extended reading, portability, and just the sheer pleasure of it. By comparison with even the very best laptop computer, a well-made book is light, tough (you can drop a book without damaging it, but not a laptop), comfortable in the hand, and usable anywhere. It has an extremely high-contrast, high-resolution display, and the access mechanism (turning pages) is a lot nicer than using a mouse and cursor to scroll text down a screen. Indeed, I have often thought that, if Gutenberg had invented the personal

computer and printed books had not appeared until the 1980s, we would now be hailing paper and print as a major technological advance!

As forward-looking computer technologists will be quick to point out, things won't stay this way. Computers will become lighter, less fragile, and more portable. The quality of displays will improve. Sophisticated home and office printers will allow production of high quality, personalized print copies on demand. We may even see the emergence of programmable "smart paper" , allowing development of devices that combine the virtues of the portable computer and the book. But, for the moment at least, the hardback, the paperback, and the electronic no-back have significantly different properties and roles.

#### Getting the Reader's Attention

The first task of a book , especially a trade book that's supposed to attract an audience , is to get itself picked up and read. So the hardback City of Bits has a vivid, colorful dustjacket to catch the reader's attention; it's carefully designed to stand out on a bookstore display or a library shelf. When you take it in your hand, you find a brief description and author biography on the flyleaf. Then you can flip through it to see what's inside.

The Web version clearly had to attract attention in very different ways, and making sure that it did so was a key to success. Several strategies were used.

First, a hot-link was made from the entry in the MIT Press's online catalogue to the City of Bits site. So , much as bookstore browsers can pick up a copy of the hardback , Web-surfing catalogue browsers can immediately get their hands on the online version. And the first thing that the online version presents is a Welcome page with links to a Synopsis, the author's Home Page, and the Table of Contents. Thus, to provide one path into the online City of Bits, the metaphor of an "electronic bookstore" was fairly closely followed.

Hot-links from other Web sites provide a second way in. City of Bits was quickly listed in many online, classified Internet and Web guides, "Cool Sites" collections, online newsletters and magazines, home pages of organizations and individuals who wanted to draw attention to it, and online reading lists for classes of various kinds. Some of these links were sought and negotiated by members of the City of Bits WWW team, but many appeared spontaneously. Most were one-way, from the other site to City of Bits, but some were reciprocal , a fixed "you point to me and I'll point to you" arrangement. The ultimate effect was to create a very large, electronic "catchment" to collect potential readers and efficiently funnel them to the site.

The third strategy for bringing in readers is to attract the attention of Web search engines. Typically, these engines explore the Web periodically to create large indexes and directories, then, in response to users' queries, employ these indexes and directories to provide very rapid access to the relevant Web sites. They perform their explorations in a variety of ways-by looking for specified keywords in the titles or headers of Web documents, by scanning through the documents themselves, or even by searching other indexes and directories. They are usually pretty dumb, since they just look for keyword matches. So, to make sure that your site is not missed by the search engines , which have now become very important tools for finding one's way around the Web , you must make sure that the appropriate descriptors are included in titles and headers, and in the text of the opening pages. Incidentally, you can reliably

attract a lot of attention by scattering words like "sex" and "nude" through your text , but it may not be the sort of attention that you want!

A fourth possible strategy, which we have not used, is closely analogous to pinpoint direct-mail marketing. When Web-surfers access your server, it is technically possible to collect a lot of information about them , who they are, where they are from, what links they followed to get to your site, what browser they were using, what they looked at, and so on. If you are prepared to ignore the obvious privacy issues, you can use this information to target electronic advertising. So, for example, Web-surfers who looked at MIT Press online catalogue entries for other books on related topics might get email promoting City of Bits.

#### Reading Tools and their Effects

In traditional fashion, the hardback version of City of Bits is a narrative divided into chapters on different sub-topics and it has a table of contents and an index to guide the reader through the material. This allows for multiple styles of reading; you can follow a continuous thread straight through from beginning to end, you can jump immediately to particular chapters that interest you, you can use the index to find passages on particular topics, and you can even cruise the index (or the endnotes) to look for entries that may pique your interest. You can skim quickly or you can read more slowly and attentively. You may make notes as you go, or you may not. You may read in strict sequence, or you may jump back and forth.

The physical book is not only a repository of the textual information, but also a reading tool that allows you to pursue these strategies efficiently, and gives you context and feedback as you do so. Its size and shape tells you roughly how much information it contains, and you always know how far through it you are from the relative thicknesses of the stacks of pages under your left and right thumbs. The springiness of the paper allows you to scan quickly by riffling through pages with the book half open, but the mechanical properties of the binding assure that you can also leave it open, flat on a desktop, for more extended and careful study.

Typography signals the hierarchy of information by visually distinguishing headings, sub-headings, and body text. A Table of Contents right at the front, an Index at the very back, and numbered pages, provide effective search and navigation capabilities. Endnotes, with numbered references from the text, allow backup information to be provided without disrupting the flow of the narrative.

The online version provides very different reading tools. Most dramatically, there is no index; it is replaced by an internal search engine that locates instances of user-entered keywords in the text. From the author's viewpoint, this eliminates the intellectual drudgery of creating an index. From the reader's viewpoint, it provides greater freedom; you can search for anything, and you don't have to rely on the author's judgment about what was worth including in the index. (I'm told, for example, that many readers immediately type in their own names to see if they're mentioned anywhere!)

The hierarchy of information is also handled differently in the online version, since the screen can only display a limited amount of text at one time, since current bandwidth constraints make it undesirable to download large text files to your browser all at once, and since scrolling through a long segment of text doesn't work nearly as effectively as flipping the pages of a book. The complete text is organized into a hierarchy of small

segments, with internal hot-links providing the interconnections among them. At the top of the tree is the Table of Contents page providing entry points to each of the chapters. Within each chapter, there is the introductory section of text followed by hot-links to the subsections that it contains. Finally, there is the relatively short text of each subsection. To allow for sequential reading of the narrative, without having to go up and down the hierarchy, there are "previous" and "next" hot-links at the end of each subsection.

Endnotes, of course, are handled by hot-links; click on the endnote mark and you immediately get the corresponding note. (Cross-references within the text could be handled in a similar way, but there aren't any.) To maintain consistency with the print version, and continuity with tradition, the notes are numbered-but, of course, they no longer really have to be, since there's never any ambiguity about which note relates to which point in the text.

Overall, the reading tools provided with the online version have a very interesting effect; they privilege the hierarchical structuring of the book's content and the operation of searching while they make sequentially following the narrative more cumbersome and difficult. (It's no accident, then, that CD-ROM and online books that have these sorts of reading tools have tended to emphasize modular, classified and indexed chunks of content as in encyclopedias and dictionaries, to provide dense cross-referencing within the material, and to construct multi-threaded and branching narratives-in other words, to focus on anything other than long, continuous narrative sequences.) The hardback, on the other hand, privileges skimming, random jumps back and forth, and the continuity of the main narrative thread. So it's probably optimal to read the hardback first, to gain an overview, then to go to the online version for more detailed study and for ongoing reference.

#### Fixed-Format and Personalized

Good graphic designers exert very considered and precise control over the look and feel of a printed book. Certainly this was the case with *City of Bits*. The designer, Yasuyo Iguchi chose to set it in Bembo and Meta. She arranged elements on the various different sorts of pages, and deployed white space with care.

She gave consideration to its size, shape, proportions, weight, and rigidity. She chose the paper, the cloth for the cover, and the matte varnish of the jacket so as to create a particular relationship of feels and textures. All of this matters. It all adds up to something that has the characteristic look of a MIT Press book, and that signals something about the product's style, content, and level of sophistication.

But the client-server architecture of the Web does not allow a designer such precise control of the online version; it may be downloaded to many different types of display devices, by many different types of browsers, with many different settings of their various options, to produce screen displays that vary enormously. This can be seen as a disadvantage (and typically is by graphic designers, who don't like the loss of control), and the producers of Web servers and browsers can try to eliminate as many sources of unwanted variation as possible. Or it can be seen as an advantage-opening up the possibility of adapting content intelligently to different contexts and to the needs of different readers; perhaps every reader of *City of Bits* could have a uniquely personalized version.

The issue of producer-control versus user-personalization is a philosophical rather than a technical one; it is technically feasible to implement systems that support either one or both, and to design online productions that either go for a consistent look or encourage personalization. In the online version of *City of Bits*, we tried to exert as much control as possible, to assure a reasonably high level of graphic quality, to remain consistent with the print version, and just to keep things simple for ourselves. But, as personalization tools become increasingly sophisticated, it will become more interesting to try to take advantage of them.

#### External Hot-Links

Perhaps the most obvious and striking difference between the hardback and the online version is that the text of the online version contains hundreds of hot-links to other Web sites with relevant information on the topics that are discussed. When I discuss online shopping malls, for example, you can just click to go and visit one. And, when I refer to Aristotle's *Politics*, you can immediately access the relevant passage, online, in either English or Greek. Thus the *City of Bits* site becomes a conveniently organized entry point for exploring an enormous quantity of related information.

Some of these external hot-links are to sites that I or my research assistant discovered and consulted when *City of Bits* was being written, but the vast majority have resulted from systematically going through the text, picking out key words, and sending search engines out on the Web to find what was out there.

Whenever a search engine discovers a relevant site, we link it in. (You can think of this as a new form of bricolage.) This process has to be repeated at regular intervals, since the Web is growing explosively, and relevant new sites are continually appearing. So the structure of intertextual linkages in which *City of Bits* embeds itself is a very dynamic thing, and it looked very different, after the site had been up for a few months, than it did when it first went online.

The converse process is to combat link-rot by identifying and removing hot-links to sites that have died, shifted to new locations, or become irrelevant. (If this is not done, a site quickly loses its charm, like an untended garden.) To facilitate this, we employ a software tool that automatically runs through the text, checks all the hot-links, and reports all those that don't seem to be working.

Superficially, adding these links may just seem to be a more convenient way to provide endnote citations to related publications. But, on closer inspection, there are some important differences. One is the dynamism that I have noted; print endnotes can only be updated, all at once, when there is a reprint or a new edition, but hot-links can be updated incrementally and at any time. Furthermore, you cannot add too many endnotes to a printed book without making it bulky and unwieldy, but there is no practical limit to the number that you can embed in an online text.

But the most important difference is the shift in scholarly responsibility, and correspondingly in the reader's use of the text, that the substitution of hot-links for endnote citations entails. Recall that endnote citations are normally to printed documents that have been formally published and do not change. A responsible scholar is expected to check the relevance, quality, and usefulness of a cited document, and to give publication date and page numbers; scholars who cite irrelevant or poor-quality publications are not highly regarded. But the author of an online publication cannot attempt to take the same responsibility, since the contents of an externally linked site

may change unpredictably, at any time; I might, for example, discover a site containing the text of Aristotle's *Politics*, check it out and assure myself that everything was in order, and then make the link from City of Bits, only to discover, some time later, that the operator of that site had subsequently substituted several hundred pornographic GIF files for the philosopher's words. So, external hot-links are very useful, but they have their dangers. Caveat surfer!

As the Web and similar structures mature, there will undoubtedly be an increasing number of sites providing stable, "guaranteed" content, and scholars will have less of a problem. There are, for example, already some refereed online technical journals. But the medium does not automatically enforce document stability in the way print does, so special institutional arrangements will be needed in contexts where such stability is necessary.

#### Marginalia and Readers' Comments

Sometimes readers like to scribble their comments in the margins of printed books, and sometimes subsequent readers see these comments and may even add their own responses, but this usually isn't encouraged (particularly with library books) and it isn't a very effective form of discourse. By contrast, online versions of books can easily provide for readers to add their comments, and for these comments to be widely available.

In the online City of Bits, readers can enter an electronic "agora" directly from the site's front door, or from the foot of any page of text. There, they can read the (comments) that other readers have posted. They can also use a simple form to add their own comments. And they can even insert hot-links to other sites that they consider relevant. This agora is organized as a collection of newsgroups, and provides all the usual features of newsgroup support software.

Over time, then, the online version of City of Bits has become encrusted with commentary. It has succeeded in provoking, capturing, and making visible a discourse in a way that is impossible with print. And, in the process, the seed provided by the original text has grown into a considerably larger and richer textual structure.

This evolution is fascinating and exciting to see, but it creates some theoretical conundrums and practical difficulties. The continually growing, transforming structure is actually the work of many hands, yet it has my name on it. In the beginning, it was mostly mine, but it becomes less and less so as time goes on and the online comments accumulate. At what point does it become inappropriate to say that it is "my" text? When does it become more reasonable to call it a collective work?

Who bears moral and legal responsibility for it? Should I treat the agora as a zone in which complete freedom of speech is permitted, or should I, as the author, take responsibility for actively moderating and shaping the discussion? Should I delete blatantly irrelevant and self-serving comments? What if advertisements are posted? What if a reader were to post comments that I found personally offensive and insulting? (Am I obliged to provide that person with a platform?) What if a posting were found to contain slanderous or obscene material, or a neo-Nazi diatribe? These are not the sorts of questions that arise about scribbled marginal comments in printed books, but they have been hotly debated in relation to online newsgroups and bulletin boards. A book becomes a thing of a different kind when it systematically internalizes

and reports back the discussion that it has provoked, rather than standing distinct, closed, and aloof from it.

These seem difficult questions, and general answers will probably have to be worked out through experience and debate. In the case of City of Bits, the team that maintains the site has taken a rigorous "hands off" attitude; we occasionally go through and clean out the completely irrelevant postings that sometimes appear, but we leave everything else there. Generally, comments so far have been serious and responsible, so we have not been forced to confront any really troublesome dilemmas. Perhaps we have just been lucky, though.

#### Reviews, Mentions, and Translations

Any successful book soon generates a growing body of thematically related, secondary, and derivative texts, reviews, commentaries, news articles, mentions in other works, and translations. The City of Bits site keeps a running record of this sort of material (to the extent that the team can keep up with it) and, where possible and appropriate, provides links to it.

As it turned out, City of Bits generated a lot of interest, and quickly received many reviews in both the specialist and mainstream media. Perhaps naively, we had hoped that we might add the full texts of all reviews to the site as they appeared. That would have made accessible another, extremely interesting, layer of commentary and elaboration. But the world is not quite ready for that; after a few attempts to secure permissions to reproduce complete reviews online, and generally getting rebuffed or asked to pay exorbitant fees, we retreated to the position of posting short extracts, much as they have traditionally been reproduced in jacket copy and advertisements. In future, though, it may not be so difficult to achieve our original ambition; when the majority of reviews appear in online editions of newspapers and magazines, and the like, it will only be necessary to link to them.

As translation rights have been sold, details on the forthcoming foreign-language editions have been posted in a Translations section of the site. When the translations are completed, we will explore further possibilities. (This will require making new and unusual types of agreements with the overseas publishers, and it is not yet clear how these will work out.) For example, we might simply add online texts of the foreign-language versions to the City of Bits site. We might go further, and provide structures of cross-linkages among the English and foreign-language versions so that multilingual readers might conveniently move back and forth, a particularly useful capability where words and phrases do not have very exact equivalents in other languages, or where there might be ambiguity or debate about the best way to translate things. Or we might encourage the foreign publishers to develop their own Web sites for the translations, then build links to and from. In the more distant future, it is easy to imagine online books existing as multilingual, geographically distributed sites in which you are asked, on entry, what language you want to use—as in American Express cash machines.

#### Online Appropriation

In effect, the various external linkages from the City of Bits site appropriate a vast array of existing textual fragments and combine them to form a new work, something that, because of the selection and organization that goes into it, is significantly greater than the sum of its parts. The original City of Bits text, as published on paper, is just

one of these constituent fragments , though, to be sure, a privileged one. (This shifts to a radically new context the old idea, recognized in intellectual property law, that a collection can be a creative work.)

This strategy of textual appropriation and collage does not run into the sorts of intellectual property difficulties that would arise in creating a large, cross-referenced print collection, since the constituent fragments are merely pointed to rather than reproduced. The author of an appropriated text does not lose anything in this way. On the contrary, authors usually post texts online because they want them to be noticed and read, so it is an advantage to attract linkages that might channel readers from other texts and sites.

In sum, an important new literary role has now emerged , that of the link- editor who locates fragments of text online and combines them into original literary structures by superimposing patterns of linkages. On a large scale, the operators of Internet guides like Yahoo! play the link-editor role by selecting and classifying online material and providing convenient point-and-click access from a topic list. Pedagogues play the game when they link words in books and articles to online reference works , dictionaries, encyclopedias, and so on. Critical scholars play it when they create structures of comparisons and contrasts among texts. The City of Bits team certainly played it when they constructed the online version. And, by now, the online City of Bits has been appropriated into a great many online constructions created by others.

When I have discussed this form of appropriation with other authors, some of them have been greatly disconcerted by the idea. They do not like the possibility that their work might be used in ways they cannot control and for purposes that they never intended. (They forget, of course, that authors have never really had very much control over the uses and misuses of their published texts. But embedding in online link structures does make this possibility dramatically explicit.) Others, including myself, are excited by being able to see with new clarity the evolving roles that their texts play in ongoing discourses.

#### Stabilities and Instabilities

As we have now seen, the online City of Bits has both stable and unstable elements. The core text, which corresponds to that of the print version, does not change. But the structure of links that it carries is continually adjusted and extended, the contents of the externally linked sites evolve, and the accreted structure of comments, reviews, and translations grows. If I decide to do new print editions, I expect to add the text of those to the online version, and to preserve the earlier edition texts as well. Thus any change in the core text will be carried out in well-defined, modular increments.

A more radical possibility would be to make continual small changes to the core text to reflect new developments and to respond immediately to comments and criticisms. (There is no technical difficulty in doing so.) That way, the text would be kept constantly up to date; there would be no need to keep using an increasingly obsolete and unsatisfactory text while waiting for the right moment to put out a complete new edition. But this would destroy the logical integrity of references within the overall structure. What if, for example, a reader's comment refers to a specific paragraph in the core text and that paragraph is subsequently deleted or significantly altered?

Perhaps the most satisfactory approach would be to preserve successive versions as incremental changes are made. Some fairly straightforward software could then

automatically relate comments and other linked material to the appropriate versions. So far, though, we have not had the energy or the disk space for that.

Whatever the balance between stable and unstable elements, though, you never read the same text twice. (Heraclitus would have loved it!) Even the internally stable elements are continually being recontextualized, and so shift in their meaning, as the huge structure that embeds them transforms itself. Furthermore - an alarming thought for historians, it is quite impossible to preserve more than a very partial record of the past states of that transforming structure; it has no distinct boundaries, it is distributed over many different machines in widely scattered locations, and it is far too large and complex to back up on tape. The printed book appeared to give scholars stable, repeatable text modules to work with. Perhaps that was always a myth. With online books, certainly, that myth is increasingly difficult to sustain.

The End

Hardback and paperback books eventually go out of print. Archival libraries selectively perform the function of preserving books after that point. But what about online books? Since it does take some effort and resources to keep them around, and even more to keep them growing and changing, they are likely to have quite limited lives. How long do they stay available online? What is the electronic equivalent of going out of print? Who is responsible for long-term archiving?

Answers to these questions are likely to vary with the type of book, and may change over time as online publication grows in importance, but I can give a provisional answer for City of Bits online. I regard it as a kind of extended live performance in a vast virtual theater. Eventually, that performance will end. The site that remains will not instantly disappear, but will slowly fade away, like an abandoned stage-set-as link-rot sets in and as additions and updates are no longer made. As time goes by, there will be fewer and fewer visitors.

In the end, the City of Bits will be an electronic ruin. Like Troy, it will cease to function and to live, becoming, instead, part of the archaeology of cyberspace.

**Media in Transition, Nova Iorque, 24-25 out. 1998.**