The information explosion

Moving ahead a thousand years, let us peek at what was happening in a library at a major university near the center of European civilization a century or two after Gutenberg's introduction of the movable-type printing press around 1450. (The printing press was invented in China much earlier, around five centuries before Gutenberg.) Trinity College, Dublin, one of the oldest universities in Western Europe, was founded in 1592 by Queen Elizabeth I. In 1600 the library contained a meager collection of 30 printed books and 10 handwritten manuscripts. This grew rapidly, by several thousand, when two Trinity College Fellows mounted a shopping expedition to England, and by a further 10,000 when the library received the personal collection of Archbishop Ussher, a renowned Irish man of letters, on his death in 1661.

At the time, however, even this collection was dwarfed by that of Duke August of Wolfenbüttel, Germany, whose collection had reached 135,000 imprints by his death in 1666 and was the largest contemporary library in Europe, acclaimed as the eighth wonder of the world. These imprints were purchased in quires (i.e., unbound) and shipped to the duke in barrels. The duke had them bound in 31,000 volumes with pale parchment bindings that you can still see today. Incidentally, this collection inspired Casanova, after spending seven days visiting the library in 1764, to declare that "I have sometimes thought that the life of those in heaven may be somewhat similar to [this visit]." Coming from the world's most renowned lover, this is high praise indeed.

Returning to Ireland, another great event in the development of Trinity College occurred in 1801, when an act was passed by the British Parliament decreeing that a copy of every book printed in the British Isles was to be donated to the Trinity College Library. This privilege extends to this day and is shared by five other libraries—the British Library, the University Libraries of Oxford and Cambridge, and the National Libraries of Scotland and Wales. This "legal deposit" law had a much earlier precedent in France, where King François I decreed in 1537 that a copy of every book published was to be placed in the Bibliothèque du Roi (long since incorporated into the French National Library). Similarly, the Library of Congress receives copies of all books published in the United States. But we digress.

There were no journals in Archbishop Ussher's collection. The first scholarly journals appeared just after his death: the *Journal des Sçavans* began in January 1665 in France, and the *Philosophical Transactions of the Royal Society* began in March 1665 in England. These two have grown, hydra-like, into hundreds of thousands of scientific journals today, although some are now being threatened with replacement by electronic archives.

In the 18th century, printing really took hold. For example, more than 30,000 titles were published in France during a 60-year period in the mid-1700s. Some 300 years after Gutenberg developed the printing press in order to make the Bible more widely available, the press became the vehicle for disseminating the European Enlightenment—an emancipation of human thinking from the weight of authority of the church.

In the United States, President John Adams created a reference library for Congress when Washington became the new national capital in 1800. He began by providing $5,000 "for the pur-

chase of such books as may be necessary for the use of Congress, and for putting up a suitable apartment for containing them therein." The first books were ordered from England and shipped across the Atlantic in 11 hair trunks and a map case. The library was housed in the new Capitol until August 1814, when—in a miniature replay of Julius Caesar's exploits in Alexandria—British troops invaded Washington and burned the building. Some 3,000 volumes were lost in that fire. Another fire destroyed two-thirds of the rebuilt collection in 1851. Unlike Alexandria, however, the Library of Congress has regrown—indeed its rotunda is a copy of the one built in Wolfenbüttel two centuries earlier. Today the Library of Congress contains approximately 22 million volumes.

The information explosion began to hit home in Ireland in the middle of the 19th century. In 1835, work started on the production of a printed catalog for the Trinity College Library (Figure 1.5), but by 1851 only the first volume, covering letters *A* and *B*, had been completed. The catalog was finally finished in 1887, but only by restricting the books that appeared in it to those published up to the end of 1872. Other libraries, however, were wrestling with much larger volumes of information. By the turn of the century, the Trinity College Library had around a quarter of a million books, while the Library of Congress had nearly three times that number. Both were dwarfed by the British Museum (now part of the British Library), which at the time had nearly 2 million books, and the French National Library in Paris, which had over 2.5 million.

The Alexandrian principle

In an early statement of library policy, an Alexandrian librarian was reported as being "anxious to collect, if he could, all the books in the inhabited world, and, if he heard of, or saw, any book worthy of study, he would buy it"—and two millennia later this was formulated as a self-evident principle of librarianship: *it is a librarian's duty to increase the stock of his library*. When asked how large a library should be, librarians answered, "Bigger. And with provision for further expansion."

Only recently has the Alexandrian principle begun to be questioned. In 1974, following a 10-year building boom then unprecedented in library history, the *Encyclopedia Britannica* noted that "even the largest national libraries are … doubling in size every 16 to 20 years" and gently warned that "such an increase can hardly be supported indefinitely." And the struggle continues. In the 20th century's last decade the national libraries of the United Kingdom, France, Germany, and Denmark all opened new buildings. The ones in London and Paris are monumental in scale. Standing on the bank of the Seine River, the Bibliothèque Nationale de France consists of four huge towers designed to look like open books, surrounding a sunken garden plaza (Figure 1.6). The reading rooms occupy two levels around the garden, with bookshelves encircling them on the outer side.

Sustained exponential growth cannot continue. A collection of essays published in 1976 entitled *Farewell to Alexandria: Solutions to Space, Growth, and Performance Problems of Libraries* dwells on the problems that arise when growth must end. Sheer limitation of space has forced librarians to rethink their principles. Now they talk about "aggressive weeding" and "culling," "no-growth libraries," the "optimum size for collections," and some even ask, "Could smaller be better?" In a striking example of aggressive weeding, the library world was rocked in 1996 by allegations that the San Francisco Public Library had surreptitiously dumped 200,000 books, or 20 percent of its collection,

— Resolutie van de staten generael der Vereenighde Nederlanden, dienende tot antwoort op de memorie by de ambassadeurs van sijne majesteyt van Vranckrijck.
's Graven-hage, 1678. 4°. Fag. H. 2. 80. N°. 20.
Fag. H. 2. 85. N°. 17. Fag. H. 3. 42. N°. 4.

— Tractaet van vrede gemaeckt tot Nimwegen op den 10 Augusty, 1678, tusschen de ambassadeurs van [LOUIS XIV.] ende de ambassadeurs vande staten generael der Vereenighde Nederlanden.
Fag. H. 2. 85. N°. 21.

— Nederlantsche absolutie op de Fransche belydenis.
Amsterdam, 1684. 4°. Fag. H. 2. 50. N°. 22.

— Redenen dienende om aan te wijsen dat haar ho. mog. [niet] konnen verhindert werden een vredige afkomst te maken op de conditien by memorien van den grave d' Avaux van de 5 en 7 Juny, 1684, aangeboden.
[*s. l.*] 1684. 4°. Fag. H. 2. 86. N°. 3.
Fag. H. 2. 96. N°. 8. Fag. H. 3. 44. N°. 52.

— Redenen om aan te wijsen dat de bewuste werving van 16000 man niet kan gesustineert werden te zullen hebben konnen strekken tot het bevorderen van een accommodement tusschen Vrankrijk en Spaigne.
[*s. l.*] 1684. 4°. Fag. H. 2. 86. N°. 4.
Fag. H. 2. 96. N°. 2.

— D' oude mode van den nieuwen staat van oorlogh.
[*s. l.* 1684]. 4°. Fag. H. 2. 86. N°. 12.
Fag. H. 2. 96. N°. 3.

— Aenmerkingen over de althans swevende verschillen onder de leden van den staat van ons vaderlant.
[*s. l.*] 1684. 4°. Fag. H. 2. 92. N°. 1.
Fag. H. 2. 98. N°. 16. Fag. H. 3. 1. N°. 18.

— Missive van de staten generael der Vereenighde Nederlanden, . . . 14 Maert, 1684.
's Graven-hage, 1684. 4°. Fag. H. 2. 92. N°. 10.

— Missive van de staaten generael der Vereenigde Nederlanden, . . . 11 July, 1684.
[*sin. tit.* 1684]. 4°. Fag. H. 2. 96. N°. 13.
Fag. H. 3. 44. N°. 69.

— Resolutie vande staten generael der Vereenighde Nederlanden, . . . 2 Maart, 1684.
's Gravenhage, 1684. 4°. Fag. H. 2. 92. N°. 11.
Fag. H. 3. 44. N°. 9.

— Extract uyt de resolutien van de staten generael, . . . 31 Maert, 1684.
[*s. l.*] 1684. 4°. Fag. H. 2. 92. N°. 13.
Fag. H. 2. 96. N°. 25. Fag. H. 3. 44. N°. 11.
Fag. H. 3. 44. N°. 15.

— Antwoort van de staten generael der Vereenighde Nederlanden op de propositie van wegen sijne churf. doorl. van Ceulen, Maert 23, 1684, gedaen.
's Gravenhage, 1684. 4°. Fag. H. 2. 92. N°. 12.

Figure 1.5: A page of the original Trinity College Library catalog

Figure 1.6: The Bibliothèque Nationale de France. Dominique Perrault, architect; © Alain Goustard, photographer

into landfills, because its new building, although lavishly praised by architecture critics, was too small for all the books.

The notion of focused collections is replacing the Alexandrian model of an ideal library that is vast and eternally growing. The notion of service to library users is replacing the idea of a library as a storehouse of all the world's knowledge. These movements will surely be reinforced by the experience of the World Wide Web, which amply illustrates the anarchy and chaos that inevitably result from sustained exponential growth. The events of the last quarter century have even shaken librarians' confidence in the continued existence of the traditional library. Defensive tracts with titles like *Future Libraries: Dreams, Madness and Reality* deride "technolust" and the empty promises of the technophiles.

Early technodreams

Let us, for a moment at least, consider the technophiles. Around 1936, science fiction writer H. G. Wells promoted the concept of a "world brain" based on a permanent world encyclopedia that

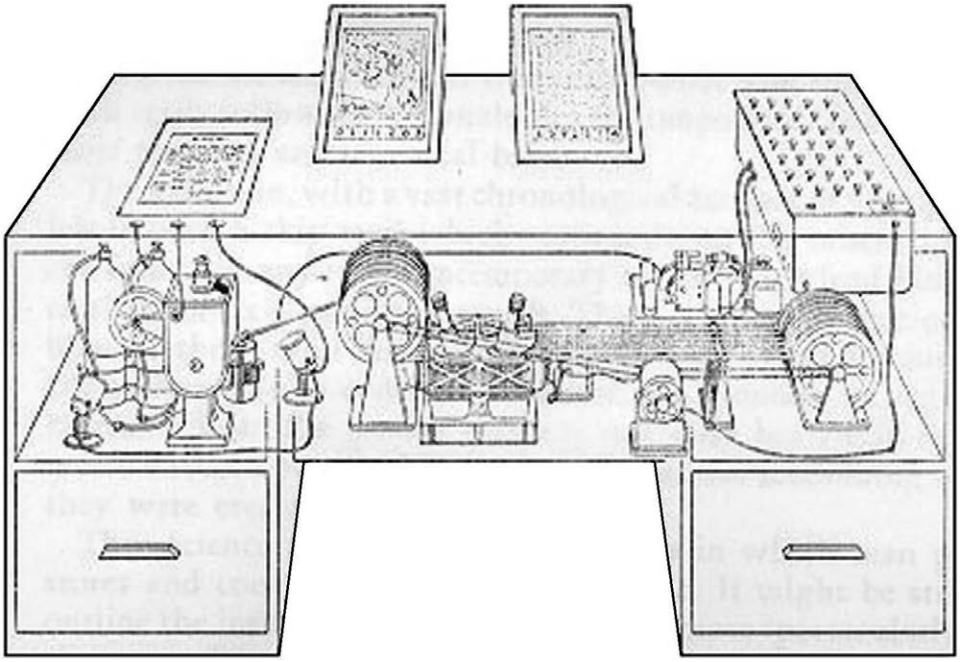


Figure 1.7: Artist's conception of the Memex, Bush's automated library. Courtesy of Mary and Alfred Crimi Estate

"would be the mental background of every intelligent [person] in the world. It would be alive and growing and changing continually under revision, extension and replacement from the original thinkers in the world everywhere," and he added sardonically that "even journalists would deign to use it."

In 1945, Vannevar Bush, the highest-ranking scientific advisor in the U.S. war effort, urged us to "consider a future device for individual use, which is a sort of mechanized private file and library … a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility" (Figure 1.7).

In 1960 J. C. R. Licklider, head of the U.S. Department of Defense's Information Processing Techniques Office, envisioned that human brains and computing machines would be tightly coupled together and supported by a "network of 'thinking centers' that will incorporate the functions of present-day libraries together with anticipated advances in information storage and retrieval."

Toward the end of the 20th century we became accustomed to hearing similar pronouncements from the U.S. Presidential Office, rising above the road noise of the information superhighway.

The library catalog

Wells, Bush, Licklider, and other visionary thinkers were advocating something very close to what we might now call a *virtual library*. To paraphrase the dictionary definition, something is *virtual* if it

exists in essence or effect though not in actual fact, form, or name. A virtual library is a library for all practical purposes, but a library without walls—or physical books.

In truth, a virtual representation of books has been at the core of libraries right from the beginning: the catalog. Even before Alexandria, libraries were arranged by subject and had catalogs that gave the title of each work, the number of lines, the contents, and the opening words. In 240 B.C. an index was produced to provide access to the books in the Alexandrian library that was a classified subject catalog, a bibliography, and a biographical dictionary all in one.

A library catalog is a complete model that represents, in a predictable manner, the universe of books in the library. Catalogs provide a summary of library contents. Today we call this *metadata*. And it is highly valuable in its own right. As a late 19th-century librarian wrote, "Librarians classify and catalog the records of ascertained knowledge, the literature of the whole past. In this busy generation, the librarian makes time for his fellow mortals by saving it. And this function of organizing, of indexing, of time-saving and thought-saving, is associated peculiarly with the librarian of the 19th century."

Other essential aids to information-seeking in libraries are published bibliographies and indexes. Like catalogs, these are virtual representations—metadata—and they provide the traditional means of gaining access to journal articles, government documents, microfiche and microfilm, and special collections.

A possible interpretation of "digital library" that is quite different from the concept developed in this book is a system designed to automate traditional library functions by helping librarians to manage physical books. Online public access catalogs (OPACs) are standard in libraries today. And many other functions are automated: acquisitions, loans, recalls, interlibrary services, and library planning. However, this kind of library automation system is not closely related to the digital libraries we describe in the examples that open this chapter.

The changing nature of books

The technophile visionaries whose dreams we shared above were not talking about a virtual library in the sense of an automated catalog. They wanted the full text of all documents in the library to be automated, not just a metadata surrogate. They took it for granted that books are adequately represented by the information they contain: the physical object is of no consequence.

The information in library catalogs and bibliographies can be divided into two kinds: the first having reference to the contents of books; the second treating their external character and the history of particular copies. Intellectually, only the abstract content of a book—the information contained therein—seems important. But the strong visceral element of books cannot be neglected and is often cited as a reason why book collections will never become "virtual."

Portable writing began with the clay tablet, invented by the Sumerians in 3500 B.C. A well-known example is the Phaistos Disk from the ancient Minoan civilization (Figure 1.8a). A thousand years later, Egyptians began writing on scrolls made of papyrus plants (a practice that was adopted by various cultures over the centuries, see Figure 1.8b). In other places, animal skins were used for

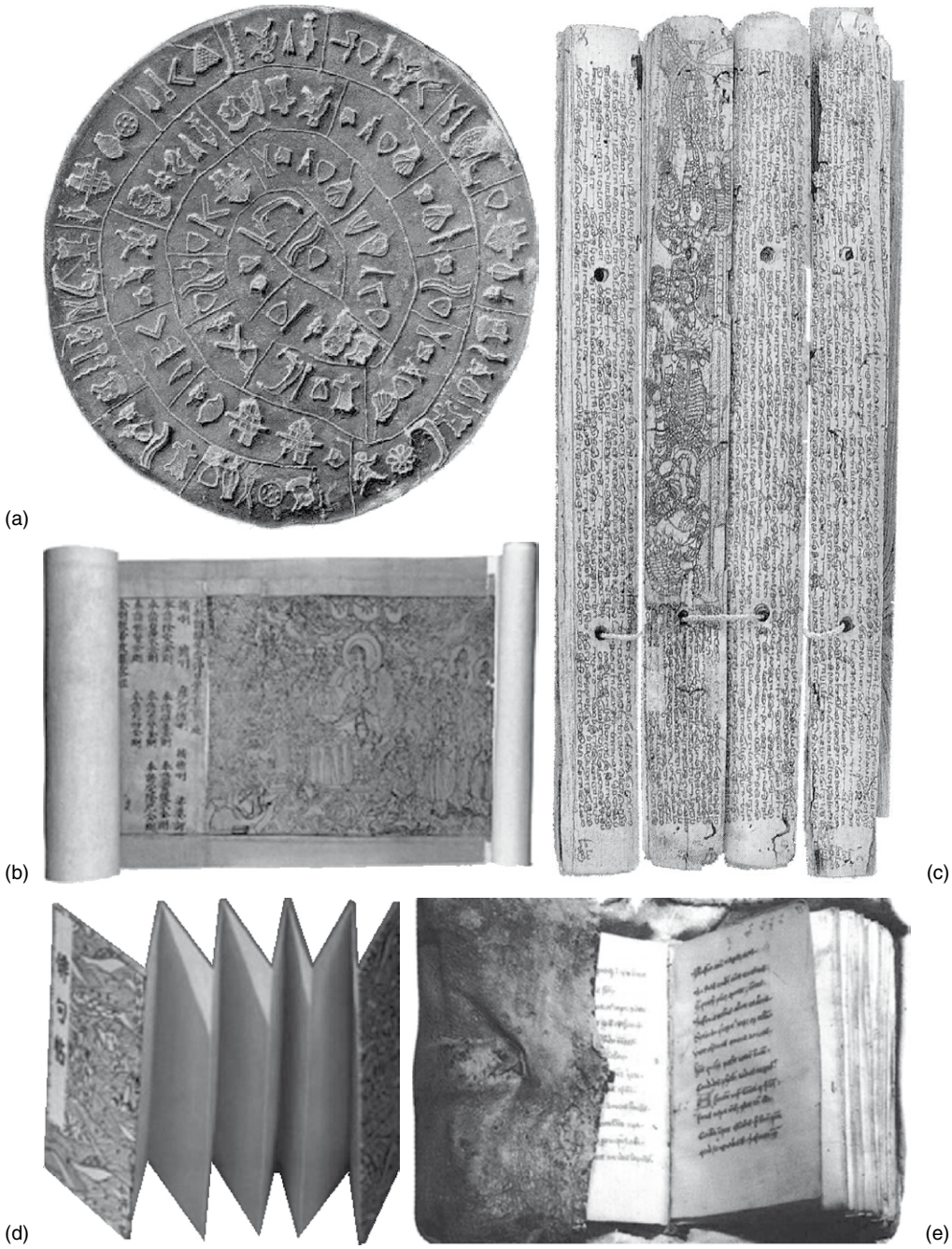


Figure 1.8: Ancient devices for portable writing: (a) clay tablet; (b) scroll; (c) palm leaves (from Thanjavur Maharaja Serfoji's Sarasvat Mahal Library, Thanjavur, Tamil Nadu, 1995); (d) concertina; (e) codex

parchment or vellum; elsewhere, bamboo, silk, and even palm leaves served as writing surfaces (Figure 1.8c). But lengthy scrolls become unwieldy—some, like the Great Harris Papyrus composed by King Ramses IV, were as long as 135 feet. A simple solution was to fold the scroll into a concertina (Figure 1.8d). Folded scrolls are more compact, making them easier to handle while reading and to store afterwards. Information can be readily accessed: when seeking particular passages, readers need not unroll and reroll the document. An unfolded concertina book is essentially a scroll, providing backward compatibility.

The book form or "codex" superseded scrolls and concertinas (Figure 1.8e). Around 200 B.C., Greeks and Romans began to write on wax tablets backed with wood, sometimes connected with cords (resembling a ring binder). Later the Romans substituted the more durable vellum for the wood panels.

Because of its numerous virtues, the codex quickly became the preferred format for all literary works. Writers can use both sides; readers can access pages randomly. Codices are easy to read, store, carry, and search. The content is well protected, and thicker pigments can be used for decoration and illustration. The only real disadvantage is that readers can't view more than two pages at once. Books quickly became the standard document format, although concertinas are still in use today for brochures and maps.

Surprisingly, the evolution of computer output paralleled the book format's development. Early printers used paper rolls (scrolls); later, line printers used fanfold paper (concertinas). The latter innovation was spurred by accelerated print mechanisms, but old-timers still recall the advantages of fanfolds over unruly rolls of paper. Indeed, paper rolls were eventually perforated so that they could be folded, boxed, and perused more easily. Today we print on pages and bind them into books. The parallel isn't confined to print technology. Early computer display monitors scrolled, as do Web pages, which originated in the early 1990s. Adobe Reader and Microsoft Word Print Preview provide a paginated concertina-like view: readers scroll through the document page by page, using a scrollbar or Page Up and Page Down keys.

Bibliophiles love books as much for the statements they make as objects as for the statements they contain as text. Indeed, early books were works of art. The steles in Xi'an are a monumental example, studied as much for their calligraphic beauty as for the philosophy, poetry, and history they record, a priceless, permanent record of earlier civilizations. The *Book of Kells* in Ireland, produced by Irish monks at the scriptorium of Iona about 1,200 years ago, is one of the masterpieces of Western art. Figure 1.9 shows part of a page and illustrates the extraordinary array of pictures, interlaced shapes, and ornamental details. Giraldus Cambrensis, a 13th-century scholar, fancifully wrote that "you might believe it was the work of an angel rather than a human being."

Beautiful books have always been highly prized for their splendid illustrations, for colored impressions, for heavily decorated illuminated letters, for being printed on uncommon paper or uncommon materials, for their unusual bindings, and for their rarity and historic significance. In the castle library of Königsburg there are 20 books bound in silver, richly adorned with large and beautifully engraved gold plates. Whimsical bindings abound: a London bookseller had Fox's *History of King James II* bound in fox skin.

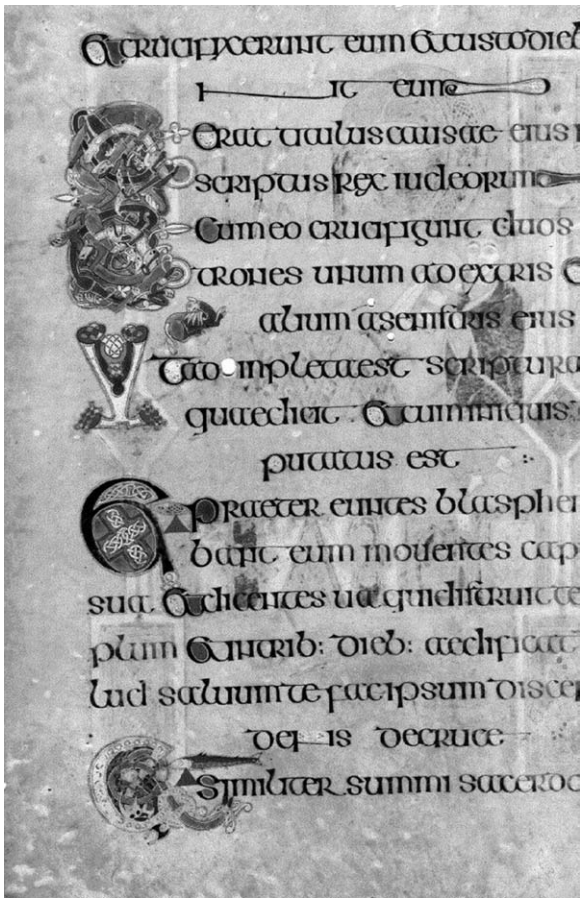


Figure 1.9: Part of a page from the *Book of Kells*

Catalogs and bibliographies are metadata: virtual information about books. In the kind of virtual library sketched in the early technodreams above, the very concept of the book as an individual physical entity seems to be at risk. However, technology has advanced to the point where it need not be: surrogates can substitute for physical works. A picture of the cover may be displayed as a "tangible"—or at least memorable—emblem of the physical book itself. Users can browse the collection using graphical techniques of virtual reality. Maybe they will even be able to caress the virtual cover, smell the virtual pages. But although it is unlikely that readers will love simulated books the way that bibliophiles love real ones, what really matters in libraries is information, even knowledge. Ask the Kataayi community.

1.3 Searching for Sophocles

Pergamos, a digital library at the University of Athens, is representative of many interesting resources that have appeared on the Web over the last decade. It contains four collections spanning

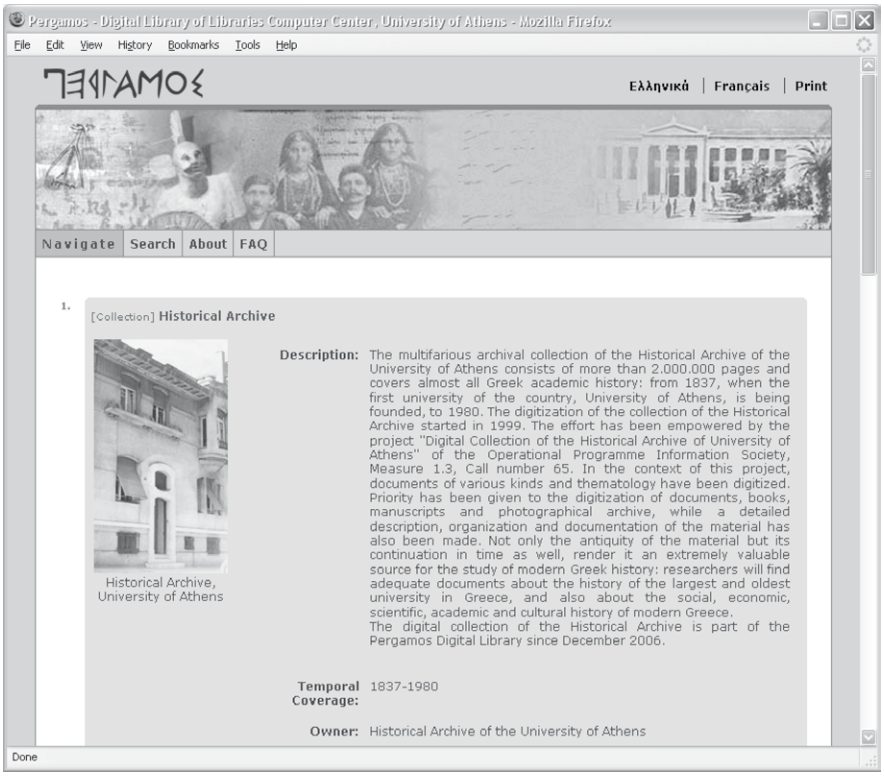
historical archives, folklore, the theater, and sheet music; and more collections are being developed. The name is an alternate spelling for Pergamum, which, as already mentioned, was the site of a significant library in its day. Here we use Pergamos to exemplify features of a real-world digital library.

The system provides an interface in Greek, English, and French. Figure 1.10a shows the home page (in English), which lists the collections available. For each collection a fitting image is shown, along with a summary of what is provided: a description of the resource, the time span it covers, who owns the material, their contact information, and access rights for the user. Some collections, such as the Historical Archive collection, include subcollections: printed documents and photography from the university archives, the Senate Secretariat, and five of the university's faculties. These can be accessed individually or collectively.

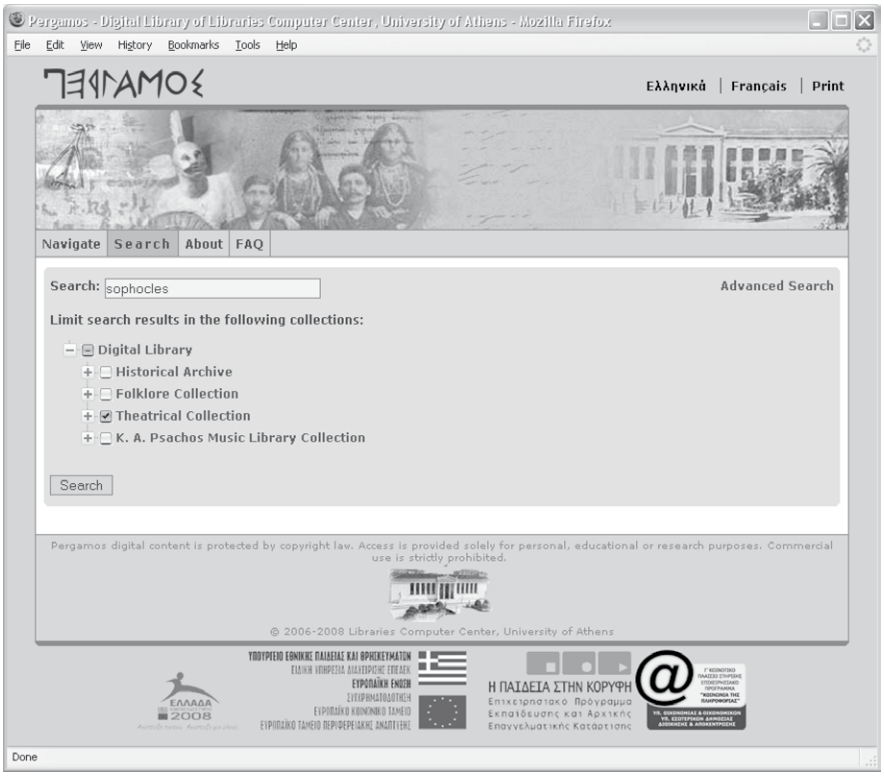
Suppose a user is interested in finding out which of the seven surviving plays by Sophocles have been performed at the university over the years. He brings up the library's search page shown in Figure 1.10b and selects the Theatrical collection from a nested list of possibilities. Next he enters *sophocles* in the search box. The result of the query, shown in Figure 1.10c, is disappointing: only one matching document is found. This is an unexpectedly sparse result for a writer of such prominence, and our user mentally runs a few checks. Was there a typing mistake in the query? Could the uncapitalized initial letter have prevented matching? In both cases the answer is No: the name is correctly spelled, and in this digital library search is case-insensitive. Why then was only one match returned?

It is hard not to notice the size of a physical library. But establishing equivalent cues in the digital realm is tricky. Perhaps, wonders our user, the Theatrical collection is rather small after all? To get a handle on its scope and coverage, he clicks the Navigate button in the interface's main navigation bar, which appears on each page as a row of words beneath the masthead. A few clicks bring up Figure 1.10d, which lists all items in the collection alphabetically by title. There are 633 items in all, and items 1–20 are presented in summary form: title of the show, thumbnail of the program, table displaying author, director, group, theater, and period. Not all items in the list have all these components. The page also contains navigation links: next page and direct entry to a particular page. The sort order can be changed from ascending to descending, and the sort field can also be chosen.

Clicking through a page or two, our user quickly realizes that author names, titles, and so forth are predominately entered in the Greek alphabet. When the user repeats the search in Greek, *Σοφοκλῆς*, twelve matches are returned (Figure 1.10e), which accords better with the user's expectations. He scrolls through the returned matches (Figure 1.10f) and selects the play Αντιγόνη (Antigone). Figure 1.10g shows the digitized version of the actual program produced for this show. The thumbnail shown on the left is quick to load, and the main area of the page is taken up with a larger version. A zoom control allows the user to alter the display resolution in order to see different levels of detail. There is a link to the second page (and, where relevant, subsequent pages) of the program. Digital libraries like this can be accessed by anyone, from any corner of the world. This makes it impossible to predict the uses that people will want to make of the material. For instance, research students might study patterns in scribbled annotations left on the paper documents, which would normally be viewed as ephemera, in order to help investigate a hypothesis they have developed. Most visitors are



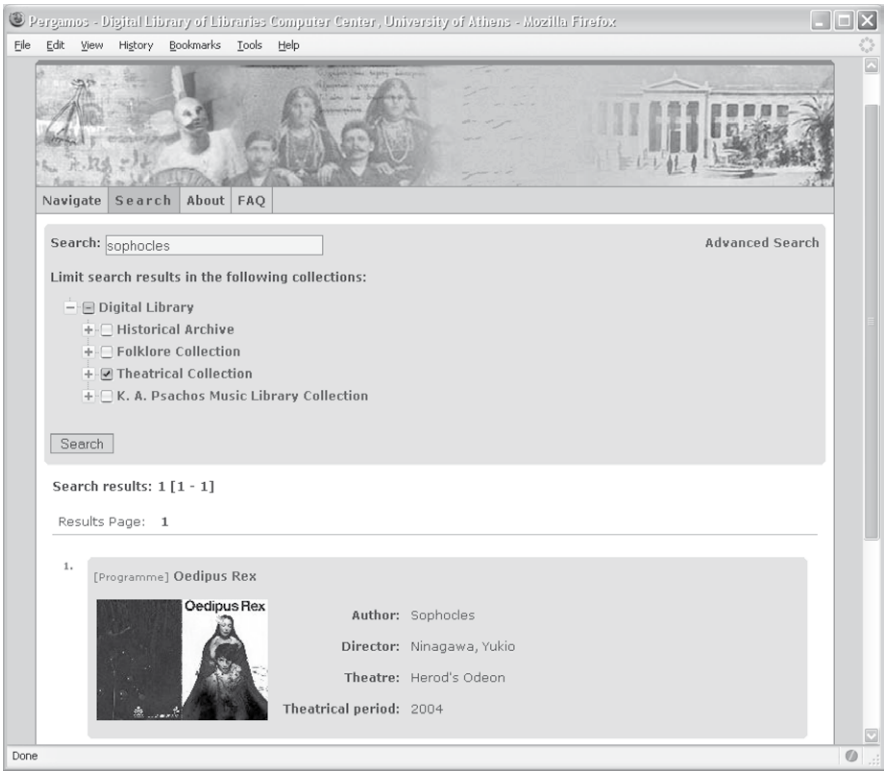
(a)



(b)

Figure 1.10: The Pergamos Digital Library: (a) home page (in English); (b) searching the Theatrical Collection for *sophocles*;

(c)



(d)

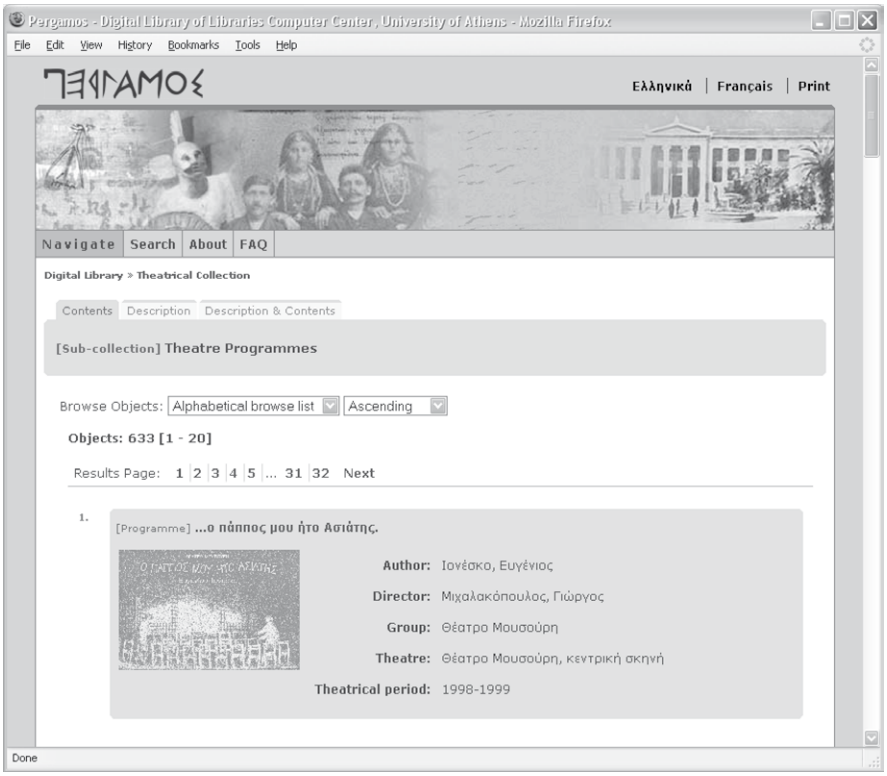
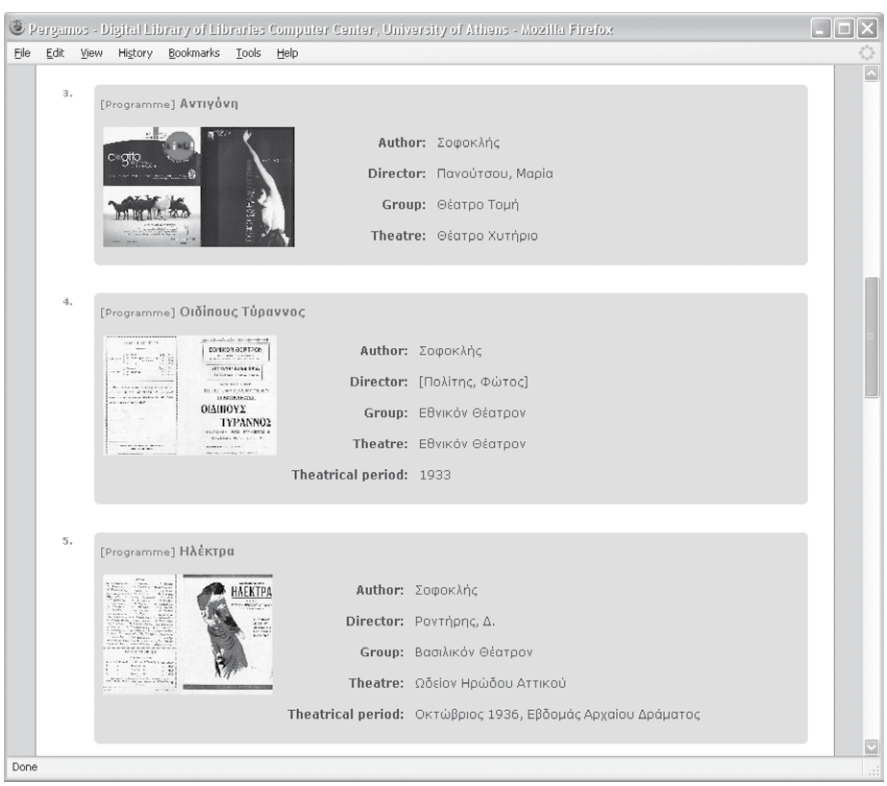


Figure 1.10, cont'd: (c) search result; (d) browsing titles alphabetically;

(e)



(f)

Figure 1.10, cont'd: (e) searching for *Σοφοκλῆς*; (f) search result;

(g)

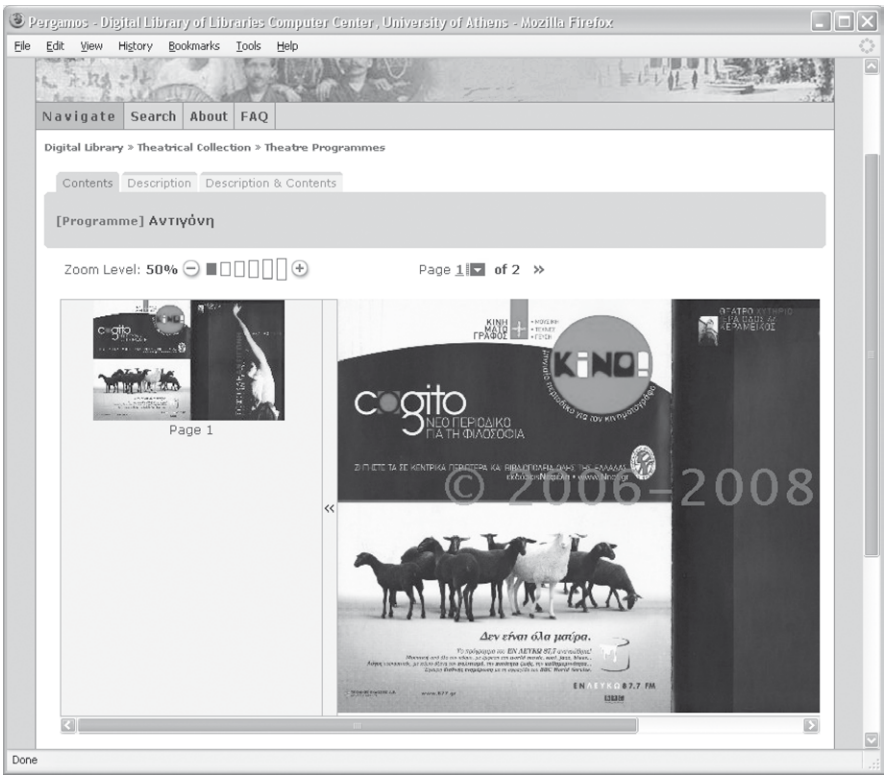


Figure 1.10, cont'd: (g) viewing the item *Αντιγόνη* (*Antigone*)

more interested in theatrical aspects. They might feel that the highest levels of magnification are overkill, but to researchers seeking faded pencil annotations they are a godsend.

Countless digital libraries on the Web adopt the basic form presented in this walk-through. They offer searching and browsing operations to locate items of interest, and ultimately show digital representations of each document, along with ancillary information (metadata). In this example the documents are digitized images (theater programs), to which access is provided through fields of descriptive text, such as title, author, and date. In the case of text-based documents like word-processed reports or Web pages, the full text can be searched as well.

1.4 Digital Libraries in Developing Countries

It sometimes happens that technological advances in developing countries leapfrog those in developed ones. This occurs because established infrastructure, a strong and necessarily conservative force, is absent. For example, in developing countries alternative energy sources, such as solar energy, are widely used in place of traditional power generation and distribution, and these countries have experienced far higher levels of mobile phone growth than developed countries have. Digital

libraries provide another example, compensating for the failure of traditional distribution mechanisms to address local requirements and to deliver information where and when it is needed.

Many current technology trends are not benefiting developing countries—indeed, some bring serious negative consequences. Just as industrialization and globalization have increased the gulf between haves and have-nots, so information and communications technology is creating a chasm between "knows" and "know-nots." By and large, developing countries are not participating in the information revolution, although knowledge is critical for development. The knowledge gap between rich and poor is widening.

In the developing world, digital libraries provide perhaps the first really compelling *raison d'être* for computing technology. Priorities in these countries include health, food, hygiene, sanitation, and safe drinking water. Though computers are not a priority, simple, reliable access to targeted information meeting these basic needs certainly is. Digital libraries give system developers a golden opportunity to help reverse the negative impact of information technology on developing countries.

Disseminating humanitarian information

Traditional publishing and distribution mechanisms have tragically failed the developing world. Take medicine, a field of great importance in this context. Whereas a U.S. medical library subscribes to about 5,000 journals, the Nairobi University Medical School Library, long regarded as a flagship center in East Africa, received just 20 journals in 1998 (compared with 300 a decade before). In Brazzaville, Congo, the university has only 40 medical books and a dozen journals, all published before 1993, and the library in a large district hospital consists of a single bookshelf filled mostly with novels.

By decoupling production and distribution costs from intellectual property charges, digital libraries offer a lifeline. A wealth of essential humanitarian material is produced by international organizations such as the United Nations, as well as national units like the U.S. Peace Corps. Being produced by internationally oriented, nonprofit organizations, funded by all people on the planet, this information is—at least in principle—in the public domain: it could be made freely available in the form of networked digital libraries. While those 5,000 medical journals cannot be distributed for free because they are produced by commercial publishers, this problem does not arise in many areas of physics, as we have seen. The world is changing, and the rate of change will accelerate.

Disaster relief

Natural disasters like earthquakes and hurricanes and human-made ones like terrorist attacks and nuclear accidents demand an immediate and informed response. Disaster relief situations are complex and are addressed by a broad range of players in a variety of organizations acting in parallel. They present an overwhelming need for information: information that is tailored for the problem at hand, organized so that it can be accessed effectively, and distributed even in the absence of an effective network infrastructure. The response to a crisis is characterized by the generation of large amounts of

unstructured multimedia data that must be acquired, processed, organized, and disseminated sufficiently rapidly to be of use to crisis responders.

Digital library technology allows the rapid creation of organized collections of information, graced with comprehensive searching and browsing capabilities. Intelligence specific to the nature of a disaster, the geographical region, and the logistic resources available for the relief effort can all be gathered into a built-to-order digital library collection that combines targeted knowledge with general information about medicine and sanitation.

Preserving indigenous culture

Libraries and their close relatives, museums, have always been involved in preserving culture. These institutions collect literature and artifacts and use them to disseminate knowledge and understanding of different times and cultures. Digital libraries, however, open up the possibility of flexible and coherent multimedia collections that are both fully searchable and browsable in multiple dimensions—and permit more active participation by indigenous people in preserving and disseminating their own culture, as is illustrated by the example of the Zia Pueblo. The principal participants here are by definition the indigenous people themselves: the technological world assumes the role of catalyst, midwife, and consumer—once indigenous culture has been recorded, it will find a fascinated, sympathetic, and perhaps influential audience elsewhere.

Information about indigenous culture takes many guises: oral history, in the form of narration and interviews; artifacts, in the form of images and descriptions; songs, in the form of audio recordings, music transcriptions, and lyrics; and dances and ceremonies in the form of video, audio, written synopses, and interpretations. Multimedia digital libraries allow such information to be integrated, recorded, browsed, and searched, all within a uniform user interface.

Because language is the vehicle of thought, communication, and cultural identity, a crucial feature of digital libraries for culture preservation is the ability to work in local languages. This strengthens individual cultures, promotes diversity, and reduces the dominance of English in the global information infrastructure.

Locally produced information

In digital library applications for culture preservation, the relevant information is, of necessity, readily available locally. But there are countless other scenarios that involve creating and distributing locally produced information collections. At first glance one might think that the Internet includes such a wealth of content that surely there must be something of benefit to everyone. However, this ignores not only the problem of language—most information is available only in major languages like English—but also the importance of the kind of content that only local communities can generate.

Teachers prepare educational material that addresses specific community problems, and they adapt published material to employ local examples. Indigenous people have medicinal knowledge based on

local plants or long-acquired knowledge of the cultivation and protection of local species. Such knowledge is vital: more than half of the world's most frequently prescribed drugs are derived from plants or synthetic copies of plant chemicals, and this trend is growing.

Local groups assemble information collections that describe and reflect neighborhood conditions, providing new material for sociocultural studies, fostering cultural exchange while retaining diversity, and increasing international understanding. Web sites for community and social development might include information on health problems endemic to a particular African community, or information on commodity prices for a particular good traded in Brazilian markets, or examples of curricular projects suitable for use in Indian schools.

The development of content that addresses the specific needs of a particular community stimulates the demand for information technology among that community. Getting learners to produce their own content is one of the best ways to exploit information technology in learning situations. It not only improves the learning experience, but also creates material that benefits the community. Teachers and students can work together to create their own content that has value for the community, and for the nation as well.

Effective human development blossoms from empowerment. As the Chinese proverb says, "Give a man a fish and he will eat for a day; teach him to fish and he will eat for the rest of his days." Dissemination of information that originates in the developed world is useful to developing countries, as Kataayi members will testify, but a more effective strategy for sustained long-term human development is to foster the capability for creating information collections, rather than the collections themselves. This allows developing countries to participate actively in the information society, rather than observing it from outside. It stimulates the creation of new industry, and it helps ensure that intellectual property remains where it belongs, in the hands of those who produce it.

The technological infrastructure

Computers are not so hard to come by in developing countries as one might think. Their extraordinarily rapid rate of obsolescence, coupled with the developed world's voracious appetite for the latest and greatest, makes low-end machines essentially free: instead of clogging landfill sites many (although certainly not enough) find their way to developing countries. A 1998 World Bank survey of developing countries found 3 to 30 PCs per 1,000 people, depending on the poverty level. Using an estimated growth rate of 20 percent per year, we conclude that at the turn of the millennium there were 50 million PCs in developing countries, serving a population of four billion.

A more serious obstacle is that network access varies widely around the globe. Whereas in 1998 more than a quarter of the U.S. population was surfing the Internet, the figure for Latin America and the Caribbean was 0.8 percent, for Sub-Saharan Africa 0.1 percent, and for South Asia 0.04 percent. Schools and hospitals in developing countries are poorly connected. Even in relatively well-off South Africa, many hospitals and 75 percent of schools have no telephone line. In African universities, up to 1,000 people can depend on just one workstation. The Internet is failing the developing world. While global satellite communication networks may eventually bring relief, this takes time and money.