

Visions of the digital library

Ian H. Witten

New Zealand Digital Library Project
Department of Computer Science
University of Waikato
New Zealand
ihw@cs.waikato.ac.nz

S.R. Ranganathan (1892–1972), an influential librarian and educator who, I am told, is considered the father of library science in India, wrote as one of his “five laws of library science” that a library is a growing organism (Ranganathan, 1931). Not only do individual libraries grow, the very concept of what it means to be a library has developed and evolved over the centuries. Before being willfully and tragically laid waste, the fabled collection of Alexandria grew to a size that would not be surpassed for two millennia. The chained books of medieval and monastic libraries gave way to classic private collections such as the great library of Duke August in Wolfenbüttel, Germany, the largest library in Europe during the 17th Century—and acclaimed as the eighth wonder of the world. During the following century national libraries gained preeminence in capital cities in Europe and, later, the new world. The public library movement took hold in the 19th Century, while progressive 20th Century librarians invented self-service and adopted the idea of open-access libraries (particularly in English-speaking countries), marking the fulfillment of the principle of free access to the contents of libraries by all—the symbolic snapping of the links of the chained book.

Today, we live in exciting times. Digital libraries, whose history spans a mere dozen years, will surely figure amongst the most important and influential institutions of this new century. The information revolution not only supplies the technological horsepower that drives digital libraries, but fuels an unprecedented demand for storing, organizing, and accessing information. If information is the currency of the knowledge economy, digital libraries will be the banks where it is invested.

Digital libraries have the potential to be far more flexible than conventional ones. Of course, they are portable: they will be with you whenever you want them to be: in the home, in the plane, at the beach, in an Indian village, on the street when you want to play your friends that new song. They will be large, giving access to your personal book collection, your town’s public library, your university library. Not only this, but they will ultimately be seamlessly integrated with national and international sources of information—inter-library loan at your fingertips.

But wait, there’s more. Flexibility will extend well beyond matters of physical convenience. Future digital libraries will surround you with information in ways that we can yet only dimly perceive. When Karl Marx wrote *Das Kapital*, he worked in the reading room of the British Museum library. Not only will future revolutionaries



Figure 1 New York Public Library Reading Room



Figure 2 Digital library in the British National Library

use their laptop instead of Marx's pen and paper, they will work "inside" their digital libraries in a stronger and more visceral sense.

This paper will look at visions of the digital library—visions from the past, present, and future. An important component of our vision today is for digital libraries in developing countries, for we believe that as far as the developing world is concerned digital libraries may be killer apps for computer technology. As for tomorrow's vision, we develop the theme of digital libraries that surround the user with information, information that is pertinent to what they are doing, and changes as they work to track the context of the tasks that are being undertaken.

1. Yesterday's visions

Over sixty years ago, science fiction writer H.G. Wells (1938) had a vision of a "world brain" based on a permanent world encyclopedia which "would be the mental background of every intelligent [person] in the world," and went on to echo Ranganathan's view a few years earlier: "It would be alive and growing and changing continually under revision, extension and replacement from the original thinkers in the world everywhere."

Eight years later, Vannevar Bush (1947), 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."

Fifteen years further on, J.C.R. Licklider (1960), 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 imagined this to be 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."

Towards 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.

2. Today's visions

Figure 1 shows the reading room of the New York Public Library. This is a magnificent place: spacious and airy with bright yet soft lighting, comfortable furniture, warm tones of wood, leather and books; the hushed sounds of people thinking and working. You can become absorbed in what you are reading, living in your own world shared between your head and the book you are holding, or your gaze can wander around the immense spaces above you, the massive windows, the huge chandeliers, the painted ceiling. But yet the space is impersonal, like any large physical library has to be. One size fits all. No matter how divergent your intellectual pursuits, you work in the same place as your neighbor. The most you might expect by way of personalization is an assigned carrel where you can leave your books and materials undisturbed until tomorrow. Is this the vision we seek for digital libraries?

Figure 2 shows an example digital library, one that forms part of the British National Library. It is just as impersonal as Figure 1, if not more so. You stand at the threshold and are confronted with row upon row of identical workstations. This space lacks the architectural beauty of the reading room in Figure 1—it looks utilitarian rather than spacious. We all know from bitter experience that the technology it contains, no matter how spiffy today, will appear old and jaded in just a few months. But perhaps the most striking thing about the vision is the staged nature of this publicity photograph: one man in an empty room, clearly posed as though at work. Every other workstation in the room shows the same screen—not only that, but beside each one, on a typing stand, is the same piece of paper. This Orwellian setting does not seem like a nice place to work. Is it the vision we seek?

3. Digital libraries in developing countries

A more engaging picture of a present-day digital library is Figure 3, which shows the computer room at the Kataayi cooperative in Uganda. From a western perspective the image is decidedly low-tech. The computers are ancient, the furniture is shabby, the walls are bare, the setting is plain and utilitarian. But it looks like it works. Here we see real people interacting with information in a real environment, rather than a publicity machine's conception of some kind of spine-chilling ideal. This picture serves as a salutary reminder that libraries are about connecting people with the information they need. Kataayi's library may not be flashy, but it works.



Figure 3 The digital library at the Kataayi cooperative in Uganda

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. Alternative sources such as solar energy are widely used in place of traditional power generation and distribution, while many developing countries have experienced far higher levels of mobile phone growth than developed ones. Digital libraries provide another example, compensating for the failure of traditional distribution mechanisms to address local requirements and deliver information where it is needed.

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 (Witten *et al.*, 2001a).

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 5000 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) (UN, 1999). In Brazzaville, Congo, the university has only 40 medical books and a dozen journals, all from before 1993, and the library in a large district hospital consists of a single bookshelf filled mostly with novels.

Digital libraries, by decoupling production and distribution costs from intellectual property charges, offer a desperately needed lifeline. A wealth of essential humanitarian material is produced by various international organizations, such as the United Nations, as well as national units like the U.S. Peace Corps. Being produced by internationally-oriented, non-profit 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 5000 medical journals cannot be distributed for free because copyright on the articles is held by commercial publishers, this problem does not arise in many areas of physics, for physicists have for the past decade been using automated archives to disseminate the results of their research (Ginsparg, 1996). The world is changing, and the rate of change will accelerate.

Disaster relief

Natural disasters such as earthquakes or hurricanes, and man-made ones such as terrorist attacks or nuclear accidents, demand 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 organized collections of information, graced with comprehensive searching and browsing capabilities, to be created rapidly (Witten *et al.*, 2001b). 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 medical and sanitary information.

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. The principal participants here are by definition the indigenous people themselves: the technological world assumes the role of catalyst, midwife, and consumer; for once 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; 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 there is such a wealth of content on the Internet 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 English or other major languages—but also the fact that there are many local community content issues that contribute towards effective information use.

Teachers prepare educational material that addresses specific community problems, and adapt published material to employ local examples. Indigenous people have invaluable 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 socio-cultural 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 of 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 amongst that community. Getting learners to produce their own content is one of the best ways to exploit information technology in learning situations. Not only does it improve the learning experience, it also creates material that benefits the community. Teachers and students can together create their own content that has value for the community, and for the nation too.

Effective human development blossoms from empowerment rather than gifting. 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." Disseminating information originating in the developed world is a useful activity for developing countries. But a more effective strategy for sustained long-term human development is to disseminate the capability of creating information collections, rather than the collections themselves. This will allow developing countries to participate actively in our information society, rather than observing it from outside. It will stimulate the creation of new industry. And it will help 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 (though certainly not enough) find their way to developing countries. A 1998 World Bank survey of developing countries found 3 to 30 PCs per 1000 people, depending on the poverty level (World Bank, 2000). With growth predicted at 20% per year, we estimate that at the turn of the millennium there were fifty 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 US population were surfing the Internet, the figure for Latin America and the Caribbean was 0.8%, for Sub-Saharan Africa 0.1%, and for South Asia 0.04% (UN, 1999). Schools and hospitals in developing countries are poorly connected. Even in relatively well-off South Africa, many hospitals and 75% of schools have no telephone line. Universities are better equipped, but even there up to 1000 people can depend on just one terminal. The Internet, as Arunachalam (1998) puts it, "is failing the developing world." While global satellite communication networks may eventually bring relief, this takes time and money.

Because of the difficulty of network access, the structure and organization of digital libraries should be separated from their distribution media. Physical distribution of information on recordable devices can provide an attractive alternative to networks. Compact disk read-only memory, CD-ROM, is a practical format for areas with little Internet access. Their 650 Mb capacity can hold a useful volume of information, such as the 1200 fully-illustrated and fully-indexed books in the Humanity Development Library. Most of the space in a collection such as this is consumed by pictures: several times as many books could be included if they were not so lavishly illustrated. CDs are giving way to digital versatile disk, DVD, which can hold from 5 to 20 Gb of data. A year's supply of those 5000 medical journals mentioned above could fit, fully indexed, on a single DVD. And save lives.

4. Tomorrow's visions

So much for the present: what about the future? By way of—literally!—comic relief, Figure 4 shows a sci-fi artist's image of a digital library taken from a Marvel comic, the living computers of Xandar (Marvel Comics, 1979). On this planet they have stored the still-functioning brains of the population for more than ten thousand millennia, giving—as our guide explains—a complete record of all their history, all their science, all their *knowledge*. Is this the long-term prognosis that we sought at the outset? (Note that our 25 centuries of libraries represent a mere eye-blink in Xandarian history.)

Here is a vision that emphasizes preservation over access. Xandarians can put the brains in, but can they get the knowledge out? It certainly appears from the illustration that their computer scientists have some work to do on the user interface. But the problem runs deeper. In storing living brains, what is lacking (or may be lacking—it's

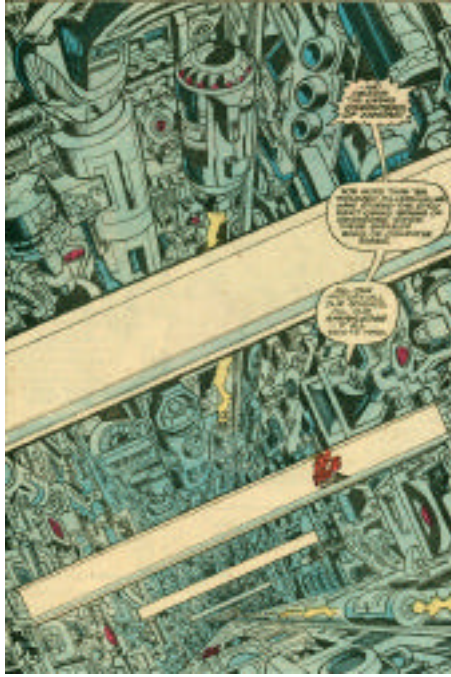


Figure 4 Xandar's digital library



Figure 5 Carpenter's workshop

hard to tell from the picture!) is *librarianship*: the selection, organization, and maintenance of information, the wisdom that librarians put into the library by making value judgments about what information is to be included and how it should be organized. Figure 4 is reminiscent of an historical archive of the Web (Kahle, 1997). There is no organization, no quality control: a mere repository. This is not a library.

Our own vision of the digital library of the future is that it will be a personalized, comfortable, space to work in. Think of it as more like a kitchen than a library—a kitchen for knowledge preparation. If you like cooking, you will have arranged your kitchen to suit what you do: utensils ready to hand, pots by the stove, spices for the kind of dishes you like to cook, and placed just where you need them. Left-handed?—change your kitchen around. Short?—adjust the work surfaces. Like wine?—keep a glass handy. Stir-frys?—here's the wok.

Figure 5 shows another workspace that illustrates the vision we are trying to convey. This carpenter's workshop—though perhaps a little stiff and idealized—is all set out as a comfortable, productive, workplace. Hand tools are arranged in convenient clusters, bench tools such as vise and drill press are conveniently located, nails and screws are boxed and labeled, a well-positioned lamp illuminates the current job. This man knows where everything is, and the physical arrangement almost exudes productivity. Imagine how much better you could work here than in the messy corner, littered with assorted junk and half-finished projects, that most of us have in our garage. Imagine how this carpenter would feel if he had to carry all his tools into the

New York Public Library Reading Room in Figure 1; unpack them, sort them, and arrange them before he began work; and pack them up again before lunch to leave the space clear for someone else.

Just so with the digital library. Inside the computer it will be *your* library, arranged the way you like it, personalized for the kind of things you do. Externally it may look like Figure 1 (just bring your laptop and plug it in), Figure 2 (hopefully a little more welcoming and ergonomic, with concealed computers, flat-panel screens, and wrist-rests), or even Figure 3 (though with more powerful equipment). Inside it will not only give access to the world's recorded knowledge as Wells's vision did, but it will feel like Figure 5, arranged just for you and the kind of things you do.

You will need to invest in this personalization, just as our carpenter invested time and money establishing his workplace. And there are pitfalls aplenty, for with the potential for flexibility comes the potential for confusion. A physical environment makes the possibilities that it opens up for interaction—its “affordances”—openly manifest. There is no user manual for Figure 5: you can step into this environment and see immediately how it works, use it yourself. It will be harder to make the intellectual environment your digital library provides so accessible and transparent.

But we're dreaming of the future: these problems will be solved. The computers, so prominent in Figure 2, will disappear. There will still be an interface—perhaps a screen, though it may disappear into the wall or into your spectacles; perhaps a keyboard, though it may disappear into finger sensors or a microphone; perhaps a mouse, though it may be a wand or a wave of the hand. But you will conceptualize this interface as a *library* rather than a *computer*: the computer will become invisible just as the countless electric motors in your house have become invisible, disappearing into hairdryers, fans, electric razors, kitchen appliances, CD players, VCRs. Tomorrow's digital library will feel less like a computer, more like a kitchen or workshop.

As well as being personalized, your digital library will be dynamic. And not just in the sense that the information it contains will be bang up to date. The library will work alongside you, tracking your activity, unobtrusively rearranging itself to put what you might need in the context of what you are doing just there where you can see and read it. When you leave off for the day it will continue to work for you, locating pertinent information, classifying and categorizing it, working through the implications, researching on your behalf, so that tomorrow when you recommence you will start well ahead of where you were when you left off today.

5. Working inside the digital library

Digital libraries are libraries without walls, though they do need boundaries—we argue that the very notion of a collection implies a boundary. Paradoxically, perhaps, in the future we will work *inside* the digital library in a new sense of “in”-ness that we can barely glimpse today. The library will be an environment that surrounds you in an intellectual, not a physical, sense. But virtual reality means that intellectual experiences can easily be translated into physical ones. More or less immersive (you can choose), the library will be an environment that reacts and responds to what you

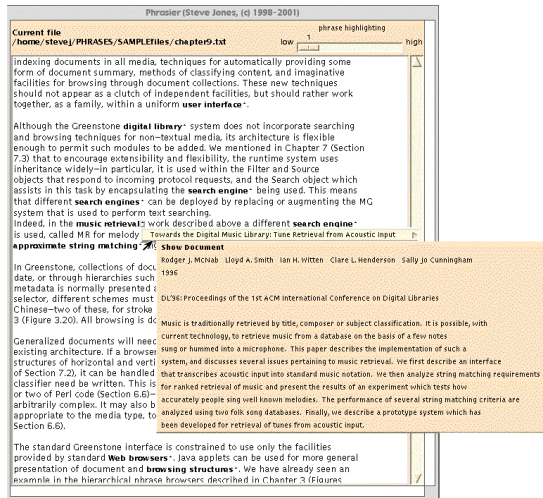


Figure 6 Reading a document in a digital library

are doing, making the right kind of information available to you as and when you need it, and in an appropriate form. It will surround your head.

Lest you feel you are being carried away by empty rhetoric, a system called Phrasier, conceived and constructed by Steve Jones of Waikato University, gives a glimpse of what we mean (Jones, 1999; Jones and Paynter, 1999; Jones and Staveley, 1999). Phrasier is an environment for reading and writing within a digital library. Figure 6 shows us at work writing this article. We are working in a digital library, and associated with every document in it are a handful of keyphrases, perhaps assigned manually, perhaps extracted automatically from the text. It is these that are used to connect the article being written with the documents in the library.

As Figure 6 shows, certain parts of the article's text are highlighted in bold face. These are phrases that appear as keyphrases of other documents in the library. If a subject-matter thesaurus were available, phrases in it would be highlighted too, wherever they appeared in the article. The user can control the amount of highlighting, and the tone of the non-highlighted text, using the slider at the top of the page. Studies are underway to determine whether people can skim text faster, yet still gain some comprehension of it, if the text fades away into the background and only the keyphrases are clearly visible.

Mouse buttons can be used to focus on a phrase of interest that appears in the article and examine the documents in the library for which it is a keyphrase. A pop-up window contains titles of documents for which this is a keyphrase. In this case the keyphrase is *music retrieval* and the list contains just one item. That document has been selected and is shown just beneath the item, in a second pop-up window.

In Figure 7, we have brought up another window (in the background) in which to examine related literature. We have focused on a particular area of the article—the



Figure 7 Focusing on part of the document and finding pertinent literature



Figure 8 Focusing on part of the document's subject matter

three paragraphs in the middle—by highlighting it with the mouse (the highlighting is only faintly visible in the picture). In the background window there appears, on the left, a list of the keyphrases that appear in that region (there are three), along with their frequency and the number of documents for which they are a keyphrase. The system has used this set of keyphrases as a query into the digital library, and retrieved a list of documents that relate to all them, sorted into relevance order using the same kind of ranking heuristic that is used during full-text search. If the selected part of the article is the focus of interest, this list shows the relevant literature in the library. It appears in the right-hand panel: clicking on a document brings it up in a separate window.

As authors, we often want to focus on a conceptual subtopic of the article's subject matter rather than a spatial region like a particular paragraph or section. In Figure 8 the phrases in the background window's list are keyphrases in the digital library that are mentioned anywhere in the article's text, not just in a particular sub-area as before. Some are highlighted because we have selected them manually by clicking on them. This selection effectively defines a subtopic, or group of subtopics, that has been chosen as being of special interest. As before, a list of related documents appears in the right-hand window; but now it is the manually chosen keyphrases that are used to select and rank them. This provides a reading list, ranked by relevance, for the conceptual subtopic that has been defined. Again, of course, the full text of each document is just a click away.

We have chosen to illustrate the Phrasier interface by describing how we might have written this article. As we type new material into the document window in Figures 6, 7 and 8, everything described above happens interactively. The new article is not only born digital, it is born in a library, fully contextualized and linked to the existing literature at birth.

Exactly the same system supports reading. You can load an article into the document window and read it "in" the library. The article could be taken from the library, or downloaded from elsewhere. The entire contents of the library are available with no

effort, in context, at your fingertips, on the fly, as you read and write. This is a library that works along with you.

6. Conclusions

We began this essay by quoting one of Ranganathan's five laws of library science. It is fitting to conclude with the complete set:

1. Books are for use
2. Every reader his book
3. Every book his reader
4. Save the reader's time
5. The library is a growing organism.

We like to think that he would have shared our visions (which a forthcoming book—Witten and Bainbridge, in press—develops in far more technical detail). It is because books are for use that we prefer the active working environment depicted in Figure 3 to the sterile publicity image of Figure 2. The aim of the reading and writing environment in Figures 5–7 is to match the reader with the documents he needs, and vice versa—except that digital libraries can provide a dynamic, context-dependent match that no physical arrangement of material could ever rival. Saving the reader's time by providing convenient access mechanisms is a principal concern of all digital libraries—except that we would prefer to couch this in terms of providing the best intellectual return for whatever time the reader chooses to invest.

Finally, a library is indeed a growing organism. Individual digital libraries grow, and the field itself grows and develops. We look forward with eager anticipation to whatever visions the future brings.

Acknowledgements

My visions of digital libraries have evolved out of the stimulating research environment of the New Zealand Digital Library project, and I would like to acknowledge the profound influence of all project members—and particularly Steve Jones, who conceived of the Phrasier system and supplied Figures 5–7. I wrote this article while visiting Google (google.com), whose support is greatly appreciated.

References

- Arunachalam, S. (1998) "How the Internet is failing the developing world." Presented at *Science Communication in the Next Millennium*, Egypt; June.
- Bush, V. (1947) "As we may think." *The Atlantic Monthly*, 176(1): 101–108.
- Ginsparg, P. (1996) "Winners and losers in the global research village." Presented at UNESCO Conference on *Electronic Publishing in Science*, Paris; February.
- Jones, S. (1999) "Phrasier: an interactive system for linking and browsing within document collections using keyphrases". *Proc IFIP Conference On Human-Computer Interaction*, Edinburgh, Scotland, pp. 483–490; August–September.

- Jones, S. and Staveley, M. (1999) "Phrasier: a system for interactive document retrieval using keyphrases". *Proc ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, California, pp. 160–167; August.
- Jones, S. and Paynter, G. (1999) "Topic-based browsing within a digital library using keyphrases." *Proc Fourth ACM Conference on Digital Libraries*, edited by E.A. Fox and N. Rowe. Berkeley, CA, pp. 114–121; August.
- Kahle, B. (1997) "Preserving the Internet." *Scientific American* 276(3):82–83, March.
- Marvel Comics (1979) *Fantastic Four*. Vol. 1, No. 25.
- Licklider, J.C.R. (1960) "Man-computer symbiosis." *IRE Trans Human Factors in Electronics*, HFE-1: 4–11.
- Ranganathan, S.R. (1931) *The five laws of library science*. Madras Library Association, Madras.
- World Bank (2000) *World Development Indicators 2000*. World Bank.
- UN (1999) *Human Development Report 1999*. UN Development Programme.
- Wells, H.G. (1938) *World Brain*. New York: Doubleday.
- Witten, I.H., Loots, M., Trujillo, M.F. and Bainbridge, D. (2001a) "The promise of digital libraries in developing countries." *Communications of the ACM*, Vol. 55, No. 5, pp. 82–85; May.
- Witten, I.H., Bainbridge, D. and Boddie, S. (2001b) "Power to the people: end-user building of digital library collections." *Proc ACM Digital Libraries*, Roanoke, VA.
- Witten, I.H. and Bainbridge, D. (in press) *How to build a digital library*. Morgan Kaufmann, San Francisco, CA.