Journal of Zhejiang University SCIENCE
ISSN 1009-3095
http://www.zju.edu.cn/jzus
E-mail: jzus@zju.edu.cn

The qualitative advantages of quantities of information: bigger is better

LESK Michael

(*Department of Library and Information Science, Rutgers University, New Brunswick, NJ 08901, USA*)

E-mail: lesk@acm.org; lesk@scils.rutgers.edu

Received Aug. 5, 2005; revision accepted Sept. 10, 2005

Abstract: Digitization projects should focus on quantity rather than quality. Increasing quantities of information produce qualitatively more valuable services. Online writing and searching are now common, and it is only online reading that is still limiting our use of online books. New interfaces might increase our willingness to read online, which should be encouraged rather than fought, since it represents an increase both the amount of information available and the participation of more people in the writing and exchange of information.

Key words: Digital libraries, Search engines, Mass digitization
doi:10.1631/jzus.2005.A1169 **Document code:** A **CLC number:** TP391

SUMMARY

Critics of large digitization projects say that the quality of information available to students is being lost in a rush to increase its quantity. But in reality the larger amounts of material mean that it is more likely students can find appropriate and relevant documents. What is happening is that people react to the large quantities of material on the Web and the power of the search engines by seeking the right places to read, rather than being limited to a few items by scarcity of material.

The next President of the American Library Association, Michael Gorman, wrote in December 2004 (Gorman, 2004) that "massive databases of digitized whole books, especially scholarly books, are expensive exercises in futility" and that "a snippet from Page 142 must be understood in the light of pages 1 through 141 or the text was not worth writing and publishing". He was referring specifically to Google Print, but his comments would apply to Project Gallica (100 000 French works), the Million Book Project, Project Gutenberg, the book conversions of the Internet Archive, projects such as the Making of America, and others. Similarly, in 1996 Ian Irvine, then head of Elsevier, said that what people found on the Internet were the manuscripts his journals rejected (Irvine, 1996).

In reality, however, the enormous size of the Web means that even for scholarly queries, there is likely to be a better answer on the Web than in a professionally edited information service. Looking at a few sample queries confirms that less specialized queries find more appropriate documents on the Web, and more specialized queries are often answered only on the Web. The students who use Google for everything should not be criticized; their behavior is rational and sensible. Instead, we should be trying to help them by increasing still further the materials they can use.

As for the suggestion that books have to be read from cover to cover, this may have made more sense when we had no way to search them. If we took Gorman's logic further, we would say that books should not have indexes or tables of contents. People find searching so convenient that they use it for almost any task, even tasks designed to make searching difficult.

Technology has meant that searching is now easy, while tasks such as classification, reviewing, and editing are still hard. This means that putting primary materials online is easier, while preparing edited scholarly editions is still hard. Libraries throughout the world have found that providing digital access to previously obscure works has greatly increased their use.

Yes, there are still times when users need summaries, evaluations, and context. We are now finding that there is an enormous resource in people all over the world willing to contribute their effort to make Web resources still more useful. For example, Project Gutenberg's "Distributed Proofreaders" group checks thousands of pages a day, entirely with volunteer labor. The various Wiki groups are even more remarkable in their ability to generate evaluated and readable summaries and discussions.

We should not be trying to fight mass digitization, by arguing that people should limit their reading in the way that they had to when books were expensive and scarce. Instead, we should be encouraging mass digitization, confident that people will be able to find what they want, and will be willing to help other people who follow them down the same information paths.

The technologies for digital libraries are now in place, and even some of the economic issues are being solved (Lesk, 2004). The major issues now are legal and public acceptance; and a recognition that adding to the Web will benefit us, and if done correctly will not hurt publishers.

INTRODUCTION

There was a time when we had an information economy based on scarcity; only a few items were available, and finding what you wanted was hard. In those circumstances, people hoarded information, and they paid great attention to maximizing the use of whatever they had. We are now moving to an information economy of abundance, in which we have enormous opportunities to read whatever we want, on any subject. We also have a new, and historically unanticipated, ability to search for information. In an economy of scarcity, there are a few things you know about, and you use them; in an economy of abundance, a great many things are exploited.

The accumulation on the Web is enormous. Google now claims to be searching 8 billion Web pages and 2 billion images and the Web was estimated two years ago at 170 Terabytes. By contrast Lexis-Nexis, collecting many traditional publications, had 4 billion documents in 2002, or about 4 Terabytes of text, and 73 million images. Looking at traditional book lists, Worldcat, the unified file of OCLC (the Online Computer Library Center) has one billion holdings records, collected over 61 million items. That would probably be about 60 Terabytes of text, still less than the Web.

Even the traditional marketplace is being transformed by online services. Historically, it was very difficult to succeed as a small specialist publisher. Without marketing staff and funding, the chance that a publisher who issues a dozen books a year would find one of them on the most visible shelf at Barnes & Noble was vanishingly small. But an online bookstore does not have any limits on how many books it can stock and books found by searching are much more equal than books found by looking to see what is at eye level.

Last year Chris Anderson wrote that half of Amazon's sales were from other than the 130 000 best selling books, implying that they were sales of books that a paper bookstore does not even stock (Anderson, 2004). Anderson's numbers have been challenged and he has retreated to a claim that perhaps one-fifth to one-third of Amazon's sales are from outside the 130 000 best-sellers, but even that is remarkable: Amazon is selling, each year, \$0.5~\$1 billion worth of books that no ordinary store even stocks (based on total sales of \$2.5 billion) (Rosenthal, 2005).

Similarly, online libraries do not have the same limits on how many books they can provide. Major research libraries tend to have about 10 books per square foot or 100 books per square meter of building space. The cost of adding storage space for another book ranges from a few dollars in an offsite warehouse to a few dozen dollars in a central facility. In a central city or central campus library, built with full services and some attention to architecture, building the shelf space to hold another book costs more than the book does. To add even 100 Megabytes for a full set of high-resolution page images of a book costs about 4 cents, as of mid 2005.

That means that for an online library, selectivity is a luxury. It is hard to imagine any manual process for deciding to discard a book that costs less than 4 cents. So why bother? What do we gain by rejecting some items as unworthy or uninteresting?

Yet, commentators still write that a key ingredient in any project is selection and evaluation. People have sneered at the Million Book Project for buying books discarded by libraries, and Raj Reddy has answered by suggesting that discarded books, typically books of which a library bought more than one copy, are thus doubly valuable. But in reality the cost of choosing books is soon going to exceed the cost of scanning them.

In 2001 the US spent $686 million building libraries (Friess, 2002). This far exceeds the cost of scanning one copy of each book they own, since there are only some 30~40 million different books in US libraries. The new San Francisco Library building alone cost $137 million almost ten years ago and it has 154 242 linear feet of shelving to hold some 1.5 million books; that is a price of $100/book held, far above the conversion cost of creating digital images. Today it is the copyright law, rather than either technology or economics, that most constrains the digitization of old books. Even Google has run into copyright issues, and as of the summer of 2005, is still negotiating with publishers about the details of their project.

One of the objections to the Google Print has been a charge of "cultural imperialism." The head of the Bibliotheque de France, Jean-Noel Jeanneney, called Google Print a "confirmation of the risk of crushing American domination in the way future generations conceive the world (Jeanneney, 2005)." European libraries came together to propose a multinational plan countering Google, claiming that a single source of information was a danger for cultural plurality ("une seule source d'informations est un danger non négligeable pour la pluralité culturelle", according to Jean-Frederic Jauslin, until recently the Swiss national librarian (Jauslin, 2005)). Yet surely the provision of additional information does not make European libraries any less accessible or smaller than they are now; the real danger is that those who cannot read the language of a Web page are blocked from learning what it contains. This is actually a greater danger to Americans, who are rarely fluent in multiple languages, than to Europeans.

Fortunately, the European plan, to improve scanning of their own resources, will benefit everyone. Even if it does not, about half the books in major American research libraries (including the ones Google is working with) are written in languages other than English. With luck, and success at handling the rights management issues, worldwide information will become ever more available.

Digital libraries are also expanding into areas of content that were not available or only available with great difficulty in the past. Most recently, the services of Google Maps, Microsoft Virtual Earth, and Amazon "maps.a9.com" have provided visual representations of United States cities never before available. The Google product (until recently known as Keyhole) gives views from above with resolutions of 1/3 meter or even better; Cambridge, Massachusetts is available at a resolution of 10 centimeters. Google has also provided outlines of major buildings so that one can get a 3-D view either in mountains or cities. Amazon, driving a van around a number of cities, will let you see what the streets look like at ground level.

There are even historical aerial photographs available, for example a 1939~1941 view of much of Illinois at UIUC and a 1934 view of Connecticut from the Connecticut State Library. Figs.1 and 2 show the vicinity of the New Haven railway station in 1934 and today respectively, courtesy of the Connecticut State Library and Google Maps. You can see that the steam locomotives, and the roundhouse they used are gone as are many other railway and industrial buildings. There are some similar examples in (Lesk, 1997), but they meant that I had to go to used bookstores and buy

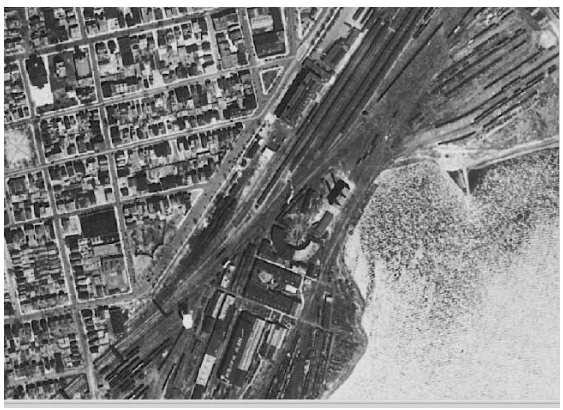

Fig.1 New Haven, Connecticut, 1934

Fig.2 New Haven, Connecticut, 2005

some old maps. Today I merely needed to spend a few minutes on the Web.

Similarly, many primary documents are appearing on the Web. Ten years ago when I wanted an example of 18th century handwriting I bought an old letter in a bookshop. Today anyone can find George Washington's correspondence at the Library of Congress website.

What is remarkable is that all of this material is being used. Every library has stories about material that sat in the basement relatively idle, has been put on the Web, and is getting hundreds of hits and downloads. Digital technology is greatly expanding the range of material that students can use. And this in turn means that student research projects are no longer limited to a few books that professors assign, but include all the resources of great libraries around the world. When I was in college in the 1960s it was still expected that there were "undergraduate" libraries and "research" libraries, and I needed special permission to use the "research" library. Over the next few decades most universities admitted undergraduates to their largest library; now even the special collections, once digitized, are easy for everyone to use.

Once we change from an attitude that information is scarce to an attitude that it is abundant, we can stop hoarding it and become more willing to share it. In the information context, this has made many individuals and organizations much more willing to distribute information freely and easily, without trying to ration access to resources or restrict users.

SEARCHING

The ability to find items inside books and journal articles is a major technological change, which has taken place over more than forty years. Even in the 1930s, people were imagining machines to scan text in the form of bar-coded microfilm. Searching software and hardware started in the 1950s. It was really the 1960s, however, which created a boom in systems for searching and retrieving text; a SIGIR (Special Interest Group in Information Retrieval) conference in the 1960s would attract more attendees than a SIGIR conference today. In this decade the research community switched from Boolean search techniques that largely mirrored the actions that were possible with paper-based systems to searching by coordination level or to vector models. This change, led by the late Prof. Gerard Salton of Harvard and Cornell, created the search methods we still use today. They were not adopted by the industry for several decades; to this day online systems such as Dialog still offer the traditional Boolean search interfaces.

More recently, the Internet search engines, most notably Alta Vista and Google, found out how to use multiple processors and huge cache memories to search queries through billions of Web pages in less than a second. After a brief interval in which everyone complained that there were too many responses to any query and that it was too difficult to find the good ones, Google introduced quality rankings based on web links and solved the problem.

One of the enormous attractions of the search engines is the enormous size and scope of the Web, plus of course the convenience of having it at your desk (or lap). My students are enrolled in a library school; they have often chosen librarianship as a career because they like books. Nevertheless, they do their research on the Web and they prefer reading assignments from the Web to reading assignments from the library. All of us have access to the standard online bibliographic services through the Rutgers library, but I only a few of the students routinely use them in my digital libraries course.

The students are behaving very sensibly. For most queries, the Web is a better solution than the edited and selected services, since it has so much more information available.

A few years ago I looked at a small sample of queries using both professional indexing services and the Web. The first comparison used the ACM Digital Library, and I chose topics that seemed very technical, such as "neural nets", or "RSA cryptography". The ACM Digital Library represents some of the most respected computer science journals and I expected these queries to be well suited to it. Below are the first four results in the searches: remember one is a search in refereed and printed journals, the other is a Google search across the full Web (in 2003).

Query: "neural nets"	
ACM: 554 hits	Google: 131 000 hits
Bounds for the computational power and learning complexity of analog neural nets, 1993; Neural networks and open texture, 1993; Efficient simulation of finite automata by neural nets, 1991; Parallel construction of minimal perfect hashing functions with neural nets, 1993.	Lecture notes from an MSc course on neural nets, 2003; Neural networks at Pacific Northwest National Laboratories, 2001; Old neural net FAQ, 1995; FAQ from comp.ai.neural-nets, 1995.

Query: 'RSA cryptography"	
ACM: 12 hits	Google: 117 000 hits
Hardware speedups in long integer multiplication, 1990; Dynamically reconfigurable architecture for image processor applications, 1999; Representation of ASN.1 in APL nested structures, 2000; Architectural tradeoff in implementing RSA processors, 2002.	RSA Laboratories cryptography FAQ, 2003; RSA Laboratories algorithm simulate center (Javascript), 1999; RSA Cryptography Today FAQ, 1997; RSA Cryptography specifications version 2.0, 1998.

Despite the technical nature of these queries the Google results are more useful to an undergraduate; the ACM results are extremely specialized. Part of the problem, in these cases, is that the ACM library does not contain monographs, but then a hypothetical undergraduate probably does not want to take the time needed to read a whole book. If the queries were more specialized, then Google is even more likely to have a good answer: it simply has so much more material that a random topic is more likely to be represented.

It is actually rather difficult to find a topic that the ACM Digital Library does better than Google (other than picking a known paper and asking about it). To do so I had to resort to asking about topics now obsolete, since the Web basically starts in about 1995 and the ACM library goes back several more decades, thanks to a retrospective conversion project at ACM.

The same results appeared when I looked at Wilson's Art Abstracts. Again, the tables below compare two searches in this professional index of refereed journals with a Web search.

Query: "paleography"	
Art Abstracts: 72 hits	Google: 21 100 hits
Cuneiform: The Evolution of a Multimedia Cuneiform Database; Une Priere de Vengane sur une Tablette de Plomb a Delos; More help from Syria: introducing Emar to biblical study; The Death of Niphururiya and its aftermath; Fruhe Schrift und Techniken der Wirtschaftsverwaltung im alten vorderen Orient.	Manuscripts, paleography, codicology, introductory bibliography; Ductus: an online course in Paleography; BYZANTIUM: Byzantine Paleography; Texts, Manuscripts and Palaeography; The medieval paleography tutorial has moved to...

Query: "Raphael, fresco"	
Art Abstracts: 15 hits	Google: 8 950 hits
Sappho, Apollo, Neophythagorean theory, and numine afflatur in Raphael's fresco of the Parnassus; Accidentally before, deliberately after (Raphael's School of Athens); Raphael's Disputa: medieval theology seen through the eyes of Pico della Mirandola, and the possible inventor of the program, Tommaso Inghirami Raphael's use of shading revealed (restoration of the Parnassus in the Stanza Della Segnatura almost completed).	Raphael: The School of Athens; WebMuseum: Raphael: the nymph Galatea; OnArt Posterstore: Art Photography Music Film Posters; Raphael: Olga's Gallery.

Search engines make possible the abundance of information that we now have. Without them, we would find ourselves spending too much time figuring out what to look at and be back in the world Vannevar Bush knew where large libraries could only be nibbled at by a few. Where we cannot search

content, as with sound files, we are generally dependent on knowing what we want, and the abundance of information is not as useful as we would like.

READING

During the end of the dot-com boom, in 2000~2001, we had a sudden enthusiasm for portable book readers. The best-known device was probably the Rocket E-book, but there were other devices (Softbook) and there were methods of reading ebooks on ordinary computers and handheld PDAs. Various publishers issued such ebooks, under labels such as "AtRandom", "iPublish", and Mightywords (subsidiaries, respectively, of Random House, Time Warner, and Barnes & Noble). This particular fad came and went very quickly, although to my surprise the Rocket e-Book is still worth around $100 on eBay, suggesting that there are users out there even in 2005.

What went wrong? There are a wide set of explanations, including hardware, format, content, and price.

Hardware explanations, for example, include limited screen size, insufficient battery life, poor readability in bright light, and excessive weight.

Format explanations would include the inability to see several pages at once in some kind of tabbed mode, disappointment with control options, lack of a way to search across multiple books at once, or other software-related problems.

Content problems were the limited number of books available, since only a small number of in-print books were issued in e-book format.

Price, of course, reflects not just the cost of the reader but the fact that ebooks had prices comparable to paper books. Readers and newspaper columnists often suggested that ebooks should be cheaper than paper books. It is not that making and shipping physical books is that expensive, but that in the electronic world much of the wastage (something like half the paperback books printed in the United States go unsold) and the distribution costs can be avoided.

Perhaps most instructive are the objections based on content. Ebooks from commercial publishers came with "digital rights management" software that limited the ability to pass the book on to a friend or read it on multiple devices. Reader choice was severely limited (although the University of Virginia and others made a variety of public domain texts available easily).

Today the industry is trying again, with the Sony Librie. This product has greater display quality and readability under a variety of lighting conditions, since it is based on an "electronic ink" technology rather than an LCD display. Again, however, the choice of books is limited and the material comes with the encumbrance of digital rights management software.

Prof. Reddy has, from the beginning, hoped that the Million Book Project would be a way of stimulating publishers to solve the problems of providing content. The Million Book Project would define a standard format and would also provide a lot of books that would be easily available. Commercial publishers, realizing that users would be reading other books instead of theirs, would be prompted to try to sell their books as online objects. Million Book Project users would be accustomed to reading online, and so the various format and hardware objections would in practice have been overcome.

Again, quality at the expense of quantity is probably a bad choice. This is not just a matter of selection. Excessive demands for quality in conversion can make projects too costly to complete. For example, the Audio Engineering Society published a report recommending that preservation of analog recordings in digital format be done with sampling rates of at least 88 000 samples per second (AES, 2002). This corresponds to frequencies up to 44 kHz. That is not only well above human hearing (which tails off around 22 kHz even for people who have never listened to loud rock music) but it is even beyond what a dog can hear. Cats can hear 60 kHz, but their interest in electronic systems is usually limited to the heat coming from the cabinet. Unfortunately if someone really does want to do conversion at a 88 000 sampling rate, normal consumer grade electronics equipment cannot be used; it is all designed for CDs at 44 100 samples per second. Nor can such a project do quality control by having people listen to the results. The extra cost of conversion at qualities beyond what users can hear will mean that much less conversion can be done. And, of course, it is extremely unlikely that whatever analog source is being converted had captured signals at 44 kHz anyway; old

microphones, amplifiers, and recorders would not have had that kind of bandwidth.

Surprisingly, there has been little direct comparison of on-screen and on-paper reading. Cornell University and collaborators (including Bellcore) did some studies in the early 1990s (Egan *et al*., 1991), showing that for some tasks the ability to search really helped the users. Today at Rutgers University, Nina Wacholder, Lu Liu, Ying-Hsang Liu, and I have been experimenting with people reading on both paper and screen. We have tried to bias these experiments in favor of paper: we give the users topics that are difficult to search for and designed to encourage "browsing" in the traditional way. Yet we find that people with a book in PDF format routinely search and do as well at rating the book as people with a paper copy, spending only half the time.

Yet when I ask my students how much they will read on screen before they print it out, I typically get answers in the range of 3~4 pages. I believe that this reflects a combination of the hardware ease of reading paper and the lack of organization in long online documents; if one does not have a specific topic to look for, and expect to read the entire document, the simplicity of moving through a series of pages is attractive compared with keeping track of your location in an online browser.

Attempts to build an interface which helps the user navigate through a long document such as a book have been frustrated by the inability to extract significant phrases from a page. The next figures show a possible interface design, with multiple windows showing successively more detailed views of a book. The book used in these experiments deals with English political history from 1815 to 1835 (McCarthy, 1899). In Fig.3 the phrases used are taken from the author's table of contents. The left column shows the chapter titles and the author's phrases, and the right column the text.

Fig.4 shows the same kind of interface but this time with automatically selected words; in general they are not suitable for people who do not already know what the book is about. The problem of selecting phrases from a text to give an idea of the con-

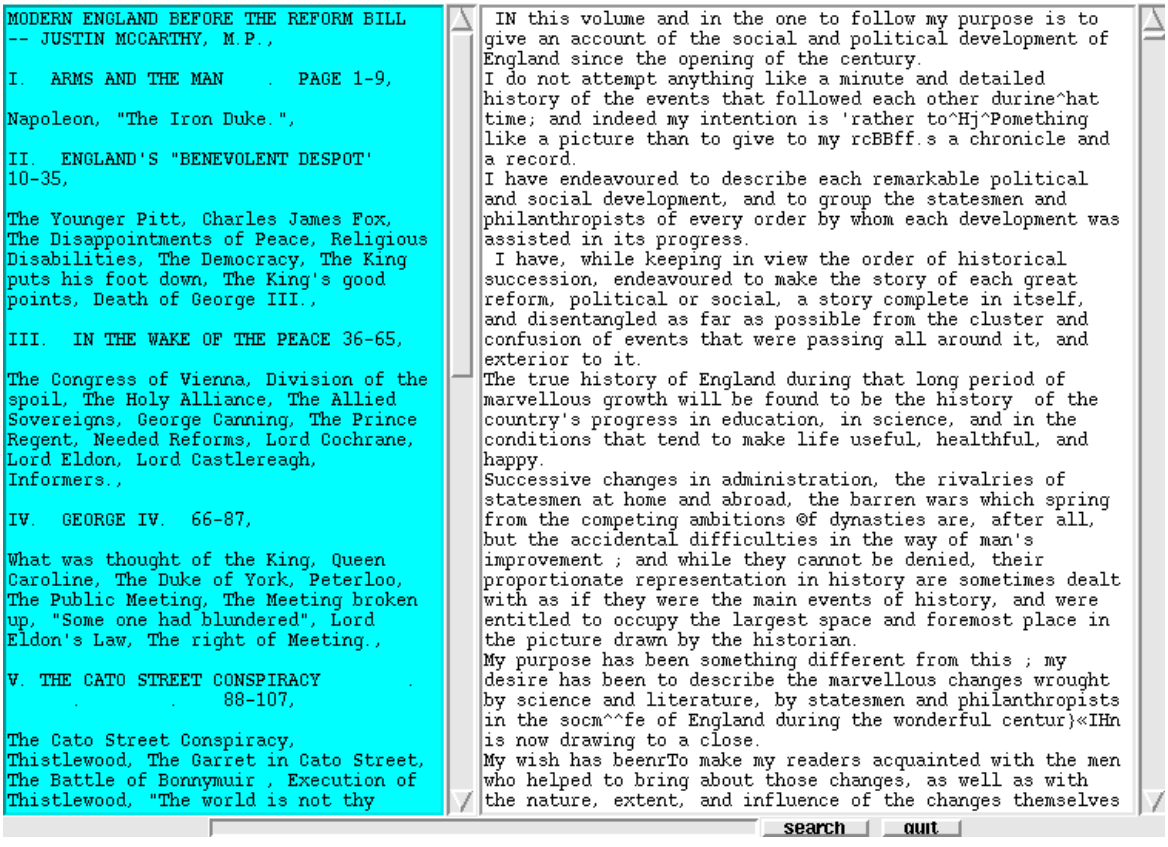

Fig.3 Guidance by author phrases

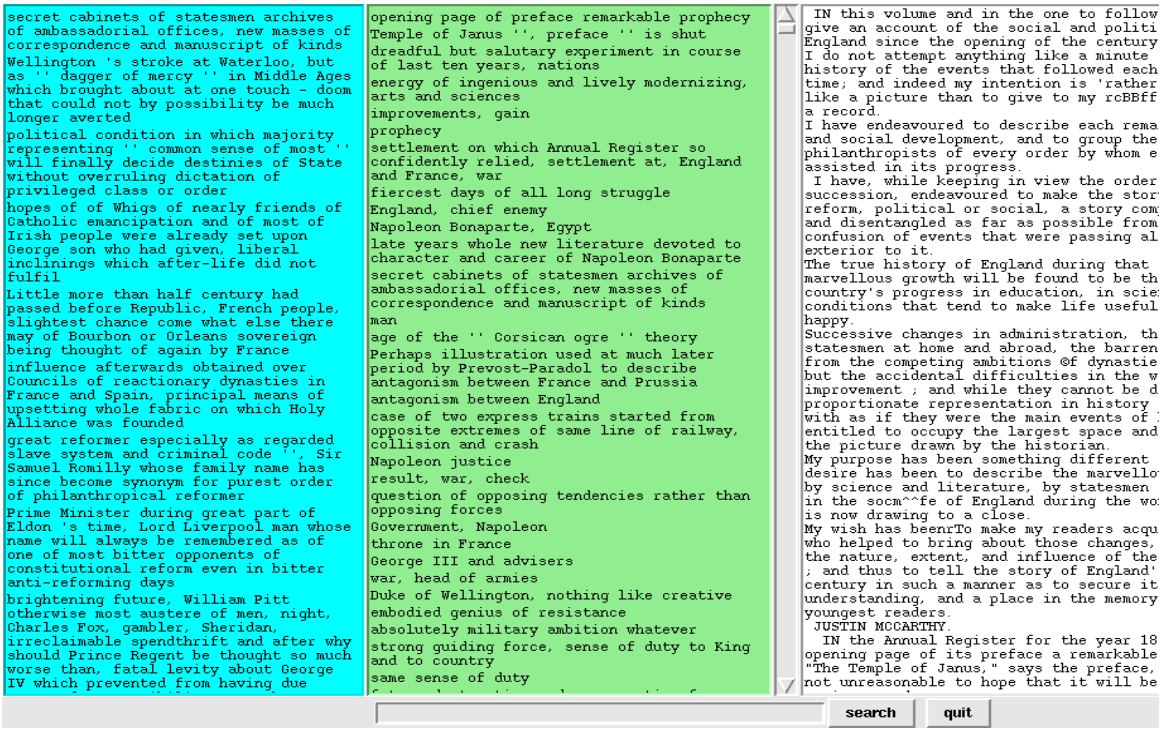

Fig.4 Guidance by automatic phrase selection

tent on a page proves to be too similar to the problem of summarizing the page. Automatic summarization is still a challenging research problem.

Will people read on screens in the future? Some of the answer may be hardware, but I often see people in their offices printing long documents, even though battery life and screen size are not problems. Additional navigation tools are arriving steadily; but it may be that we should be focusing effort on capturing descriptive information about sections of the book as described by the author, rather than trying to generate useful phrases automatically. Another alternative, which we are exploring, is the possibility of exploiting previous user experience to guide new users from page to page.

SOCIAL IMPACT

The traditional publishing system ranges from very slow (academic journals) to daily (newspapers). It involves a fairly restricted list of contributors; we know that the authors of journal papers are usually academic professors, the authors of newspaper articles are journalists, and the authors of computer manuals are technical writers. Merely my ability to list the job titles of each such person confirms that only a select set of people see their writing in print.

The Internet, of course, is readily available to anyone. At times this is bad; spam and obscenity afflict us all. Nevertheless there is an incredible amount of useful, voluntarily contributed information. For years I relied for systems administration for my desktop on the idea that there was somebody else on my corridor that knew more than I did; now I can nearly always go to the Web with whatever problem I have and find that somebody else will have provided a solution.

Volunteer postings are now supplemented by volunteer editing. The advent of the “wiki” documents has shown the ability of a community to create and maintain documents which are relatively free of commercial spam, obscenity, and deliberately provoking nasty remarks. The “Wikipedia” encyclopedia, created by Jimmy Wales and Larry Sanger, has more than 700 000 articles, all accumulated in four years. There are ten languages with more than 50 000 arti-

cles; and there are more than 70 languages which have at least 1 000 articles (Wikipedia, 2005).

Another example of world-wide collaboration by volunteers is the "Distributed Proofreaders" effort of Project Gutenberg. This group, led by Juliet Sutherland and Charles Franks, takes scanned images of books and corrects the transcription. More than 30000 people all over the world are registered, with around 1 000 doing something each week. Several hundred books are finished each month (Franks and Sutherland, 2005).

Traditionally, volunteers in many organizations are at least partially motivated by the social reward of meeting the other volunteers and feeling a sense of belonging and participating. Distributed Proofreaders has participants who typically never meet anyone else on the project; they are working alone at home.

There are now many such projects, although most involve people donating computer time rather than their own time. Some of the first were projects in cryptography, starting with an effort in 1988 by Arjen Lenstra and Mark Manasse (Lenstra and Manasse, 1990) to use spare cycles on individual workstations to win an RSA challenge competition. Today, perhaps the best known such project is "SETI@home" which asks people to donate cycles on their computers to analyze signals from space looking for signals which might suggest the presence of extra-terrestial intelligence (Anderson *et al*., 2002). Other and perhaps more practical examples are distributed computing projects aimed at understanding protein folding, finding AIDS drugs, or searching for compounds which might be active against cancer.

Why do people participate in such "non-social" volunteer efforts? Some actually welcome this aspect of the activity. Distributed Proofreaders, for example, says that some of their participants are disabled individuals who can not do many conventional volunteer activities, but who have been helped by others and wish to repay society in some way. Some of these people feel happy that there is something that they can do for a social purpose that does not require them to be able to leave their house.

Some people want to see their words distributed; some may want to show off their expertise. And some people get satisfaction just from contributing to an effort, and find it convenient that they can do so on their own schedule and with minimal travel. In the end, however, it may be enough that a great many more people get to write and a great many more things get read.

It is hard to see, in fact, how many more specific needs could be filled other than by large volunteer efforts. The total number of users for many sites is limited and many have limited funds; it is not likely that commercial services could profitably serve all of the needs. The great expansion in breadth and depth of information sources is only possible because of the number of people participating.

Perhaps more interesting are the social effects of large scale volunteer information efforts. People will be encouraged to be active rather than passive, and some of the trends encouraged by broadcasting and the mass media might be reversed. Some might fear a loss of social cohesion if we do not all watch the same television programs every night, but the world existed for thousands of years before mass media. One hopes that the online help and sharing will translate into a generally more supportive society. A greater variety of information will be available, more specialized needs can be served, and more people will be participating rather than remaining by the sidelines.

CONCLUSION

As we move from information in short supply to information in great abundance, we will have fewer problems of hoarding, jealousy and conflict over information. Richard Titmuss, in a well-known book (Titmuss, 1970), compared the supply of blood for transfusions in the US, where it was typically paid for, and in the UK, where it was generally given by volunteer donors. Less blood was wasted in the UK, because it was not hoarded until it spoiled. In the same way, digitizing books will turn us from a world in which books sit unread on shelves in research libraries into a world where they are used and enjoyed, whether by scholars, students, genealogists, or just the idly curious.

References

AES (Audio Engineering Society), 2002. Recommendation for Delivery of Recorded Music Projects. http://www.aes.org/technical/documents/AESTD1002.1.03-10_1.pdf.

Anderson, C., 2004. The Long Tail. *WIRED*, Oct. 2004.

Anderson, D.P., Cobb, J., Korpela, E., Lebofsky, M.,

Werthimer, D., 2002. SETI@home: an experiment in public-resource computing. *Commun. ACM*, **45**(11):56-61.

Egan, D.E., Lesk, M.E., Ketchum, R.D., Lochbaum, C.C., Remde, J.R., Littman, M., Landauer, T.K., 1991. Hypertext for the Electronic Library? CORE Sample Results. Hypertext 91, Proc. 3rd Annual ACM Conference on Hypertext, San Antonio, p.299-312.

Franks, C., Sutherland, J., 2005. http://www.pgdp.net (Distributed Proofreaders' site).

Friess, S., 2002. The Web Didn't Kill Libraries. *Christian Science Monitor*, July 25, 2002.

Gorman, M., 2004. Google and God's Mind. *Los Angeles Times*, Dec. 17, 2004.

Irvine, I., 1996. Quoted by Jim Milliot in "Publishers still searching for profits in new media." *Publishers Weekly*, **243**(1):22.

Jauslin, J.F., 2005. Statement by European Librarians on the Web at http://www.hasgard.net/article.php3?id_article=668.

Jeanneney, J.N., 2005. See Agence-France Press, April 28, 2005.

Lenstra, A., Manasse, M., 1990. Factoring by Electronic Mail. *In*: Advances in Cryptology. EUROCRYPT'89, p.355-371.

Lesk, M., 1997. Practical Digital Libraries: Books, Bytes and Bucks. Morgan Kaufmann, San Francisco.

Lesk, M., 2004. Understanding Digital Libraries Morgan Kaufmann (2nd Edition of the 1997 Book). San Francisco.

McCarthy, J., 1899. Modern England Before the Reform Bill. T. Fisher Unwin, London.

Rosenthal, M., 2005. North American Book Market. http://www. fonerbooks.com/booksale.htm.

Titmuss, R., 1970. The Gift Relationship. Allen & Unwin, London.

Wikipedia, 2005. http://en.wikipedia.org/wiki/Wikipedia (it seems only fair to cite their own article about themselves, although there are now many others).