Int J Digit Libr (2006) 6(2): 192–209 / Digital Object Identifier (DOI) 10.1007/s00799-004-0113-9

Appearance and functionality of electronic books

Lessons from the Visual Book and Hyper-TextBook projects

Fabio Crestani[1], **Monica Landoni**[1], **Massimo Melucci**[2]

[1] Department of Computer and Information Sciences, University of Strathclyde, Glasgow, G1 1XH, Scotland, UK
e-mail: {f.crestani,m.landoni}@cis.strath.ac.uk
[2] Department of Information Engineering, University of Padova, Padova, Italy
e-mail: melo@dei.unipd.it

Published online: 8 March 2005 –

Abstract. We present the results and the lessons learned from two separate and independent studies into the design, development, and evaluation of electronic books for information access: the Visual Book and the Hyper-TextBook. The Visual Book explored the importance of the visual component of the book metaphor in the production of "good" electronic books for referencing. The Hyper-TextBook concentrated on the importance of models and techniques for the automatic production of functional electronic versions of textbooks. Both studies started from similar considerations on what kinds of paper books are suitable for translation into electronic form but differ on the prominence given to book appearance and functionalities. The results of these two research projects are critically presented in this paper, with the aim of helping designers and implementers to better integrate appearance and functional aspects of books into a more general methodology for the automatic production of electronic books for information access.

Keywords: Electronic books – Electronic publishing – Design – Evaluation

1 Introduction

How many times have you been on the road and wished you had one of your books with you? Whether for work or for fun, you might want to revisit something from an old college text, reread a passage from a favorite author, or look up something in a technical reference source. Similarly, in the ever-changing Web environment, wouldn't it be nice to know that books covering time-sensitive subjects will be regularly updated to reflect recent industry developments?

Without a doubt, the least-available type of text on the Web today is that from books. We can get just about any newspaper or magazine online, but the contents of millions of printed books remain scarce. Online books currently consist mostly of works whose copyrights have expired[1] or those used to launch volatile hardware devices such as NuvoMedia's Rocket eBook and Librius' Millennium E-Reader [52]. Although hardware products will eventually be important (especially to children now required to carry backpacks of books to school each day), the real revolution is much more one of delivery than devices. Just as MP3 has grown from an obscure audio-compression format to become one of the hottest topics in the music industry, so are electronic books (henceforth called e-books) about to shake up the traditional book publishing business, and e-books will change book publishing much more than e-retailing. While Amazon.com does a great job using the Web to sell existing books and a limited selection of fiction e-book titles, the emerging e-book business will create whole new markets for book content and eventually expand our very idea of what a book is.

Many of the advantages of dealing with e-books are similar to those in other Web businesses. Because there is no physical inventory, storage capacity is huge, e-books can be printed or downloaded on demand and thus are never out of stock, and electronic reader-reader and reader-author interaction is enabled. In addition, e-books let us access a single chapter or even a single page, are searchable, can have versions, and are linkable, while the ability to provide regular updates enable an ongoing relationship with the reader, especially in many nonfiction subjects [49].

Because of their many advantages – such as familiarity, resolution, portability, disposability – paper books

[1] For example, those collected by the Project Gutenberg, available at: `http://www.promo.net/pg/index.html`, and the Project Runeberg, available at: `http://www.lysator.liu.se/runeberg/`.

will clearly be with us for decades to come, but e-books are becoming more and more popular, and they can be with you wherever you go. As the scientific literature on this subject suggests, it is still unclear what kinds of books could be transformed into electronic form by preserving the same effectiveness of the paper-based counterpart. We do not believe that all kinds of books could be translated into electronic form with equal success as their paper-based versions. On the other hand, we believe that some books could be translated into electronic form and reach levels of effectiveness that are higher than their paper-based counterparts.

For the above reasons, we believe much research is needed into the design, implementation, and evaluation of e-books to assure that they will be as effective as possible for the purpose for which they were designed.

In the first part of this paper, we introduce the main notions of e-books. Then, in the second part, we summarize two examples of studies into the design and production of e-books carried out in the context of two independent projects: the *Visual Book* project [19] and the *Hyper-TextBook* project [35]. Both studies started from the assumption that not all paper books are suitable to be translated into electronic form and that scientific publications, which are more used for referencing to than sequentially read, are better candidates for this translation. Despite this common starting point, the studies took two completely different research directions. The Visual Book project studied the importance of the book metaphor to produce e-books in order to retain as much as possible the many powerful features that printed books have. The Hyper-TextBook project explored ways of producing e-books in a fully automatic way using information retrieval techniques to enable fast and effective access to information. The conclusions drawn by these two studies show that these two lines of research complement each other and provide very interesting insights on how e-books for information access (e.g., referencing, fact finding, etc.) should be designed, implemented, and evaluated, despite the differences in focus. Thus, the results of these two research projects are critically presented and compared in the third part of the paper, where we provide some guidelines for a better integration of the two main dimensions of electronic books: appearance and functionality.

The paper is structured as follows. Section 2 presents a detailed view of the book metaphor and how this has been interpreted in e-books. The following two sections describe two projects that have taken completely different views of the book metaphor. Section 3 presents the Visual Book project. Section 4 describes the Hyper-TextBook project. Section 5 compares, through a critical interpretation, the results of the two projects, describing the lessons learned and directions of future work. Section 6 presents the distilled results of the experience in the two projects in the form of some guidelines for future similar studies concerned with the fast and cost-effective production of good e-books for consultation and self-referencing. Section 7 ends the paper with the overall conclusions.

2 The book metaphor in electronic books

The result of integrating classical book structure, or rather the familiar concept of a book, with features that can be provided within an electronic environment is referred to as an *electronic book* (or e-book).

From a conceptual point of view, it is an attempt to overcome the limitations of paper books by adding a series of useful features made possible through the nature of the electronic environment. The main features of e-books are that they are dynamic, reactive, and can be made available in different formats and/or editions in a short time; in addition, with laptops and e-book readers becoming increasingly lighter and more powerful, and Internet connection being so popular and reliable, e-books are easy to access from anywhere.[2] However, the translation from paper to electronic environment is not appropriate for every type of publication and for every type of reader; the process of reading and the tasks readers are attempting to accomplish have a central role in judging the suitability of this translation. In any case, the cognitive overhead that results from the special environment chosen (i.e., the computer) represents a valid reason for carefully considering the appropriateness and the method of realizing this conversion. The fact that technology is able to represent documents on the screen is not a sufficient reason for translating every piece of paper into electronic format. It is important to consider the subject matter and usage of the specific paper book in order to decide whether an electronic version will be useful or not. Observations of the behavior of the market for paper and electronic publications show that currently only a very small percentage of paper publications (up to a maximum of 10% for extremely successful productions, such as encyclopedias and dictionaries) are sold in electronic format, and that in any event users prefer to keep a paper version in addition to the electronic one.

From the technological point of view e-books need powerful processors, large amounts of storage, and the capability to manage combinations of different types of data. The motivations for producing e-books are more complex and deal with economic factors, related to the market requirement for e-books, as well as with cognitive issues related to the ability of readers to use, appreciate, and even prefer books in electronic format to paper ones. The quality aspect becomes crucial, and for this reason the design process for producing e-books has to be reviewed in order to provide "good" e-books, that is, electronic information with some sort of additional value that paper cannot provide readers with.

[2] See `http://www.openebook.com` for the latest product reviews.

To effectively perform a translation from paper book to e-book, it is necessary to introduce an appropriate formalism that is based on the use of metaphors. *Metaphors* are rhetorical tools that are used for introducing and defining new concepts by using similar ones that are already familiar to the public [17]. They have been widely used in the design of human-computer interfaces [30]. A representative and familiar example is the desktop metaphor (used in both Macintosh and PC interfaces) where the onscreen work area is graphically represented as a desktop with documents and folders that are ready to be opened and used. It is clear from this example that very often a metaphor consists of a number of interlinked metaphors, and the way they look and work together is a crucial issue in determining their effectiveness. In the case of electronic documents, the market offers a large choice of products, each of which emphasizes different aspects or interpretations of the original document.

A conceptual classification can be made with respect to e-books that implements the book metaphor in different ways. It is possible to delineate a hierarchy of e-books starting with those that imitate the paper book in all its physical components to the so-called "cyberbooks," which have nothing more in common with the paper book than being a tool to present information to readers. Among the book imitators it is useful to identify additional subclasses each of which focuses on a different aspect of the book metaphor and gives greater or lesser emphasis to the features inherited from the paper book.

Thus our classification ranges between:

- Page-turner books
- Scrolling books
- Portable books
- Multimedia books
- Hypermedia books
- Cyberbooks

2.1 Page turners

Page-turner books can be divided into those that imitate the original paper book and those that have no paper counterpart and imitate the general idea of a book, i.e., the book metaphor. Among those with a paper counterpart there are different levels of closeness, from those that maintain all the visual features of the original book by using a picture of the original pages, to those that use a graphic template to imitate the original book but do not allow the reader to interact with it in the same way as a paper book. We are referring to the most classic and widely adopted static form of e-books, i.e., Portable Document Format (PDF). Publishers appreciate PDF for a number of reasons: it helps preserve the original form of the text, it is really easy and cheap to produce, it is supported by a reliable software provider, and it does not facilitate copying and pasting. At the same time it allows simple interaction with the content such as annotations, bookmarking, and highlighting, depending on the version of the reader used. But still it is far from ideal from a usability point of view [44].

The Bodleian Library has developed a system that presents the user with an online picture of the page, as it appears in the traditional paper publication, with hypertext links to other documents.[3] Another example of a page turner has been developed in the Dante project,[4] which provides pictures of the originals from the "Inferno" and other parts of the *Divina Commedia*. The same sort of approach is found in *The Dead Sea Scrolls*. All of these are mainly presented as high-resolution pictures because the appearance of the original manuscript has a clear artistic value, which has greater significance than just the textual content, or the document is damaged, thus making a definitive transcription difficult and allowing scholars to have access to the source material to verify the transcription themselves.

A large family of page turners was based on HyperCard, a hypertext authoring software application released by Apple in 1986. This family of e-books was more hypertext oriented than the previous group and pays particular attention to the definition of links and browsing techniques inside the text presented in the book. In this way the translation to electronic form adds some value to the original book. Examples of systems based on Hypercard are Vortex [15] and Hyper-Book [18].

Another example in the page-turner class is the book emulator, which incorporates an electronic bookshelf on which are displayed the spines of the electronic books [12]. The electronic representation of each book on the shelf is directly related to the number of pages in the book, providing a search cue for familiar books. It is envisaged that publishers would provide distinctive spine designs to capture the attention of the casual browser and provide a further cue for familiar books and for those belonging to a familiar series. The book has a structure consistent with its paper equivalent, i.e., chapters are in progressive order, the contents of the book are listed on the front, and the index and references are located at the back of the book. If the contents page is displayed, the user can note the page number, jump to the location within the book, and flip through pages until the required section is reached. Various functionalities are available within this kind of book that can be referred to as reader services. For example, pages may be turned, and selecting within the black bands enables the reader to make large approximate jumps; pages may be annotated either with a normal pen or with a highlighting pen, and such annotations may be removed with an eraser; bookmarks may be inserted into the book so that when the marker is selected the book turns to the marked page.

A project that uses page turners is the Mercury project [7]. The project aimed at demonstrating that

[3] `http://www.bodley.ox.ac.uk`

[4] `http://etcweb.princeton.edu/dante/`

the (then) current technology (high-speed networks, high-resolution screens, multimedia facilities) and techniques for processing electronic documents made it possible to build full-text electronic libraries. Very similar to the Mercury project is CORE [24], which produced a prototype for the storage, searching, and displaying of scientific journals in electronic form together with related graphical information such as tables and graphics.

2.2 Scrolling books

In scrolling books text is presented according to a scroll metaphor. This was the classic way to write text on parchment, and in modern times this metaphor is very close to that used in word processor environments. This strategy lets the designer determine the page size based on the space available and dimensions of the screen. The page no longer exists as a logical and physical unit, nor does any reference to page numbers or to the page sequence. The text scrolls almost without any physical limitation. As a result the electronic scrolling book is portable on different platforms and is in no way dependent on screen dimensions. However, readers lose one of the classical and fundamental keys to accessing information in a book, the page number, and they can easily get lost in the flow of information. The book metaphor is kept in its logical structure; the information is presented according to a book-style hierarchy, made of chapters, subchapters, paragraphs, and sections. Another important observation is that information in this class of electronic book is made of text and graphics, even if there may be some hypertextual features such as the presence of links to browse the book. This is still a very traditional way of interpreting the concept of book in electronic terms, and for this reason it is closely related to the paper book. Examples can be found in extensive text repositories such as those organized and maintained by the Gutenberg Project or the Library of Congress.

Other examples of scrolling books with more sophisticated hypertextual interfaces are Dynatext and SuperBook [25]. Both of these provide full-text indexing, links, navigation, and orientation through a dynamic table of contents and a multiwindow text display. An interesting aspect of these two systems is the fact that they provide the capability for automatically importing text that is available electronically in different formats.

The WEB Book project [37, 38] is yet another example of scrolling book to be presented on the Web. The aim was to study whether focusing on the appearance of the content when preparing a scientific textbook for electronic publication had a positive impact on its usability. The WEB Book project concentrated on scannability/legibility issues, applying Morkes and Nielsen guidelines to the presentation of textbook material on the Web [43].

2.3 Portable books

Portable books are becoming more and more common as appropriate technology has developed. They imitate the book as a portable tool for providing information. A side problem is that they have to deal with limitations in screen size, resolution, and efficiency, but these are all technological rather than conceptual aspects. An example of a portable electronic book is the one presented by Feldman [26]. The main feature of this electronic book is its portability based on the use of a laptop computer. A similar project is the Virtual Book developed at Digital, whose main component – Lectrice – is defined as "a prototype reading appliance" by its authors [32]. The main effort has been put into digitizing, recognizing, and compressing the original pages of the documents to be viewed. The result is a very pleasant, easy-to-read, portable tool that allows users to read and flip through the pages of the chosen electronic document without frustrating delays. This still has to be considered as more of a viewer than as an example of an electronic book as its authors have decided not to be involved in any design processes of the document to be shown; instead they have concentrated on the functional interface. The same philosophy has been followed by Sony for the Data Discman and the more advanced Bookman, both providing a small and light high-resolution monitor for presenting electronic books that can be prepared by using the Sony Electronic Book Authoring System [13].

The area of portable electronic books received a boost from the interest shown by Microsoft, and a series of brand new products hit the market in the late 1990s. These all shared the same philosophy and could be better named portable electronic book readers. Although some of these products have since disappeared from the market, it was clear that the competition was on how well these products were implemented, rather than designed, and how much they would cost users. While many joint efforts were made to achieve an even higher screen resolution in order to increase competitiveness of future e-books [39], a project at the MIT Media Lab and Xerox PARC presented an innovative approach [40]. This project aimed at producing an electronic book in the form of a bound set of pages using "digital ink" that may soon replicate the visual clarity of permanent black ink on white paper. Users would move forward in the information by flipping pages just as they do in a printed book. However, the electronic book uses special flexible "paper" that can be rewritten under computer control. Thus, a single physical book can turn into any desired work simply by downloading the corresponding file. In this way the e-book becomes a sophisticated transportable high-tech object where it is possible to store and reaccess various sorts of information, but still it is not expected to become a substitute for the paper book.

2.4 Multimedia books

Multimedia books represent a further step away from the paper book. The contents of such books are no longer simply electronic text or pictures but a mixture of different media such as video, sound, animation, text, and pictures. It is no longer possible to keep this enriched form of information inside the physical border of a classical book. That is why the electronic environment is the natural one for this class of book. Such books still borrow essential features from the book metaphor by either imitating its physical appearance or keeping the same logical structure. The metaphor is enlarged to consider this new form of information as generic book contents and to organize it in a new book container according to new needs and presentation paradigms. See [9, 10] for an example of such books.

It is also worth mentioning multimedia books that are essentially composed of pictures and various contributions from different media. An interesting approach has been made by Barker [8] in connection with children's books. In this type of book visual information is crucial, and particular attention must be paid to how this information is presented to the reader. This requires special attention to the definition of the logical screen, a conceptual place where information in its visual form is displayed. Depending on the nature of the source information, different instances of this logical screen are generated by the authoring system. For each kind of information the most expressive and powerful representation has to be chosen and at the same time the representation has to be dynamic in order to allow users to change it depending on their individual needs.

2.5 Hypermedia books

Hypermedia books present textual material and integrate it with other related sources, such as video, sounds, and pictures, and provide the reader with alternative reading/browsing paths. The resulting book is an augmented version of the original. Hypermedia books inherit all the problems related to the use of hypertext and hypermedia such as orientation problems and the risk of confusing users with too much information. A rich subset of hypermedia books are those that are available on the Internet, an increasingly common phenomenon. An example is "Lady Freedom among us," a poem readable or, better, read on the Internet as part of a program celebrating the acquisition of the four millionth volume at the University of Virginia Library [22]. It is possible to see and listen to the poem read by its writer. Another group of electronic books that are widely available are those on tape where actors read for the reader/listener pieces from classical literature.

2.6 Cyberbooks

Cyberbooks are completely free from any physical/conceptual dependence on the paper book, as they have only appeared in electronic form. In this context the term "book" is used in its broadest sense as a repository for information. On the other hand cyberbooks depend very much on the dynamic nature of their context, the computer. In this sense they can be defined as active books with which readers can interact. They are part of an alternative new line in modern literature, called postmodern literature, which closely integrates the computer culture with the classical human one. An example of these books is [34].

3 A study of e-book visual appearance: the Visual Book project

The main assumption of the Visual Book (VB) project was that paper documentation could be organized on the basis of the translation of the concept of a library into an electronic form.

The original aspect of the Visual Book experience is the importance given to the visual components of the physical book when designing e-books together with the interpretation of an e-book as part of an electronic library intended as an informative system with specific and innovative features [35]. A new object, the Visual Book, has been studied within the context of an electronic library by following an original approach that has highlighted and exploited its relation with the real object it imitates. In particular, a new aspect, which has not been exploited to date, is that of *visual rhetoric*, which is important to the design of both paper and e-books.

The definition of visual rhetoric is tied, at least in a first instance, to the concept of text rhetoric, from which it is derived, as this is a well-established and popular example of rhetoric applied to written information. The idea is to define those parts that are more important for the comprehension of the meaning of the text. To achieve this purpose, verbal and/or graphical techniques can be used. In fact, a document can be interpreted as a visible representation of a text according to its semantic contents [51]. Thus visual rhetoric is simply the translation into graphical terms of the text rhetoric that results from both the logical structure of the text and its pragmatic component. It provides the reader with a graphical markup language that is immediately recognizable on the basis of previous reading activity. Different graphical presentations suggest different readings and affect deeply the interpretation of the contents of the same text. These observations lead one to conclude that visual rhetoric is a crucial aspect for both reading and browsing a document, as the findings of the Visual Book project have confirmed.

The *Visual Book project* focused on the importance and use of visual rhetoric when presenting information on a screen and in particular on the influence of visual rhetoric in the presentation and use of e-books. The work was carried out during 1995-1997 at the Department of

Information Science of Strathclyde University (Glasgow, UK). The study started with a comparison of the effect of visual rhetoric on the paper book and its impact on the presentation of the same book when translated into electronic form. A further step was to consider whether visual rhetoric could also have an impact on the design and presentation of electronic publications that had no paper counterparts.

3.1 Design issues

We believe that, thus far, the emphasis in the production of electronic books has been mainly on the final product rather than on defining a paradigm or a set of rules in order to achieve an abstract object worthy of the label "e-book." The definition of what an e-book is, or has to be, is quite loose, and unfortunately this is unavoidable because of the nature of the object and the difficulty of defining something with so many aspects and functionalities. A possible solution was to collect heuristic rules and principles as a result of careful analysis of examples of e-books that already exist and use the reader perspective as a discriminating factor. In general the perspective of the whole Visual Book project was centered on readers and their needs, expectations, and requirements. This was the first guideline and the main concern while designing the Visual Book system. After making this basic assumption, it became easier to find some elementary and basic guidelines for producing e-books. The first step was the definition of three different basic types of e-books:

- Those created from books that have only existed on paper;
- Those originally generated in electronic form for subsequent printing on paper;
- Those available only in electronic form.

The Visual Book project focused on the first type and mainly on the process of translation of existing paper books from paper to electronic form. The book metaphor plays a crucial role in the definition of guidelines for the design of the Visual Book. These guidelines fall into two broad categories. The first category is concerned with establishing whether a book should be produced in electronic form. In this contest, the reasons for producing books in electronic form need to be addressed. The second category considers when the term "e-book" can be applied properly with reference to some basic requirements that need to be satisfied in order to make the use of this label/term appropriate. In order to be called an "e-book," an electronic product has to resemble, be consistent with, and work according to some or all aspects of the book metaphor with no ambiguities, conflicts, inconsistencies, or confusion. Situations must be avoided in which the page metaphor is not respected, or the logical structure of the book is not considered, or again the book template is used to present different kinds of information, not necessarily book related, the result being a heterogeneous system that has to be designed very carefully in order not to confuse users.

Thus, the Visual Book is just an example of how the study of existing trends in producing e-books, together with the observation of readers' reactions, can lead to the definition of heuristic rules for the production of "good" and "readable" e-books.

3.2 The Visual Book

The Visual Book represents a particular interpretation of the e-book, based mainly on the visual aspects of the real book, its physical features, such as dimensions, thickness, page form, and general design style [35]. The Visual Book is the result of a process of conversion of existing paper books into electronic form; it is comprised of two main components:

- The Visual Book Builder (Book Laboratory plus Book Compositor), the authoring system for building the Visual Book;
- The Visual Book Viewer (Book Browser), the module for presenting the Visual Book to readers.

In this way the Visual Book system is a practical approach to real book translation. The physical aspect of the real book is the key to the construction and representation of the Visual Book. The focus is on the book's appearance, interpreted as a way of conveying the cognitive background the reader already has. Book functionalities are studied in relation to the book's real use, so information is presented in a natural and familiar way, and at the same time the reader's activities are supported in a comfortable environment.

An important point in Visual Book development is the choice of the original paper book to translate [36]. Not every kind of publication is suitable for electronic translation. In the case of the Visual Book project, scientific publications were chosen as they are used and consulted more than read by users who are already familiar with computers and could better appreciate the additional features and functionalities offered by the electronic translation. The main idea of a Visual Book is to offer the reader an object as similar as possible to the paper book, replacing its essential physical features such as size and quality with visual ones. Another important principle in Visual Book design has been to determine which paper book features are more useful and familiar to the reader and to reproduce them in an electronic environment. Particular attention has been paid to the functionalities a Visual Book could acquire from electronic support and the choice of those that would be most useful and consistent with the specification of the book concept. One of the main advantages of electronic support is its dynamic nature, which allows the modification and updating of data in a simple way. On the other hand a set of functionalities already available in paper books but not currently in their electronic translation are widely used by conven-

tional readers, such as: bookmarks, notes on the margins or elsewhere in the text, highlighting of interesting parts, easy access to pages frequently consulted following different clues, information about the ratio between pages already read and those remaining, control of reading progress, and browsing for interesting sections.

All of these functionalities have been translated into electronic terms and are essential features of the Visual Book Builder and the Visual Book Viewer, as shown in Figs. 1 and 2. The project chose and used *Information Retrieval* by C.J. van Rijsbergen for its first prototype.

Besides, the Visual Book model takes as its basis the paper book and translates it into electronic form while focusing on its visual components. The general structure is the same as for the paper book, and it is possible to recognize an ordered set of units called visual pages as a set of specific visual elements composed of images of the original cover page, back cover page, subject index, table of contents, and body pages.

The interface for the Visual Book consists of two parts: a graphical simulation of the shape and dimensions

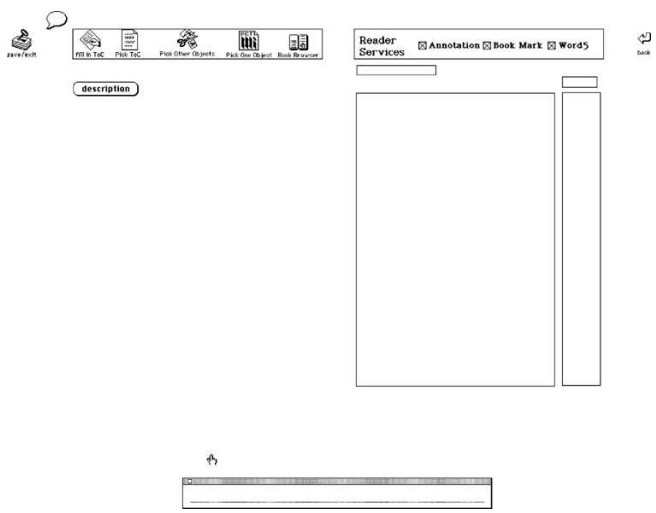

Fig. 1. The Visual Book Builder

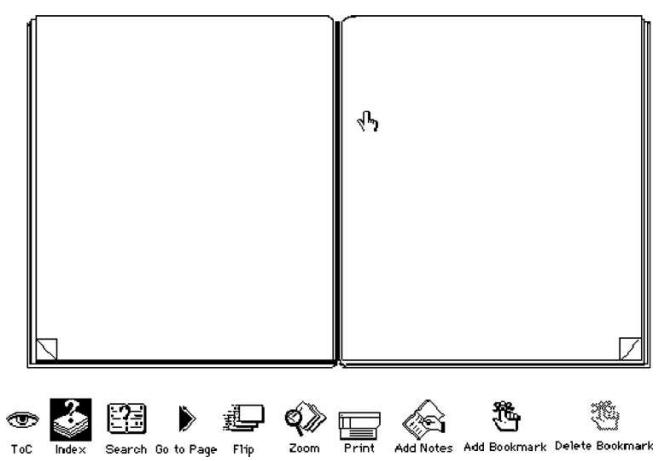

Fig. 2. The Visual Book Viewer

of the book and a set of tools that allow the reader to use the Visual Book by imitating the normal actions that are performed on a paper book. The representation of the content of the original book is particularly important. The image of each page is presented on the screen in an attempt to reproduce exactly its original format. In this way the information maintains the format resulting from the initial design process on paper. This choice allows readers to exploit a memory of the original page that they may already have and provides a presentation paradigm for electronic information based on conventional use and cognitive relevance.

The importance of the various components of the Visual Book, the impact of their visual representation, and the effectiveness of the provided functionalities in helping users to find and use the information they need has been evaluated, and the next section discusses the findings of the Visual Book project.

3.3 Conclusions from the Visual Book project

The evaluation of the Visual Book has concentrated on various aspects of the book metaphor when translating paper books into electronic format. In particular, the importance of maintaining the same typographical features, the design rules, the pagination format, and the paper appearance have been considered carefully in order to retain the whole meaning of the original text. The physical aspects of the page are intended to be encapsulated in a set of presentation rules through which visual clues give the reader semantic information about the context. This primary feature of the book metaphor has been considered to be the most relevant and to be an original contribution to the future development of the electronic design process. The effectiveness of traditional style in presenting information through visual clues such as a specific choice of font, style, typefaces, headers, footers, justification, and spacing rules can also be considered to be valid in an electronic environment through the validation of this prototype.

The evaluation method applied to the Visual Book was the cognitive jogthrough [46]. In this approach, users were free to express, in written or verbal form, their comments and suggestions. It was possible to extract some relevant factors from other experiences in the evaluation of reading processes from a computer [24, 31] and use them as the basis for the Visual Book evaluation:

– *Sense of directness* is the degree of feeling that the changes on the screen are the result of the user's actions. It is connected with the illusion the user has that the displayed image is a physical object that can be manipulated as a real paper book. A sense of directness helps a user learn and internalize the interface to a system because every response by the system reinforces the user's confidence and understanding. In a system with a high sense of directness, users can concentrate on the task at hand without the cognitive overload of understanding system reactions.

- *Sense of engagement* is the level of interest the system induces in the user. The result of a good level of engagement is a high level of concentration that makes users interested in their task. One source of engagement is the fun of seeing the system react and is related to the novelty of the system; tangibility and responsiveness are also responsible for a good level of engagement. Paper has generally a low level of engagement because it is not interactive and is already familiar to users.
- *Sense of text* is the feeling a user may have of the structural and semantic structure of the text being read, i.e., its spatial disposition. Readers are able to recall the position of text in a paper text [45]. This fact connects a semantic entity, the information, with a physical one, which has visual and tactile cues. Factors that can influence the sense of text are the page size, limited legibility, and a low responsiveness while scrolling when looking for more text.

In general the results of the evaluation of the Visual Book, supported by the findings of a similar project, the Hyper-Book [18], showed that the book metaphor was well accepted and fully understood by the evaluators. They also highlight the necessity for a new role in electronic publishing: the *designer of e-books*. This new professional figure corresponds to a person competent not only in the technological aspects of producing electronic documentation but with a background in understanding the importance of presentation issues such as pagination and the general format and appearance of the electronic document, which can be called electronic typography. The importance of electronic typography, where typographical rules are essential components of the cognitive model of a book and for this reason are relevant to the translation of the metaphor of the book into electronic form, has been proved by examining the state of the art of electronic publications and the reasons for their lack of success and comparing it with the history of the paper book through explicit analogies. All these issues have been summed up in the concept of visual rhetoric. The results of the evaluation of the Visual Book project show the importance of visual rhetoric as the application of a visual rhetoric to the book image facilitates the extraction of its logical structure and thus provides essential information to the designer of visual books [35].

The main conclusion from the evaluation of the Visual Book project is that the Visual Book is a powerful tool for presenting electronic text to readers who are more interested in reading and using it than in learning to use a new technology. However, the Visual Book needs to provide more powerful browsing and searching facilities to its users, taking advantage of the characteristics offered by the electronic medium.

One of the results of users' positive evaluation of the Visual Book system was a list of recommended future developments. In particular the evaluators felt that the Visual Book Browser should be expanded to incorporate more sophisticated features related to the capability of the computer. Users were happy to interact with an object that resembled a book and appreciated that its enhanced functionalities were consistent with the original paper version. The main request was for an intelligent search function to simulate and enhance the way readers search in paper books. Such a function should combine the precision provided by paper visual clues in defining the context of the search and by index terms compiled by a human indexer, with the higher recall provided by full text search that can be performed very efficiently using mainstream information retrieval software.

Another important result of the evaluation was the finding that evaluators preferred the subject index to the table of contents since they found it more book oriented and more satisfying to use. However, the table of contents then scored better in the sense of directness because of its relative simplicity when compared with the Index. It is also important to point out that the table of contents and the subject index have very different roles in the book metaphor, even if both are navigational tools. In particular the table of contents is used for skimming the content of a book when readers do not know what can be found inside, i.e., as an exploratory tool, although it is also used by readers familiar with the content to jump directly to specific sections of interest. The subject index, on the other hand, has been designed to help readers find something they know about and wish to find inside the book, i.e., it operates as a searching tool. The comparison of these two tools has to take into account this difference in roles as well as considering the interface and system aspects.

The preference for using the subject index tool can be interpreted as a tendency to use more sophisticated tools for searching for information, especially when they are fully supported by an electronic medium. Even if the table of contents has been rated as understandable and easy to use, as shown by the comparison in the case of the sense of directness, still the information contained in it is not sufficient for readers looking for a specific topic. This demonstrates that a good interface design and a successful mapping between the system and the book metaphor is not enough to make the reader happy. The information is still the main requirement, and future design for electronic publications will have to consider this primary need overall. The presentation issue is also very strong, as the comparison of the sense of directness shows; the fact that the Index is not as natural as the other tools for browsing makes it less direct to use. The same happens with paper books, where the index is usually hidden at the end of the book. However, in the case of the Visual Book the evaluators were more interested in the subject index functionality as the medium makes it a powerful and realizable tool that should be seriously considered during the production of useful electronic publications. It is a case where the computer can help to make a paper

book tool more efficient by making it more explicit and more powerful, through the interface, by moving it from the back of the book, and from the system point of view, making it active.

4 A study of e-book functionalities: the Hyper-TextBook project

The Hyper-TextBook project was carried out during 1997–2000 at the Department of Electronics and Informatics of the University of Padova (Padova, Italy). The project drew from the University of Padova's past experience with converting from textual to hypertextual form such diverse collections of documents as collections of bibliographic references, abstracts, and full text journal articles (see, for example, [3, 4]). It was recognized that different types of documents require different automatic transformation techniques and that there is quite a difference between converting an abstract, a journal article, or a book into a hypertextual format.

The Hyper-TextBook project is concerned with the design, implementation, and evaluation of a fully automatic tool for the creation of hypertextbooks from textbooks. Some results of this project were presented in [19, 20].

Why textbooks? Because most textbooks differ in many aspects from other kinds of books (e.g., fiction, biographies, etc.). The classical linear structure is more important in a book than in a textbook. A novel, for example, has a narrative structure that is inherently linear, and the content of the novel would not make sense if the book was not read from beginning to end ([34] is a notable exception to this observation). Most scientific books and articles, too, are inherently linear, following a traditional thread that goes from introduction to conclusion, via the presentation of related work, the main contribution, and the experimentation. This is not the case with textbooks. A textbook is a book containing facts about a specific subject that is used by people studying or referencing that subject. Although it is true that the order of chapters and sections may reflect a carefully designed way of learning, the information presented in a textbook does not need to be accessed in a linear way. Very often people approach a textbook searching for a specific topic, sometimes looking for a more in-depth treatment of a specific subject about which they already have some knowledge. For this reason, the material presented in a textbook is often organized in a such a way that the treatment of a topic is self-contained in a particular chapter or section. A textbook is also often used for reference, in particular after the subject dealt with by the textbook has already been assimilated. This is particularly true of scientific textbooks, with formulae and data that cannot be easily memorized by the reader but that are known to the reader and therefore can be recalled quickly once accessed in the textbook. Rather than carrying around a textbook, a digital library could make them available online for easy and ubiquitous access.

Although the previous consideration regarding the differences between books and textbooks may perhaps seem a dangerous generalization, since many textbooks are not suitable as reference material, it was believed that the construction of hypertextbooks needed to be carried out using a different methodology from the one used to construct other types of hyperbooks. In particular, a hypertextbook should provide a larger number of searching and browsing functionalities to enable a user to access in the fastest and most effective way the (and only the) information sought without spending too much time looking for it.

Even though it is easy to recognize that a general-purpose automatic technique for the authoring of hypertextual e-books is not an achievable goal, when we focus on a particular typology of books and hypertexts, it is possible to design an appropriate automatic technique.

The work in the HyperTextBook project focused on the use of information retrieval (IR) techniques for the automatic authoring of hypertexts [5]. The reason for the use of IR techniques for the automatic authoring of a hypertext lies in the fact that the IR area deals with methods and techniques for content-based management and retrieval of information [27]. Since the most difficult part of the automatic construction of a hypertext is the building of links that connect semantically related document fragments, it is natural to concentrate on IR techniques that have always dealt with the construction of relationships dependent on the mutual relevance of objects to relate. The use of IR techniques for automatic authoring is particularly useful in the case of authoring a hypertextbook since the presence of a large number of quite small and highly related pieces of text is a particular feature of textbooks.

4.1 Design issues

One of the main assumptions of the Hyper-TextBook project was that the final hypertextbook should preserve the structure and the features of the paper version. This choice would facilitate comparative studies between textbook and hypertextbook and would retain those features of the textbook (e.g., references to pages, typographical aspects, etc.) that the user was long familiar with. Thus the already available text-based features should be enhanced, not removed, by the implementation of hypertext capabilities.

In addition, as hypertext systems become widely available and their popularity increases, attention has turned to converting existing textual documents into hypertextual form.

Many projects have focused on converting existing textual documents into hypertextual form, like, for example, SuperBook [23], CORE [24], Hyperbook [18], and IRIDES [4], and much research has been devoted to this

issue in other areas as well (see, for example, [33]). In particular, research has been concerned with the effective identification of interdocument and cross-document links to enable a user to access information in a nonlinear way (e.g., [6, 14, 28, 29, 50]).

One of the starting points of the HyperTextBook project was to recognize that the complexity of automatic text-to-hypertext conversion is mainly due to the structure of the textbook, to the complex nature of the relationships existing between the different parts of the textbook and between these parts and the subject index. Two problems should be addressed in converting a textbook into a hypertextbook:

- The need to identify a textbook structure in terms of small and homogeneous text excerpts to be used effectively as hypertextbook nodes;
- The need to set up additional content-bearing links among the different parts of the textbook like, for example, full-text pages, bibliography, and subject index.

In [19] the need for detecting alternative text structures other than those given by the author was highlighted; however, the identification of an alternative textbook structure was beyond the scope of the project. Instead, the project concentrated on the construction of content-bearing links between text fragments. Text structure identification still has many open problems such as passage retrieval, theme extraction, and text structuring. To date some local solutions have already been proposed [16, 47, 48], but no general techniques are available. Therefore, to maintain the structure and the features of the original paper version, the original book pages were preserved. The original pages were translated into HTML and presented to the user through a standard Web browser (e.g., Netscape or Microsoft Internet Explorer) after having been authored through the insertion of links to other pages or data semantically relevant. Relevance relationships from subject index terms to pages were also preserved. The subject index represents the main and preferred structure for browsing since it is the set of terms the author used to index the textbook. However, the original subject index cannot be considered as a conceptual structure useful for effectively revealing the contents of a book as, for example, a thesaurus would be. This is because terms in the subject index are organized in a simple alphabetically ranked list. A term is associated to other terms by simple semantic relationships like "see" or "see also," which are not used very often by authors. The subject index would be much more useful if it were expanded with more links between terms and pages or enhanced through richer relationships between terms.

In [1] a two-level conceptual model of hypertext called EXPLICIT was presented. The main elements of the model are nodes and links. We defined two types of nodes: the list nodes and the data nodes. Data nodes are containers of information, or of links to list nodes. List nodes are lists of links to data nodes. Data nodes are specialized into auxiliary data nodes and elementary data nodes, and both are organized on the two levels of the model. The first level consists of elementary data nodes storing the individual pages of the textbook, while the second level is that of the auxiliary data nodes describing the book content, as depicted in Fig. 3. We call the former page level (P) and the latter term level (T). The T level is then the set of subject index terms, whereas P corresponds to the book pages. Figure 4 also depicts the relation between data nodes and list nodes. Note how the presence of list nodes enables the modeling of one to many relationships between data nodes.

In [2] a methodology for the construction of a hypertext organized according to EXPLICIT model was presented. The methodology makes use of IR techniques for extracting and linking semantic objects like index terms

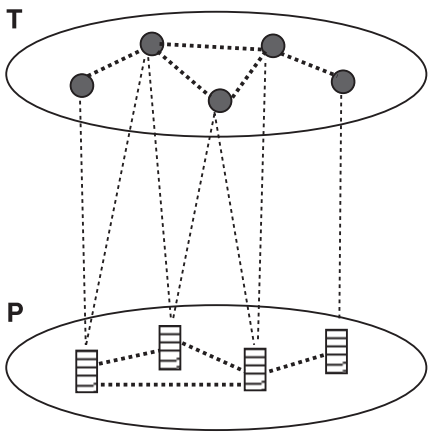

Fig. 3. The HyperTextbook conceptual structure

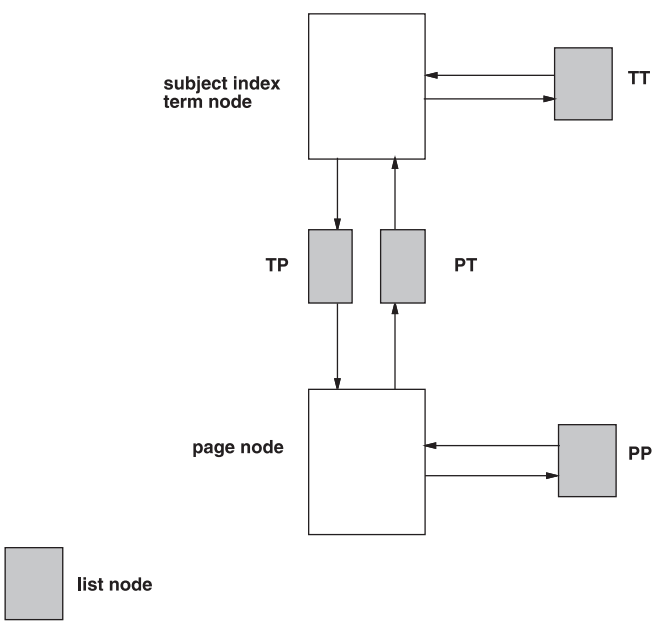

Fig. 4. The HyperTextbook conceptual structure

or parts of text. The major advantage of the methodology and of the resulting hypertext is in enabling direct access to the information one is looking for through an extensive but clear network of auxiliary data automatically extracted from and linked to the original text. These characteristics seemed to be just some of those required by the process of transforming a textbook into a hypertextbook. In [19] a case study of the automatic construction of a hypertextbook from a specific textbook was presented. In this work we made use of an automatic technique that we developed to insert links between text excerpts of a textbook and items in the subject index. These links enable one to access parts of a textbook that have not been specifically indexed by the author but that are semantically related to items in the subject index. Such links are meant to improve the effectiveness of the use of the book in search-oriented tasks. An evaluation of this work was presented in [20].

4.2 The HyperTextBook

A full description of the automatic construction of a hypertextbook is outside the scope of this paper and can be found in [19]. Here we will briefly present an example of the use of a hypertextbook produced using the HyperTextBook methodology. The example describes how a user can browse and search a hypertextbook. This hypertextbook[5] has been constructed from the textbook "Information Retrieval" by Van Rijsbergen [53] and is currently used as a teaching and self referencing tool at the University of Strathclyde.

Figure 5 shows the homepage of the hypertextbook. Three entry points for browsing are available. The first one allows access to the hypertextual version of the subject index, from which the user can go to a term. The second entry point enables access to the table of contents. The third entry point corresponds to the list of pages. In this example the user goes directly to the subject index, where s/he selects a particular term that best represents what s/he is looking for.

Figure 6 is the page for the term "cluster profile" of the subject index. From this page the user can access:

- A TP list page reporting pages relevant to the current term (according to its patterns of occurrence in the pages);
- A TT list page reporting terms similar to the current term (according to the pattern of co-occurrence of terms in the textbook);
- Other T pages with terms linked through the manually inserted "see also" links (according to the author's subject index).

Here the user chooses to see the list of pages relevant to the term. Figure 7 reports the list of textbook pages ranked relevant to the term "cluster profile." As this term was not associated to any pages by the author, the list of linked pages was built using the pages linked to the term "cluster representative." It is worth noting that the top-ranked page is one of those that were pointed out as relevant by the textbook author in the subject index, but the

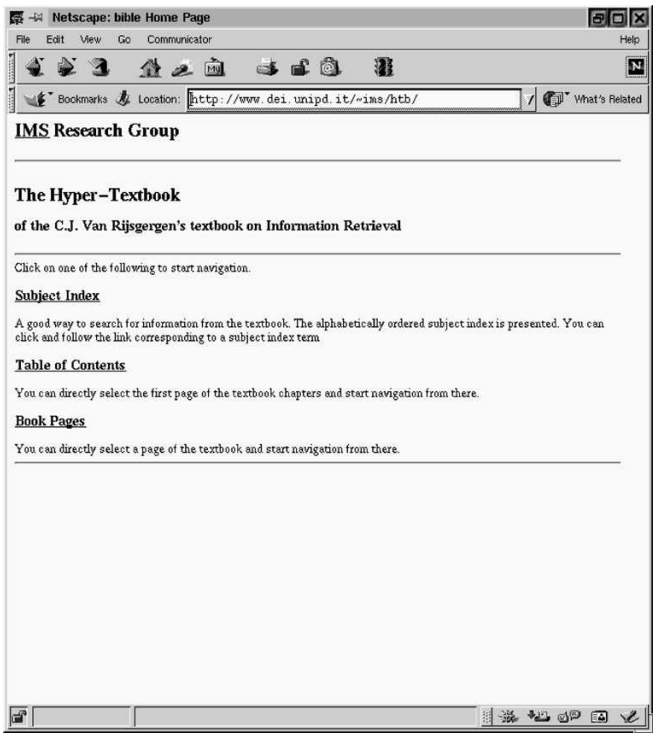

Fig. 5. Homepage of the hypertextbook

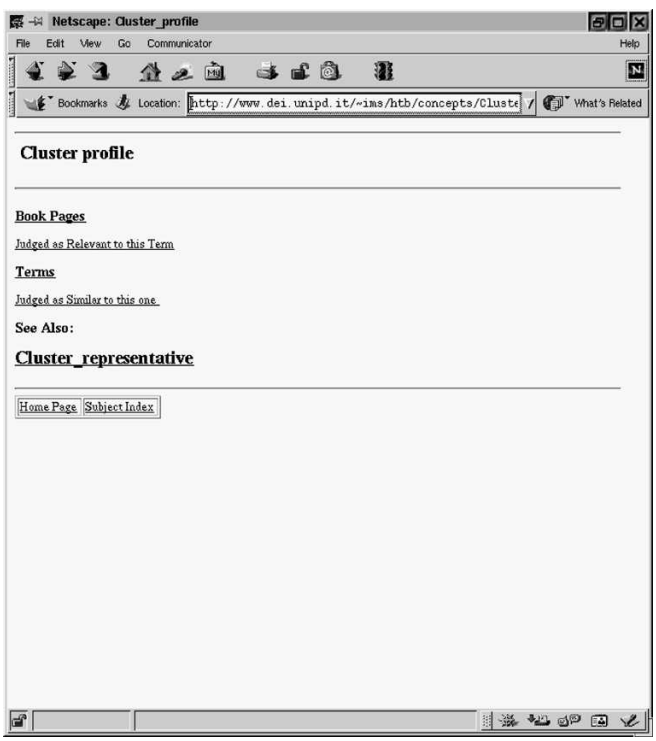

Fig. 6. Page of a term of the subject index

[5] `http://ims.dei.unipd.it/htb/`

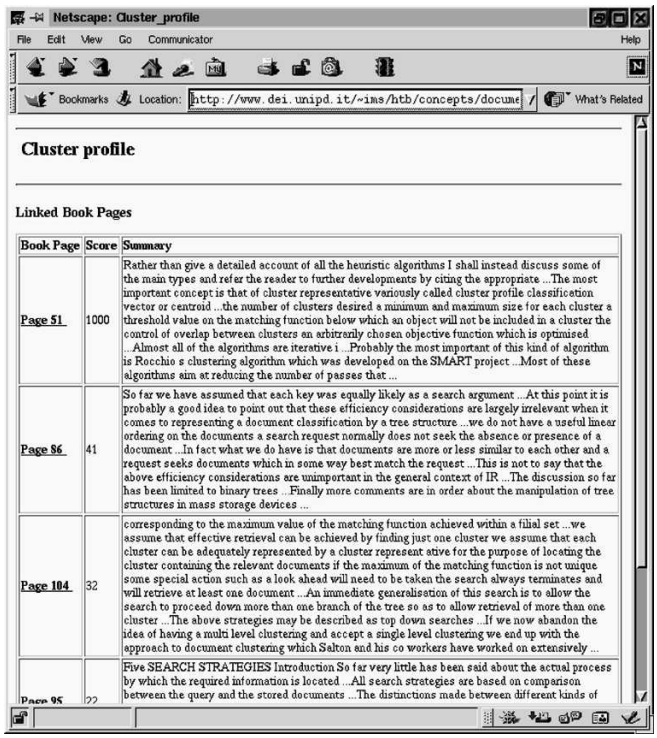

Fig. 7. List of textbook pages relevant to the term "cluster profile"

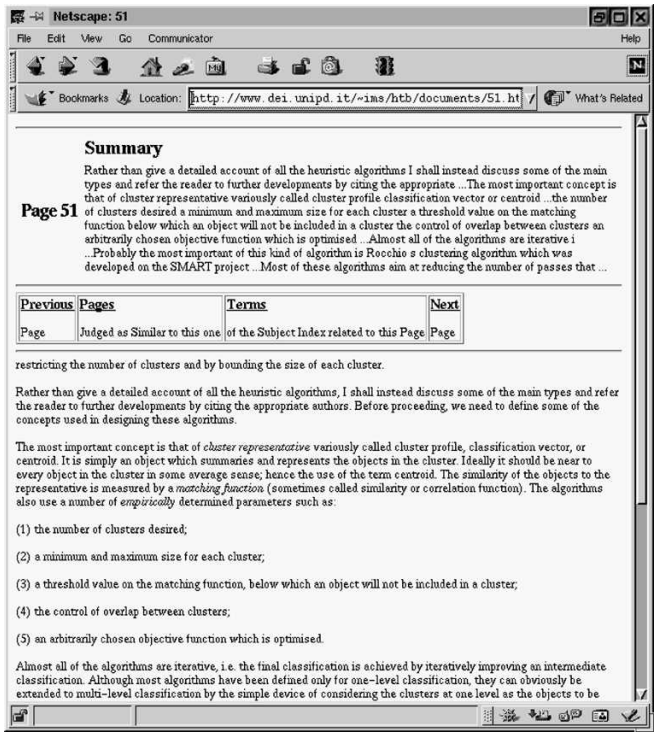

Fig. 8. A page of the textbook relevant to the term "cluster profile"

following one, for example, was not. This means that the technique used to infer semantic links can preserve the author's judgments of relevance, as well as identify some additional relevant pages [19]. Links are ranked according to the normalized term weight within the textbook page.

Figure 8 contains page 51 of the textbook, a P page. The user arrived at that page just by clicking on the page number. The hypertextbook displays, for each page, the page number, a summary, and the full text of the page. Both at the top and at the bottom of each page are located some buttons that link the current page to the next page, to the previous page, to terms describing the content of the page, and to similar pages. Links to the next and previous pages enable a linear reading of the textbook. Links to terms describing the content of the page or to similar pages (PT and PP list pages) enable a nonlinear navigation of the hypertextbook according to the EXPLICIT model. Pages can be saved locally or printed at any time, and all the functionalities of the Web browser are available, such as, for example, bookmarking or "find in page."

Figure 9 shows the list of similar pages linked to the current page (a PP node link). Links are ranked according to a normalized page similarity measure based on the distribution of terms in the pages.

The user can also see the terms that are similar to a selected term, as depicted in Fig. 10, which reports the TT list node of terms similar to the term "cluster profile." Links are ranked according to a normalized term similarity measure.

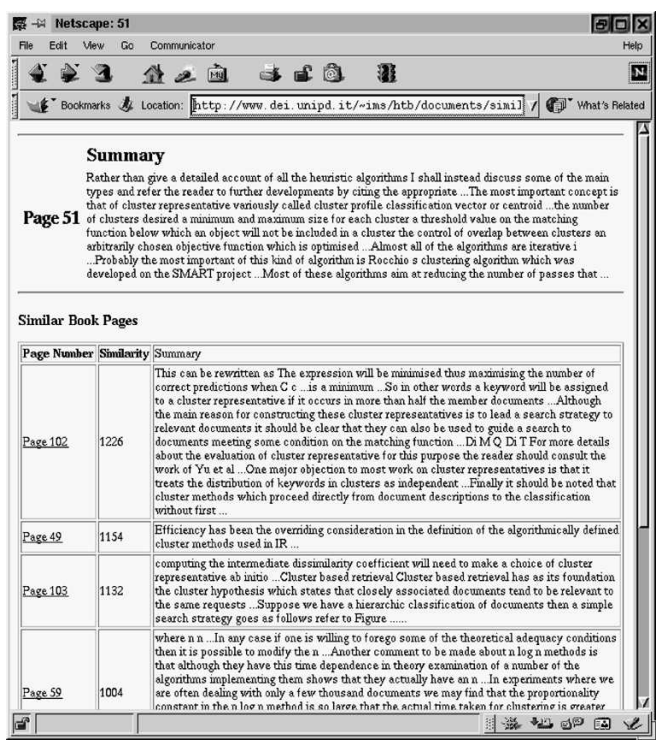

Fig. 9. List of pages similar to p. 51

4.3 Conclusions from the HyperTextBook project

A detailed presentation of the results of the evaluation of the usability of hypertextbooks produced in the context

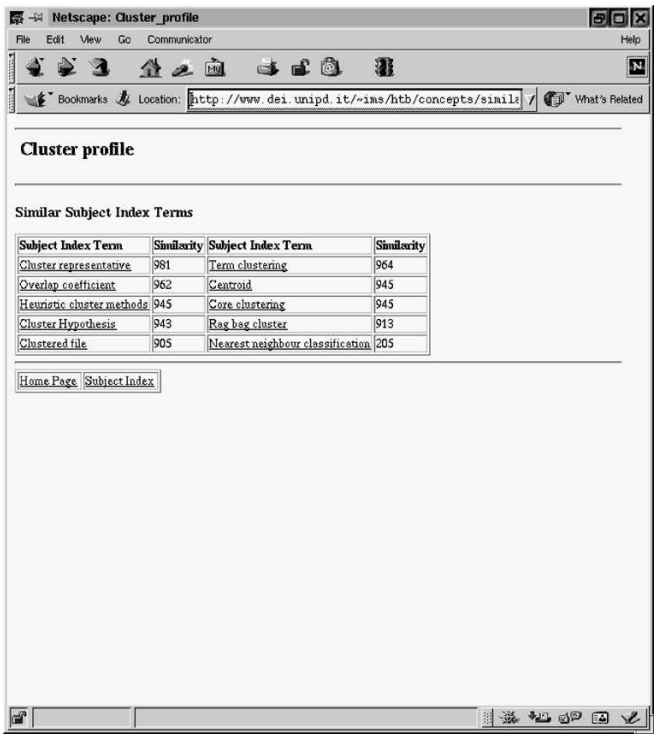

Fig. 10. List of terms similar to the term "cluster profile"

of the HyperTextBook project is reported in [20]. In that case study, a user task-based evaluation scheme was developed and a comparative evaluation of a hypertextbook against the printed form of the textbook was performed. An additional evaluation comparing the hypertextbook against an "online" form of the textbook was also carried out, but the results were inconclusive and difficult to interpret. At that stage we decided to focus our evaluation on the usability of the final product, the hypertextbook, rather than on the efficiency of the automatic authoring process.

In the hypertextbook evaluation we made the following experimental hypothesis: the hypertextbook is a more effective tool for seeking information than the original printed form of the textbook. Once the research hypothesis was established, a rigorous experimental design was adopted to test it [41]. In our hypothesis the independent variable had two levels: the presence (hypertextbook) or absence (printed textbook) of specific hypertext functionalities. Participants in the evaluation had to perform a number of tasks with and without the assistance of hypertext links and hypertextbook additional features. The dependent variable to be measured was the effectiveness of the participants in completing these tasks. Our primary objective was to prove or disprove that any variation of dependent variable, the effectiveness, between participants was to be attributed only to the change in the level of the independent variable, i.e., to the additional functionalities enabled by the hypertextbook.

Since the independent variable had two levels, two groups of the same size of participants was employed. One group performed some tasks using the hypertextbook, while the other group performed the same tasks using the printed form of the textbook. In our case, due to restrictions on both time and resources, each group consisted of ten participants. Since the evaluation of the hypertextbook was substantially user-centered, the performance of the participants under such task-based conditions could be considered an indication of the effectiveness of the hypertextbook.

The participants were postgraduate students doing a conversion course in computer science. The textbook used was the same one presented in the example reported in this section. All participants had the same experience in the subject of the textbook. Each participant was presented with seven questions, whose answers required finding information in the textbook within a time limit. Questions were of two categories: open-ended and close-ended. The answers to the questions were given by the author of the textbook. Unbeknownst to the participants, some of the questions were designed to be well suited for answering using the conventional, printed textbook, while the rest exploited, partially or fully, the functionality of the hypertextbook.

This evaluation procedure was supposed to mimic a student using the book for some homework or examination. Apart from the participants' time for completing the assigned task and the accuracy of the answers, the experimenter took note of the steps taken by the participants in order to accomplish the task of answering each question. This was done in order to reveal participants' strategies in completing the assigned task. At the end, each participant was asked to complete a questionnaire.

The outcome of the analysis of the experimental data was that, in some experimental tasks, the hypertextbook was superior in matters of accuracy and rapidity in satisfying users' information needs in comparison to the other form of the textbook, while for other tasks subjects using the printed form performed better.

Some of the most interesting conclusions pertaining to the design and implementation of the hypertextbook were the following.

- The accuracy rate in answering the close-ended questions was greater for participants using the hypertextbook than that for participants using the textbook. On the other hand, for the open-ended questions, the accuracy rate of participants using the textbook was marginally higher than the accuracy rate of participants using the hypertextbook. This partially proved the greater usefulness of the hypertextbook for fact finding.
- Participants using the hypertextbook were much faster in finding answers than participants using the textbook. This was particularly true for open-ended questions, where the increase in speed was about 10–20%. There was no significant difference in speed between the two groups of participants for close-

ended questions. An in-depth analysis of these results showed that participants using the hypertextbook were faster in finding the required pages for both open-ended and close-ended questions, but the display characteristics of the hypertextbook made spotting the sought-after information in the page slower.
- The analysis of the subjective opinions gathered through a postexperimental questionnaire submitted to the experimental subjects suggested that a better formatting of the text in each page would improve readability. The highlighting of terms when a text page was reached from a term page was considered particularly important.
- The analysis of the subjective opinions also revealed that most of the subjects did not appreciate the functionality of the summaries appearing at the top of each page and they would have preferred the pages to look more like "real" book pages.
- Most users were also of the opinion that there was a need for the integration of a global search function in the hypertextbook to enable users to locate a section in the book pertinent to an information need without browsing. We had anticipated this result, but we did not make available such functionality on the hypertextbook in order to make a fairer comparison with the textbook.
- Browsing and searching the hypertextbook was done with the aid of a common Web browser. However, such a browser was not sophisticated enough to enable the implementation of an effective interface to access the hypertextbook and fully exploit its functions. Therefore, the need for the design and implementation of a better interface, reflecting the underlying model of the hypertextbook, should be catered to in the future.

The above conclusions highlight the advantages of giving the user powerful functionalities for searching and browsing the content of a textbook to find the sought-after information. These functionalities have proved to enhance the speed and accuracy of user task solving, despite the many limitations of the evaluation. However, they also highlight how important appearance is for e-books and how appearance can also improve the usability of searching and browsing functionalities. In addition, these results showed a number of areas that future work should concentrate on in order to improve the usability and the effectiveness of the hypertextbook. These directions pertain to both redesign of the hypertextbook and its reevaluation, following a formative design evaluation approach.

A new version of HyperTextBook methodology addressing many of the functional limitations highlighted in the evaluation has been produced, and a new version of the same hypertextbook used in this evaluation has been produced. This new version has been included in the CD-ROM accompanying the book *Finding Out About: A Cognitive Perspective on Search Engine Technology and the WWW*, by R. Belew, published in 2000 by Cambridge University Press [11].

We are currently exploring how to overcome some of the limitations highlighted above that pertain to the visual appearance of hypertextbook pages.

5 Two electronic books, one lesson?

The previous brief description of the two projects and their conclusions make clear that the Visual Book and the HyperTextBook have many similarities when it comes to their approach to electronic publishing. They both take the readers as the center of the whole system and pay particular attention to their real needs. Both systems are the result of extensive studies on what sort of electronic publications can be of real use to a specific category of users (scientists for the Visual Book, students for the HyperTextBook) and what additional features they may require or profit from in the electronic version. A difference between the two projects is that while the Visual Book concentrates on the presentation issues, by looking for new presentation paradigms suitable for the electronic medium in an attempt to improve the visual quality of the electronic version to be at least comparable with the paper one, the HyperTextBook focuses instead on providing additional features related to the electronic medium in order to make the electronic version more flexible and powerful than the paper one.

The results of the Visual Book evaluation show the importance of presentation issues in electronic publishing, that appearance plays a crucial role in increasing the global value of information, and have also pointed out the need for advanced browsing and searching features to be part of e-books. On the other hand, the results of the evaluation of the HyperTextBook have shown that advanced browsing and searching features need to be complemented with a presentation of the book that is as familiar as possible to the user since these new features add considerable complexity to its use. In effect, the results of both projects have shown that the subject index is the preferred search and browsing tool for users of a hypertextbook and that such a tool can be very effective if made easy to use and integrated into the visual appearance of the book. In order to make a subject index an effective and advanced means of searching for information, both appearance and automatic generation have to be addressed.

Of course, automatic hypertext construction, which has been addressed by HyperTextBook, is unnecessary until electronic books are produced without any hyperlinks that permit the end user to cross refer across one book or diverse books. But, if you would like to provide the end user with hypertextual capabilities to search a huge digital library of electronic books, the automatic generation of hyperlinks becomes a mandatory option since manual generation would be infeasible. Thus, one

sense of the convergence is from a single Visual Book toward a Visual Book Digital Library through automatic link generation, perhaps among distinct e-books, as envisaged in [21]. The other direction of convergence is from a hypertextbook toward a digital library of electronic books that, while retaining hypertextual capabilities, are enhanced by a more familiar appearance, like that of the Visual Book system. Indeed, one outcome of the HyperTextBook evaluation is that overall appearance had to be improved.

6 Combining appearance and functionalities: guidelines for the design of electronic books for information access

A set of guidelines for the production of "good" e-books has already been defined by [35]. These have been expanded to include the findings of the HyperTextBook project with the aim of providing a more generic guidance without losing effectiveness. These guidelines will be discussed in terms of appearance (central to the Visual Book project) and functionalities (as explored by the HyperTextBook project) in the next two sections. Our aim is to provide guidance for the designers of electronic books in general and of electronic textbooks in particular since this was the area where our efforts converged.

6.1 Appearance

The design of the appearance of a good e-book follows the guidelines discussed above in Sect. 3, with strong emphasis on the importance of the visual impact of the final product. This is in line with [42, 43] on writing for the Web, and in particular with their findings on how the presence of "scannable" (i.e., easy-to-read) text enhances usability.

Great importance is given to the presentation of title and page number as they are crucial clues in guiding the reader while consulting the book, as shown by the Visual Book experience [35]. In particular, style, font, size, and position on the screen/window/page are considered to be the results of a proper application of electronic typography. The same attention is paid to the choice of labels and titles to be immediate, effective, and unambiguous for the reader. An electronic page is a container of information to be used or consulted on a screen and as such should be presented in a clear, appealing, easy-to-scan fashion. The definition of the hypertextbook page, as a visual unit of information, takes into account many different factors, among them the legibility and cognitive issues, which are related to the amount of space available for presentation, the necessity of maintaining as much as possible the same appearance as in the original paper version (as it is already part of the cognitive background of readers of this very same textbook), the possibility of presenting additional information in order to increase the value of the one contained in the original paper page, and the presence of various menus to provide navigation functionalities while browsing the electronic book. Information in electronic books can be of any sort, from traditional text to animation, video, and images, which is why the concept of legibility has to be interpreted in a more flexible way such as scannability. A printable version should be available, too, for users to take with them.

When designing an electronic page few essential guidelines should be followed in order to provide a pleasant and familiar appearance. Some general guidelines are the following.

- *Choose a readable font.* Fonts should be large enough to read comfortably for long periods of time. If possible, readers would like to choose a font style and size to suit their individual preferences, thereby satisfying the needs of those with perfect vision and those with poor vision or reading difficulties.
- *Page numbers.* Page numbers are a familiar clue for the reader, and if at all possible they should be considered as an essential component of the electronic page.

Layout provides the visual clue for the reader to follow while reading the book, and as such it plays a crucial role when it comes to overall usability. In fact, appearance plays a crucial role by making the reader feel at ease, achieving what users of the HyperTextBook called "real book" pages. Specific guidelines concerning the layout are the following.

- *Design typographical aspects carefully.* Readers expect typographical sophistication, and pagination has to be designed carefully to enhance readability.
- *Provide content clues.* Section headings, keywords, or abstracts under chapter headings in the main table of contents will inform the reader's understanding of the contents of each chapter at a glance.
- *Provide orientation clues.* Readers gain a sense of their place in a printed book via page numbers and by comparing the thickness and weight of the pages read against the thickness and weight of the pages left to read. It is important for this "sense of place" also to be present in the electronic medium, via chapter and section headings or navigation bars, which highlight the current position. These indications of a reader's progress through the book should be accurate and visible.
- *Use nontext items with care.* Readers expect images, diagrams, and formulae to be included and to look as visually sophisticated as they do on the printed page.
- *Special pages.* Special pages such as the subject index and the table of contents require particular attention as highlighted by both the Visual Book and the HyperTextBook experiences.

Finally, the results of the evaluation of the HyperTextBook project are being taken into account in order to produce an effective interface able to overcome the limitation of the Internet paradigm, which is mainly scroll

oriented instead of page oriented. The book metaphor has to be reconsidered and eventually extended to consider the new advanced features introduced in the HyperText-Book.

6.2 Functionalities

All the functionalities that have proved so valuable with paper books should be kept (e.g., table of contents or index), but they could be redesigned to take advantage of the new environment; this is where the designer is encouraged to use the potentiality of a dynamic environment at its best. New features such as search tools and hyperlinks can be added, too, in order to help the reader using (e.g., consulting, searching, personalizing, interpreting, studying, skimming, etc.) the electronic book. In particular, our experience shows that the following guidelines can help in designing effective e-books.

- *Include a table of contents.* Tables of contents are an essential feature in both print and electronic media and are used by readers to skim the contents of an unfamiliar book to gain an idea of what can be found inside. They also provide the reader with a sense of structure, which can easily be lost in the electronic medium, and can be an important navigation tool where hypertext is used to link from the table of contents to individual chapters.
- *Include an index.* The subject index proved to be a good way to search for information in both studies. But while the subject index was kept very close to its paper counterpart in the Visual Book project, the HyperText-Book study has produced an enhanced, more efficient, and more flexible version of the traditional index by using IR techniques. Terms in the subject index are still presented alphabetically, the way the user is used to. The user can click and follow the link corresponding to a subject index term. The result of clicking on any term is a page with the term as title and a list of links to book pages or a list of links to subject index terms. The score representing the importance of the term in a page or its similarity with other terms in these list pages should not be presented to the user since it only provides a distraction. Moreover, only a limited number of links should be presented in link pages.
- *Provide a search tool.* Search tools should not replace tables of contents or indexes and should be intelligent enough to simulate and enhance the way readers search in paper books. The need for the integration of a search function in the hypertextbook becomes obvious from the analysis of the experimental results. The power of searching with a universal, or limited, scope would result in faster searches and, perhaps, more accurate answers since the time to find a section in a book pertinent to an information need would be minimized. Therefore, more time could be spent on examining the actual piece of information rather than locating it in the textbook.
- *Use hypertext to enhance navigation and facilitate cross-referencing.* Cross-referencing between the pages of a book, between the main text and table of contents, index, footnotes, glossary, or references and between two or more books is considered an important property of the printed medium.
- *Provide bookmarking and annotating functions.* By applying these simple guidelines the overall usability of electronic textbooks will increase – as will reader satisfaction.

There are similarities with and differences from other research works. We wish to point out some aspects characterizing our work and distinguishing it from all others.

Research findings on Web site design found that users rarely read pages sequentially. Rather, they scan them, picking out individual words and sentences. Therefore, access to the information stored in Web pages should ease this kind of process. To this end, it was suggested to make Web pages concise and "scannable." Nielsen et al. suggested highlighting keywords, writing meaningful subheadings, using bullet lists, putting one idea per paragraph, placing emphasis on the conclusion. Moreover, it is important that Web pages be credible, using high-quality graphics, good writing, and outbound hypertext links [42, 43].

In our work we are concerned with textbooks, which are available to the user before being put on the Web, since they are often already published in paper format. In our experience we found that readers of textbooks expect them to be published in an electronic format that adheres to the content, layout, and appearance of the paper format. Moreover, readers of textbooks perform access tasks that are very different from scanning – they refer to searched information and often read as if the content were on paper. Thus while scannability, i.e., legibility, can be enhanced by proper design and layout, as the findings of the Visual Book project suggest, it is not possible to make textbooks more concise than the paper format. As an alternative, the introduction of advanced search functionalities that make information access more effective is suggested (as if the book were indeed concise and scannable). This is why we propose to integrate automatic hypertext generation and visual book design.

Despite the differences in tasks and content between working with electronic books and Web sites, we found some similar results; high-quality graphics, careful typographic design, and content/orientation clue provision are as important with electronic books as with Web pages. However, the fact that previous research on Web site design has led to similar guidelines only means that there are some similar directions to follow in designing both Web sites and electronic books. We cannot argue that the conclusions reached for Web site design apply directly to the design of electronic books just because Web sites and e-books exploit similar dissemination means. In-

deed, we found that there are some specific e-book guidelines that do not apply to Web sites directly, e.g., sophisticated fonts and vice versa, e.g., conciseness, because they play different roles for information access.

Both projects found that subject index and table of contents are two important means for effective information access because they organize the content in a way that is already known by the user, i.e., in the way employed by paper books. This suggests the veracity of the experimental hypothesis that a conceptual model for electronic books is fundamental to permitting the end user to find the desired information.

The results from the HTB project suggest that the automatic generation of hyperlinks involving subject indexes and tables of contents is a feasible solution. This is an important finding since it enables us to show that there exists a convergence of VB and HTB, i.e., that it is possible to implement automatic link generation to VB in order to rapidly produce large digital libraries of searchable and browsable electronic books that still retain their paper-book-like appearance.

In conclusion, we believe that there is a need for better e-books and that this can be achieved only by carefully evaluating existing e-books, taking advantage of their strengths, and overcoming their weaknesses. We have started doing so with the Visual Book project and the HyperTextBook project. The aim of the joint guidelines just discussed is to combine the strengths of the projects from which they originate. Our expectation is that this integration will provide very useful insights into the design and implementation of next-generation e-books.

7 Conclusions

In this paper we presented and compared the results of two research projects aiming at producing e-books that emphasized two different aspects of the book metaphor: appearance and functionalities. We argue that good e-books should have both, in particular for e-books where information access is very important, like, for example, "e-textbooks." We think that one of the reasons of the scant success of current e-books can often be found in the lack of attention to pragmatics. Based on our experience, we have suggested ways of combining appearance and functionalities to take advantage of their respective strengths. We have presented a set of guidelines for the design of e-books that retain and enhance both. Guidelines have been left pretty flexible on purpose in order to be easily tailored to different e-books and their readers. We hope that designers of e-books will find it easy to follow them and that the result will be a new generation of better e-books able to provide extra value and cater to different readers. We also believe that, while e-books will never substitute paper books in full, once they are properly designed, they could well play a crucial role in education and training.

References

1. Agosti M, Colotti R, Gradenigo G (1991) A two-level hypertext retrieval model for legal data. In: Proceedings of ACM SIGIR, Chicago, October 1991, pp 316–325
2. Agosti M, Crestani F (1993) A methodology for the automatic construction of a Hypertext for Information Retrieval. In: Proceedings of the ACM symposium on applied computing, Indianapolis, IN, February 1993, pp 745–753
3. Agosti M, Crestani F, Melucci M (1995) Automatic authoring and construction of hypertext for Information Retrieval. ACM Multimedia Syst 3(1):15–24
4. Agosti M, Crestani F, Melucci M (1998) Electronic publishing, storage, dissemination and retrieval of a scientific journal through the Web. In: Proceedings of the IEEE ADL'98 conference, Santa Barbara, CA, pp 137–146
5. Agosti M, Smeaton A (eds) (1996) Information Retrieval and Hypertext. Kluwer, Boston
6. Allan J (1997) Building hypertext using Information Retrieval. Inf Process Manage 33(2):145–159
7. Arms WY, Michalak TJ (1990) Carnegie Mellon University. In: Arms C (ed) Campus strategies for libraries and electronic information. Digital Press, Bedford, MA
8. Barker P (1991) Interactive electronic books. Interact Multimedia 2(1):11–28
9. Barker P (1996) Living books and dynamic electronic libraries. Electron Libr 14(6):491–501
10. Barker P, Richardson S, Benest ID (1994) Human-computer interface design for electronic books. In: Proceedings of Online Information 94, 16th international online information meeting, London, pp 213–229
11. Belew R (2000) Finding out about: a cognitive perspective on search engines technology and the WWW. Cambridge University Press, Cambridge, UK
12. Benest ID, Morgan G, Smithurst MD (1987) A humanised interface to an electronic library. In: Proceedings of INTERACT'87, pp 905–910
13. Bonime A (1992) Sony's Data Discman: a look at these new portable information machines and what they mean for CD-ROM developers. CD-ROM Professional, September, pp 55–65
14. Botafogo RA, Rivlin E, Shneiderman B (1992) Structural analysis of hypertextx: identifying hierarchies and useful metrics. ACM Trans Inf Syst 10(2):142–180
15. Burrill VA, Ogden JA (1989) VORTEXT: The hard-backed screen. Electron Publish 2(3):143–156
16. Callan JP (1994) Passage-level evidence in document retrieval. In: Proceedings of ACM SIGIR, Dublin, Ireland, July 1994, pp 302–310
17. Carroll JM, Mack RL, Kellog WA (1988) Interface metaphors and user interface design. In: Helander M (ed) Handbook of human computer interaction. Elsevier, Amsterdam, pp 67–85
18. Catenazzi N, Gibb F (1995) The publishing process: the hyper-book approach. J Inf Sci 21(3):161–172
19. Crestani F, Melucci M (1998) A case study of automatic authoring: from a textbook to a hyper-textbook. Data Knowl Eng 27(1):1–30
20. Crestani F, Ntioudis S (2002) User centred evaluation of an automatically constructed hyper-textbook. J Educat Multimedia Hypermedia 11(1):3–19
21. Crivellari F, Melucci M (2001) Automatic hyper-textbook web generation. In: Proceedings of the ACM symposium on applied computing (SAC), Las Vegas, pp 253–256
22. Dove R (1996) Lady freedome among us. `http://www.lib.virginia.edu/etext/fourmill/DovLady.html` [last visited December 2004]
23. Egan D, Remde JR, Gomez LM, Landauer TK, Eberhardt J, Lochbaum CC (1989) Formative design-evaluation of SuperBook. ACM Trans Inf Syst 7(1):30–57
24. Egan DE, Lesk ME, Ketchum RD, Lochbaum CC, Remde JR, Littman M, Landauer TK (1991) Hypertext for the electronic library? CORE sample results. In: Proceedings of ACM Hypertext'91, San Antonio, TX, December 1991, pp 299–312
25. Egan DE, Remde JR, Landauer TK, Lochbaum CC, Gomez LM (1989) Acquiring information in books and superbooks. Mach Mediated Learn 3:259–277

26. Feldman T (1990) The emergence of the electronic book. British National Bibliography Research, Publishers Association [Fund report 46], London, UK
27. Frakes WR, Baeza-Yates R (eds) (1992) Information Retrieval: data structures and algorithms. Prentice Hall, Englewood Cliffs, NJ
28. Frisse ME (1988) Searching for information in a medical handbook. Commun ACM 31(7):880–886
29. Furuta R, Plaisant C, Shneiderman B (1989) A spectrum of automatic hypertext constructions. Hypermedia 1(2):179–195
30. Gentner D, Nielsen J (1996) The anti-Mac interface. Commun ACM 39(8):70–82
31. Hansen WJ, Haas C (1988) Reading and writing with computers: a framework for explaining differences in performance. Commun ACM 37(9):1080–1089
32. Hayter M (1996) Lectrice. `http://www.research.digital.com/SRC/virtualpaper/` [last visited December 2004]
33. Hersh WR, Molnar A (1995) Towards new measures of Information Retrieval evaluation. In: Proceedings of ACM SIGIR, Seattle, July 1995, pp 164–170
34. Joyce M (1987) Afternoon, a story. Eastgate Systems, Cambridge, MA
35. Landoni M (1997) The Visual Book system: a study of the use of the visual rethoric in the design of electronic books. PhD Thesis, Department of Information Science, University of Strathclyde, Glasgow, Scotland, UK, May 1997
36. Landoni M, Catenazzi N, Gibb F (1993) Hyper-books and visual-books in an electronic library. Electron Libr 11(3):175–186
37. Landoni M, Wilson R, Gibb F (2000) From the visual book to the web book: the importance of good design. In: Proceedings of ECDL2000, 4th European conference on research and advanced technology for digital libraries, Lisbon, Portugal, September 2000, pp 305–314
38. Landoni M, Wilson R, Gibb F (2001) Looking for guidelines for the production of electronic textbooks. Online Inf Rev 25(3):181–195
39. Lehmann-Haupt C (1998) Creating "the last book" to hold all the others. New York Times, 8 April 1998
40. Manes S (1999) Gutenberg need not worry – yet. Forbes, 8 February 1999
41. Miller S (1984) Experimental design and statistics, 2nd edn. Routledge, London, UK
42. Morkes J, Nielsen J (1997) Concise, scannable, and objective: how to write for the web. `http://www.useit.com/papers/webwriting/writing.html` [last visited December 2004]
43. Morkes J, Nielsen J (1998) Applying writing guidelines to web pages concise, scannable, and objective: how to write for the Web. `http://www.useit.com/papers/webwriting/rewriting.html` [last visited December 2004]
44. Nielsen J (2003) PDF: Unfit for human consumption. `http://www.useit.com/alertbox/20030714.html` [last visited December 2004]
45. Rothkopf EZ (1971) Incidental memory for location of information in text. J Verbal Learn Verbal Behav 10(6): 608–613
46. Rowley DE, Rhoades DG (1992) The cognitive jogthrough: a fast-paced user interface evaluation procedure. Proceedings of the SIGCHI conference on human factors in computing systems, Monterey, CA, pp 389–395
47. Salton G, Allan J (1993) Selective text utilization and text traversal. In: Proceedings of ACM Hypertext'93
48. Salton G, Allan J, Buckley C, Singhal A (1996) Automatic analysis, theme generation, and summarization of machine-readable texts. In: Agosti M, Smeaton AF (eds) Information retrieval and hypertext. Kluwer, Dordrecht, pp 51–73
49. Schilit BN, Price MN, Golovchinsky G, Tanaka K, Marshall CC (1999) As we may read: the reading appliance revolution. IEEE Comput 32(1):65–73
50. Shin D, Nam S, Kim M (1997) Hypertext construction using statistical and semantic similarity. In: Proceedings of ACM Digital Libraries, Philadelphia, July 1997, pp 57–63
51. Southall R (1989) Interfaces between the designer and the document. In: Andre J, Furuta R, Quint V (eds) Structured documents. Cambridge University Press, Cambridge, UK, pp 119–131
52. Wilson R, Landoni M, Gibb F (2002) A user-centred approach to ebook design. Electron Libr 20(4):322–330
53. van Rijsbergen CJ (1979) Information Retrieval, 2nd edn. Butterworths, London, UK