Electronic Journal Provision using Acrobat

Judith Wusteman
Computing Laboratory,
University of Kent,
Canterbury,
Kent,
CT2 7NZ.
email jcw@uk.ac.ukc

October 2, 1994

Abstract

This paper discusses issues of electronic journal readability. It identifies some of the techniques that human beings have evolved for reading and skimming paper-based journals. An attempt is made to identify those manual techniques which could be constructively replicated in electronic form. Consideration is also given to the additional functionality provided by online journals and how this can be employed to improve on the methods currently used in their paper equivalents.

The paper's conclusions are based on an evaluation of online journal articles made available on Adobe Acrobat software as part of the CAJUN (CD-ROM Acrobat Journals Using Networks) project [22]. The Acrobat suite is described and its potential suitability as a vehicle for electronic journals is discussed.

1 Introduction

According to publishers at the 1993 Book Fair, electronic journals have now arrived [16].

The concept of the online journal has been in existence for some time; indeed, it is claimed that there are already one hundred and thirty "strictly electronic journals" in circulation [16]. Despite this proliferation, the future form and organisation of such publications is still wide open for debate.

This paper considers some of the design issues affecting the use and readability of electronic journals. Its conclusions are based on an evaluation of online journal articles made available by the CAJUN (CD-ROM Acrobat Journals Using Networks) project [22][1].

Section 2 provides a brief overview of the concept of electronic journals and their development. This is followed by a description of the CAJUN project and the Adobe Acrobat system on which it is based. Sections 5 to 8 consider some of the issues raised in attempting to provide online facilities for skimming and studying journals and their constituent articles.

2 Electronic Journals

Concern that the amount of knowledge in existence increases faster than our ability to organise and disseminate it effectively is not new. Journals emerged over two hundred years ago as a response to this concern, superseding the monograph as the most efficient method of disseminating information and ideas as widely and as quickly as possible [19]. In more modern times, the high cost and time delay involved in journal production has led to the emergence of the electronic journal.

[1] Based at the University of Nottingham and jointly funded by John Wiley & Sons Ltd. and Chapman and Hall Ltd.

Opinion differs as to what constitutes, or should constitute, an electronic journal. A journal may be defined as "any collection of learned articles which has been accepted via the peer review process for publication as part of a series" [23]. An electronic journal could then be defined as a journal in which the end product is available electronically, whether this is over the network or via storage devices such as the CD-ROM. However, this discounts many of the products currently describing themselves as electronic journals, not least because the majority do not employ the peer review process.

It should also be remembered that, whether the end product is provided in electronic form or not, electronic communication already plays a major part in the production of some paper-based journals. The Wiley journal *"Electronic Publishing – Origination, Dissemination and Design" (EP-odd)*, for example, relies on a "wide variety of support from electronic technologies in the pre-publication phase" [4]. To this end, authors are strongly encouraged to make final submission of accepted papers in one of several approved machine-readable forms.

The constituents of the "collection of learned articles", referred to in the definition of a conventional journal, may also differ for an electronic journal. It could be argued that electronic dissemination does away with the need to package articles into journal issues. Articles can be published as soon they are accepted by the journal's editorial board, thus minimising publication delays. After over two hundred years, is it possible that the journal will be unbundled again and the monograph will reappear?

As to whether the journal itself is an appropriate metaphor to carry over to electronic form, Pullinger [20] points out that a journal is more than a set of articles. Not only does it contain non-article-based information, such as indexes, it also provides a focus for the readership community and a forum for discussion. Indeed, some publishers have introduced bulletin boards for particular journal titles in order to encourage informal discussion among their readership. Readers need to be provided with some method of knowing where and how to look for the latest details of work in a particular field. If papers are not to be presented as predetermined collections, this "awareness function" needs to be provided in some other way. Treating a journal simply as a set of documents, results in the loss of richness which was introduced in the transition from monograph to journal.

Whether the concept of an issue will or should survive is less clear-cut. An issue based on the dates of article submissions may lose credibility. However, the concept of special issues on particular subjects may still be valid. With the increased sophistication of information retrieval methods, personalised journal issues, containing articles of particular interest to an individual subscriber, become a possibility. However, as Okerson [19] points out, for the possibility to become a reality, a fuller and more consistent scheme for article classification is required; current article descriptions could not adequately support complex information retrieval.

Initial experiments relating to electronic journals were hampered by the lack of comprehensive networking facilities such as email. Some of the first studies resorted to the use of computer teleconferencing systems [21, 7]. The BLEND project [7], for example, used the NOTEPAD teleconferencing system to disseminate its journal on computer human factors. The aim of the project was to consider not only the potential of the electronic journal but also the use of the computer for the facilitation of communication in general. As a result, it developed teleconferencing facilities for pre-publication drafts, a newsletter and email service in addition to those facilities directly required for the development of the journal. By the end of the research period, which covered the years 1980 to 1984, the technology used in the project had been superseded. In particular, the prevalence of wide area networks and email meant that the teleconferencing approach was no longer relevant.

A further problem experienced by many early attempts, including the BLEND project, was that of credibility; difficulty was experienced in attracting sufficient high quality articles. One of the main functions of an academic journal is the recognition it provides for an author's work. Paper journals have a more established reputation and are therefore seen to carry more authority. Compounding this problem is the fact that, as stated above, very few of the electronic journal titles currently available are formally refereed, although the number is increasing. This problem of credibility has not gone away; it appears that the technical issues involved in electronic journal proliferation may not be as important as political and social issues.

Despite these perceived problems, further advances in networking have fuelled a flurry of activity, both on the research side and in the emergence of numerous titles. Electronic journals, covering topics from fine art to computer science, have appeared and the Internet is alive with discussion on such diverse aspects as the electronic journal's potential readership and whether it should be based on commercial or non-commercial principles. Several current major research programmes involve consideration of electronic journals. They include the US National Science Foundation's Digital Library Initiative [3] the European Communities' Libraries Programme [18] and the UK Joint Funding Councils' Libraries Review Group proposals [2].

Many design and management issues remain to be resolved. However, one conclusion that can be drawn from the experience so far is that, to be successful, electronic journals will have to add value to the journal concept beyond that currently offered by the paper-based journal.

3 The CAJUN Project

The CAJUN project [22] began in January 1993. It aims to evaluate the technical feasibility of disseminating journal articles on CD-ROM media and over electronic networks. To this end, the archives of several journals have been converted to Portable Document Form (PDF). This is the page description language used by Acrobat and is described in Section 4. Experiments in disseminating and reading these papers have been performed. The journals involved in the project include Wiley's *EP-odd*. Sample papers from this journal will be used to illustrate the ideas discussed in this paper.

The CAJUN project is on-going and further work is intended in areas such as network dissemination. Smith *et al* [22] describe the two categories of network dissemination as "push" and "pull". In the former method, information is transferred to subscribers by the publisher. The latter refers to the more flexible approach whereby subscribers are able to access the information of their choice.

Initially, the CAJUN project concentrated on the "push" approach to dissemination. Previously published sample papers in PDF format were sent to chosen test sites, both via CD-ROM and using FTP file transfer methods. The CD-ROM contained seven *EP-odd* articles and five articles from the Chapman and Hall journal *Optical and Quantum Electronics*. To test FTP dissemination, the same seven *EP-odd* articles were packed together and sent over the network. The test sites were provided with tools to unpack these collections of papers.

Experiments are also being carried out into the use of the "pull" approach. The sample *EP-odd* documents have been made available via network information tools such as Gopher, World Wide Web and anonymous file transfer protocol (ftp) [9][2].

4 Adobe Acrobat

Many of the problems computer users experience in sharing electronic documents are the result of the plethora of incompatible hardware platforms, operating systems and application software used in document production [12]. As a result of these incompatibilities, it is often impossible to share any document more complex than a plain ASCII text file. This is a particular problem in the provision of electronic journals; the lowest common denominator format must often be chosen if successful dissemination to the journal's entire readership is to be ensured.

The Acrobat suite of software, developed by Adobe Systems Incorporated, aims to facilitate document transfer between different platforms. As already mentioned in Section 3, Acrobat is based on a platform-independent page description language called Portable Document Format (PDF) [13]. PDF can be regarded as an extended version of Postscript, a page description language which has become an industry standard [11].

A PDF document contains the device and resolution-independent page descriptions familiar to PostScript users. Each page description may include a combination of text, graphics and images. In addition, PDF supports extra document features such as hypertext links, thumbnail sketches, bookmark and annotation facilities. These are not part of the page itself, but are useful for online document navigation. These and some of the other main features

[2] Further information concerning access of sample CAJUN documents via Gopher, World Wide Web and anonymous ftp may be obtained from circus@cs.nott.ac.uk.

supported are described in sections 4.1 to 4.7 below. The PDF format also supports various forms of compression of text and images to reduce the size of the PDF files [13]. The compression formats available are JPEG, CCITT and LZW.

Conversion of documents to PDF is possible for most files produced on platforms that support Acrobat; that is, IBM-compatible PCs, under MS-DOS or MS-Windows, and Macintosh hardware. A Solaris/X-windows version is at the beta-test stage [22]. Documents may be converted in one of two ways. Those not in PostScript form may be converted using the PDFWriter, a pseudo printer driver which, instead of generating commands for a specific printer, produces PDF files [22]. Documents in PostScript form may be passed through Acrobat's Distiller software.

Distiller is an enhanced PostScript interpreter which translates PostScript into PDF. A new PostScript operator, **pdfmark**, has been introduced. Using this operator, the additional non-printable document features mentioned above can be described in the PostScript version of the document. The descriptions using operator are recognised by Distiller and translated into the appropriate PDF features.

PDF documents can be viewed and printed using Acrobat Reader software. A more sophisticated version of the Reader, Acrobat Exchange, allows, in addition, full editing of non-printable features. Thus, the Exchange viewer can be used to add these non-printable features directly to the PDF file, rather than describing them using the **pdfmark** in a PostScript file.

The Acrobat software suite was first announced in late 1992. The functionality of the package is expanding and improving with each release. Future versions promise, among other advances, improved search and retrieval facilities and cross-document links [22].

There is also the possibility of some form of logical document structure. As Smith *et al* [22] comment, this would enable the development of "more complex searching algorithms, content revisability and sophisticated database access".

Sections 4.1 to 4.7 introduce the main features available via Adobe Acrobat. Their potential use in skimming and reading journals is discussed in sections 5 to 8.

4.1 Hypertext Links

Hypertext supports non-linear traversal of documents[3]. The hypertext link feature incorporated in Acrobat allows users to follow links which have been set up from one section of a paper to another. For example, links can be made and followed between citations in the text and the bibliography page on which the reference appears. This is illustrated in Figure 1 in which the sources of the hypertext links are indicated by being boxed. Clicking on the citation indicated by the boxed figure 9 will cause the appropriate bibliography page to be displayed. Similarly, clicking on a boxed figure number will display the page on which that figure appears. For example, a click on the box containing the phrase 'Figure 1(a)' will result in the display of the page illustrated in Figure 2.

The default links set up in the papers distributed by the CAJUN project are from citations to references, figure or table references to figures or tables respectively and footnote citations to footnotes.

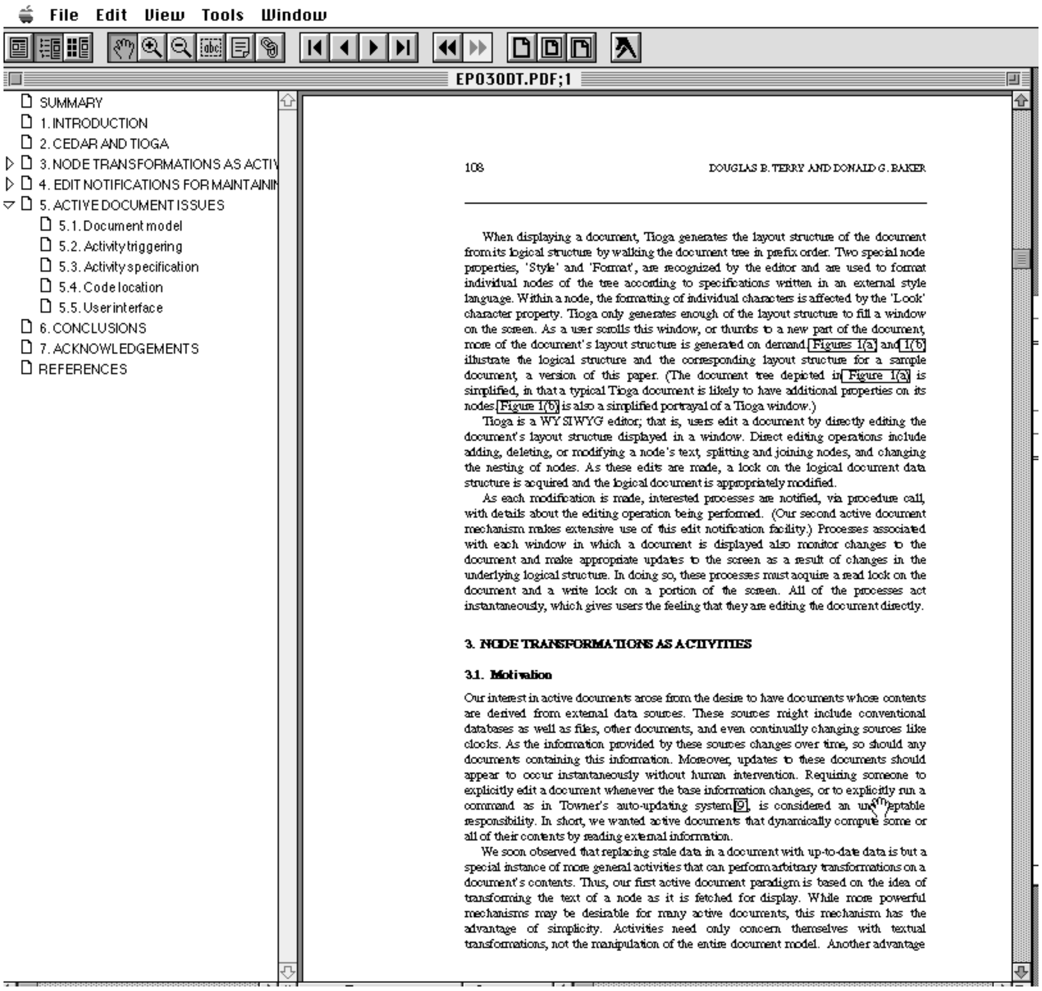
Links may be created using the links tool in the Exchange program. The tool is activated by the links button on the tool bar near the top of the Exchange display. The button can be identified by the icon representing two links of a chain, as seen in Figure 1. In Version 1 of the Acrobat software, only intra-document links are possible. Inter-document links are expected in Version 2 which is due some time in 1994.

4.2 Thumbnail Sketches

Figure 2 illustrates the thumbnail feature. This provides a gallery of miniature images of the pages of a document. As can be seen, figure outlines are clearly distinguishable. When presented online, the text of each page is represented as a grey area on the thumbnail sketches; this detail has been lost in the transition to hard copy screen dumps.

Clicking on a thumbnail takes the user directly to the page indicated by that sketch. The spot on which the user clicks determines the centre of the page display.

[3] See [17] for an overview of hypertext.



108 DOUGLAS B. TERRY AND DONALD G. BAKER

When displaying a document, Tioga generates the layout structure of the document from its logical structure by walking the document tree in prefix order. Two special node properties, 'Style' and 'Format', are recognized by the editor and are used to format individual nodes of the tree according to specifications written in an external style language. Within a node, the formatting of individual characters is affected by the 'Look' character property. Tioga only generates enough of the layout structure to fill a window on the screen. As a user scrolls this window, or thumbs to a new part of the document, more of the document's layout structure is generated on demand. Figures 1(a) and 1(b) illustrate the logical structure and the corresponding layout structure for a sample document, a version of this paper. (The document tree depicted in Figure 1(a) is simplified, in that a typical Tioga document is likely to have additional properties on its nodes. Figure 1(b) is also a simplified portrayal of a Tioga window.)

Tioga is a WYSIWYG editor; that is, users edit a document by directly editing the document's layout structure displayed in a window. Direct editing operations include adding, deleting, or modifying a node's text, splitting and joining nodes, and changing the nesting of nodes. As these edits are made, a lock on the logical document data structure is acquired and the logical document is appropriately modified.

As each modification is made, interested processes are notified, via procedure call, with details about the editing operation being performed. (Our second active document mechanism makes extensive use of this edit notification facility.) Processes associated with each window in which a document is displayed also monitor changes to the document and make appropriate updates to the screen as a result of changes in the underlying logical structure. In doing so, these processes must acquire a read lock on the document and a write lock on a portion of the screen. All of the processes act instantaneously, which gives users the feeling that they are editing the document directly.

3. NODE TRANSFORMATIONS AS ACTIVITIES

3.1. Motivation

Our interest in active documents arose from the desire to have documents whose contents are derived from external data sources. These sources might include conventional databases as well as files, other documents, and even continually changing sources like clocks. As the information provided by these sources changes over time, so should any documents containing this information. Moreover, updates to these documents should appear to occur instantaneously without human intervention. Requiring someone to explicitly edit a document whenever the base information changes, or to explicitly run a command as in Towner's auto-updating system [9], is considered an unacceptable responsibility. In short, we wanted active documents that dynamically compute some or all of their contents by reading external information.

We soon observed that replacing stale data in a document with up-to-date data is but a special instance of more general activities that can perform arbitrary transformations on a document's contents. Thus, our first active document paradigm is based on the idea of transforming the text of a node as it is fetched for display. While more powerful mechanisms may be desirable for many active documents, this mechanism has the advantage of simplicity. Activities need only concern themselves with textual transformations, not the manipulation of the entire document model. Another advantage

Figure 1: An article page viewed using Exchange

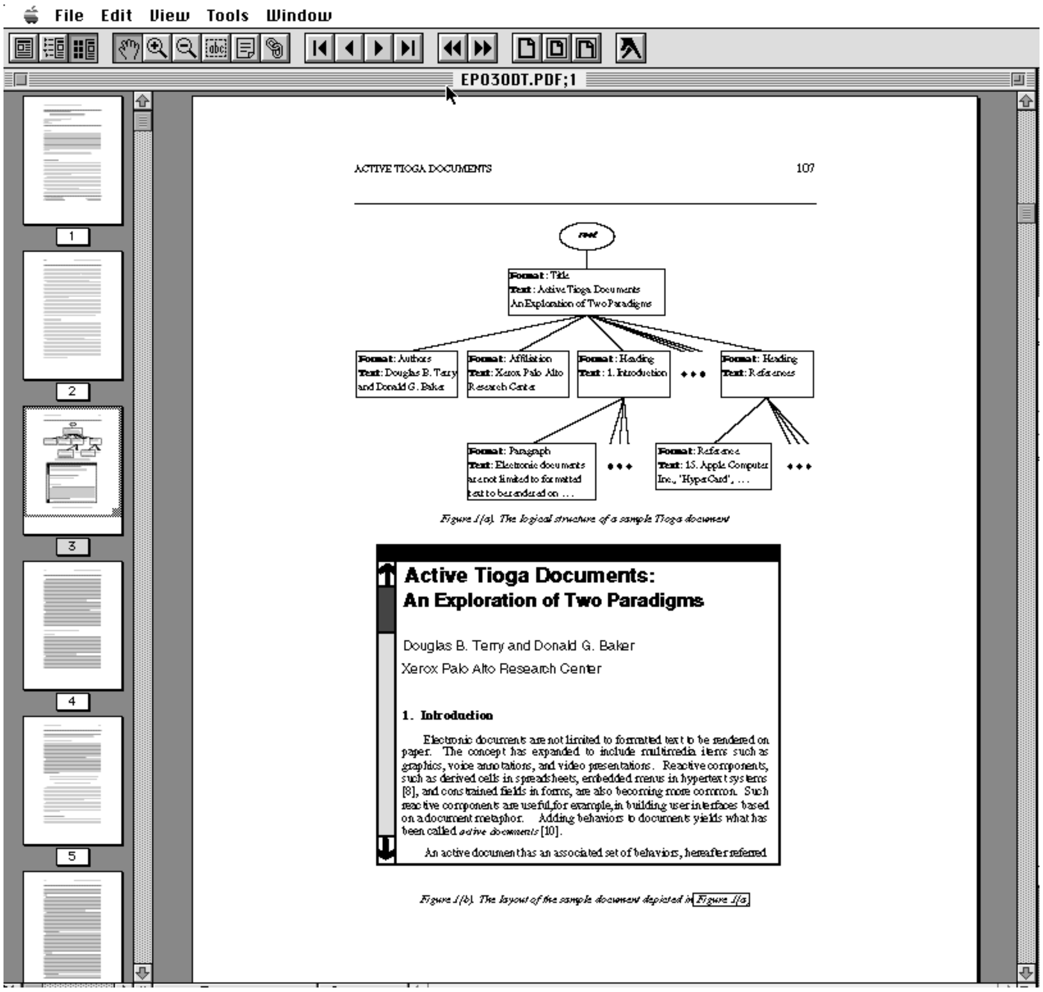


Figure 2: An article page with thumbnails

As with hypertext links, thumbnails can be added to the document either while the PDF file is being created or during use of the Exchange program. This latter option is sometimes preferred as a default in order to save file space and increase speed of access to the PDF file. To add thumbnails using Exchange, the appropriate option is chosen from the Edit menu near the top of the display. Once thumbnails have been added to a document, they may be displayed by clicking on the thumbnail icon which is third from the left in the tool bar. The width of the subwindow displaying the thumbnail sketches can be altered to show more than one column of thumbnails if preferred.

4.3 Browse and Navigation Buttons

The browse buttons are the group of four arrow-headed buttons on the tool bar in Figure 2. They allow movement to the first and last pages of a document and to the next and previous page in numerical order. The double-headed navigation buttons, to the right of the browse buttons, allow users to retrace their steps through a document, moving to each view in the order visited.

4.4 Document Display Size

The Exchange program enables magnification and reduction of the document display via the "zoom-in" and "zoom-out" tools, which appear as magnifying glasses in the tool bar of Figure 2. Alternatively, the magnification level may be chosen using the magnification box in the lower status bar at the bottom of the display. Again, this box can be identified by the magnifying glass icon. In Figure 2, the document is displayed at 85% of actual size. The default document size chosen by the CAJUN project is 120% of actual size.

In addition to these options, Exchange provides three default page viewing settings. These are activated by choosing one of the three viewing size buttons located at the right hand end of the tool bar and indicated by icons representing document pages. The first of these buttons displays the page at its actual size. The second scales the page so that it fits within the main window and the third ensures that the page width just fits the window.

4.5 Bookmarks

Bookmarks can mark points of interest, hard-to-find or often-used information in a document. As well as this indexing role, they can also be used to provide an outline of the paper, as illustrated in Figure 1. In this example, the bookmarks are nested, main headings being represented by top level bookmarks, subheadings by the next level down and so on. Papers can optionally be set up so that, on opening them, only those bookmarks above a specified level are initially visible. In this case, further levels of bookmarks can be opened out by clicking on the triangle associated with the visible bookmark entry. In Figure 1, the top level bookmarks and one set of lower level bookmarks are visible.

In the papers distributed by the CAJUN project, the bookmark facility is set up during development of the PDF file to provide a contents summary based on section headings and subheadings.

Bookmarks can also be added using Exchange by choosing the appropriate option from the Edit menu. Having been created, they can be displayed by clicking on the bookmarks icon which is second from the left in the tool bar. User-defined bookmarks may be added to any provided by the PDF document developer.

4.6 The Note Tool

The note tool is positioned to the left of the links tool in the tool bar. It can be used to place notes in documents. Such notes may be of use in various situations. For example, referees or editors may wish to include comments for document authors and individual readers may wish to annotate documents as they read them. Once the tool has been activated, the position of a note on a document page is chosen by a click of the mouse button. This opens a resizable note window as illustrated in Figure 3. Text can then be entered in the window which, on closing, appears as a small "note" icon. Again, notes may be added to the document during development of the PDF file as well as during use of the Exchange program.

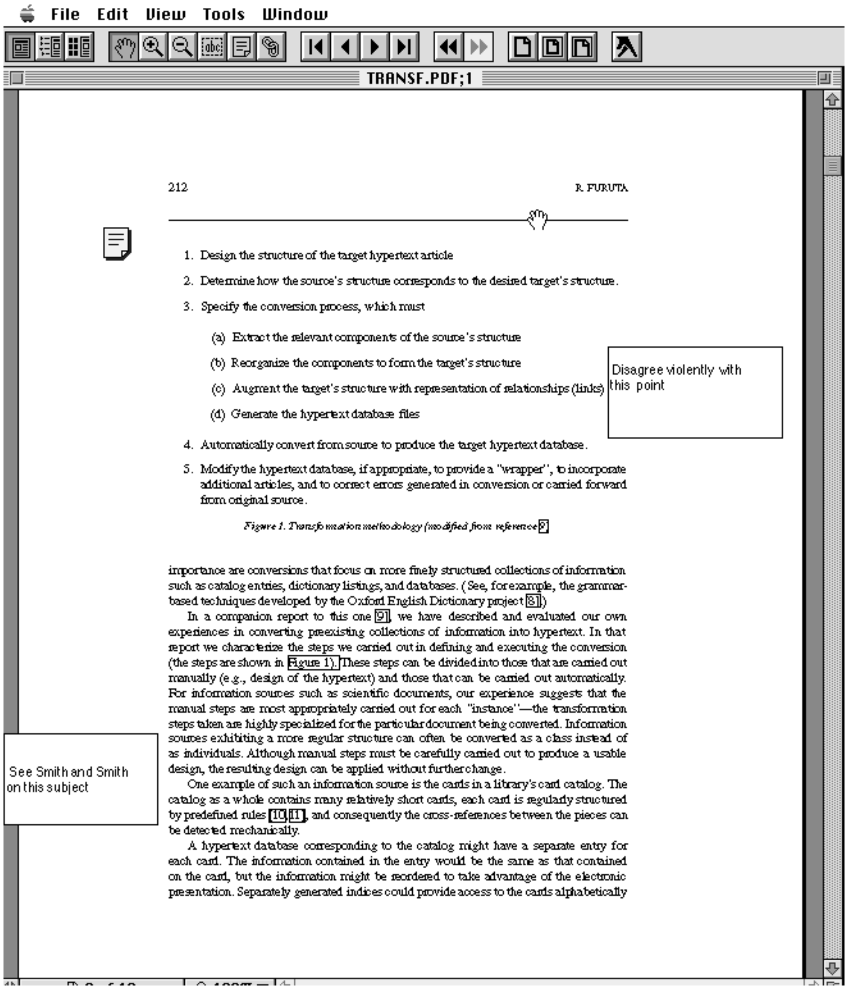


Figure 3: An article page illustrating the note tool

4.7 The Find Feature

Words, or parts of words, can be searched for using the Find feature. This is activated by the "walking fingers" icon located at the far right of the tool bar in Figure 3. The page containing the word is displayed with the word highlighted. A feature exists to find further occurences of the same word. As mentioned in Section 4 improved search and retrieval facilities are expected in future version of the Acrobat system.

5 Skimming and Reading Journals

The differing information needs of users lead to differing modes of document perusal. For example, users often skim journals for general information concerning their structure and the content of individual articles. In addition, they may study particular papers in more depth.

As human beings, we have evolved techniques for information retrieval that make the best use of our information processing capabilities within the confines of available technology. For example, as further discussed in Section 7.1, we tend to use methods that minimise the necessary hand and eye movements when skimming or reading paper-based journals. In order to determine the most appropriate form of presentation for electronic journals, we must work backwards from a consideration of how we read hard-copy journals to a determination of why we use these particular methods. Only then can we distinguish the underlying principles of readability from the limitations imposed by a paper-based technology. We can then ensure that we translate into electronic form only those hard-copy features that aid readability.

Carlson [6] suggests that the use of hypertext may eventually bring about a "paradigm shift in text delivery and in human information processing" but that, at least to begin with, the most successful approach may well be to attempt an electronic emulation of "many of the strategies a sophisticated reader uses in dealing with hard copy". Whether or not this paradigm shift is realised, it appears that many of the successful techniques for assimilation of online information stem from our experience with paper-based systems.

Sections 6 to 8 discuss some of the techniques that readers have evolved for reading and skimming paper-based journals, and the journal features that reinforce or encourage these techniques. An attempt is made to identify those features which could be constructively replicated in electronic form. The Acrobat system as set up in the CAJUN project is used as an illustration of the issues discussed. Section 6 considers these issues in relation to skimming the general information, provided by a journal, concerning its content and structure; for example, indexes and contents pages. Section 7 refers to conventions in the skimming of individual articles, and Section 8 to the more detailed study of such articles.

6 Skimming Journal Information

A well-designed hard-copy journal has a clearly defined structure at both the journal and individual article level. Although this structure may differ from one title to another, it should be easily recognisable to the user, thus simplifying the tasks of orientation, navigation and information retrieval.

Paper-based journals have developed conventions for presenting information concerning their general content and structure. Some general conventions or standards will need to be developed for the electronic presentation of such information so that readers can instinctively access the various features of an electronic journal without any previous knowledge of that journal's setup.

Sections 6.1 to 6.4 consider the general information provided by paper versions of the *EP-odd* journal and discuss possible improvements to the current Acrobat facilities.

6.1 Hard Copy Indexes

The provision of indexes is an area in which the electronic dimension can provide much added value.

In its hard-copy form, each issue of the *EP-odd* journal incorporates a contents page. Articles are identified via a combination of title and author. Other non-article features are listed under their general title; for example, Book review and Conference Diary. Access information is in the usual form of the appropriate page number.

The final issue of a volume incorporates contents details of the entire volume, and also author and keyword indices for the complete volume These contents details are set out in a similar form to that found in the contents pages for individual issues, except that the contents of each issue is listed separately. Both author and keyword indices are simple alphabetic listings with page numbers.

6.2 Online Indexes

As commented in Section 4, Version 2 of the Acrobat software is expected to provide the facility for cross-document links. A document, in this context, refers to a single PDF file. Hence, cross-document links may be simulated with the current Version 1 software by storing several concatenated documents in a single file.

The current version of the CAJUN *EP-odd* archive on CD-ROM is organised into volumes. Each PDF file represents one volume. Thus, links may be made from volume contents pages, keyword and author indexes to the appropriate papers. This allows for immediate access to articles or features via a simple mouse click on the relevant index item.

The cross-file links facility expected in Version 2 of Acrobat may encourage the reduction of file size to individual paper length. This would be of particular advantage in network access of documents where the transmission of an entire volume may be a restrictingly lengthy process. In addition, as already mentioned in Section 2, it is an open question as to whether future electronic journal articles will, in fact, be packaged in issues or whether they will be available in a different format.

6.3 Other Information

On the inside back cover of every issue of *EP-odd* are "Notes for Authors". Such notes can be long and involved and some other journals include them only infrequently, perhaps once a volume, in order to save space. This can be very awkward for potential authors; they may have only one copy of the journal to hand and it may not be one of those issues which include the relevant details. With online journals, the problem of lack of space could be overcome by including a link to the relevant information rather than repeating it for every journal issue. Again, in the case of Acrobat, inter-document links are awaited.

The *EP-odd* notes include the information that "Detailed instructions for preparing text in troff, TeX, LaTeX, are available from ..." the publishers. For an online version, a link to templates or macros for each appropriate formatter could be included here so that authors could use them in the preparation of their papers.

Details, including the editorial board and the aims and scope of *EP-odd*, are found on the inside front cover of each issue. Again, in an electronic version, there would be no need to repeat this for every issue; a link to the appropriate details would suffice.

6.4 Subscriptions

There has, as yet, been relatively little public discussion about the inevitable financial implications of the proliferation of electronic journals. Access to the majority of online information cannot remain free of charge indefinitely.

The *EP-odd* journals incorporate a tear-out subscription order form inside the back cover; again, this could be incorporated in an electronic version. Perhaps this is an example of a really "active document", one which actually takes your money! Again, whether subscription would be to the journal *per se* or to individual articles remains an open question.

7 Skimming Journal Articles

There are many possible scenarios in which a user might wish to skim or browse electronic journal articles. For example, they may wish to

- Look at papers just published (the latest "journal issue"),

- Browse a database of back-issues,
- Browse a selection of papers from this database, chosen, for example, by keyword, date, author, citation, volume or issue.

The off-line approach to skimming journal articles described below incorporates many advantageous features, some of which are supported in electronic form by Acrobat Exchange.

7.1 Skimming Hard Copy Articles

Every reader has his or her own method of skimming hard-copy journals. I tend to proceed using the following method.

When I pick up a journal, for example, a copy of *EP-odd*, I first flick through it from beginning to end. To do this, I hold the journal with my left hand and allow the pages to pass the thumb of my right hand. This means that I can start at the front of the journal and work to the back with one movement. This action gives me an overview of

- What papers are included,
- Roughly how long each one is,
- What each paper looks like,
  - What figures and tables there are,
  - The layout of sections and subsections.

I am often aware of moving from one paper to the next because the final page of one paper and the first page of the next are usually distinctive enough to stand out. However, I sometimes skim the journal so fast that I miss the significance of these two pages. In this case, I know when I have moved on to a new paper because the running headers change. At the top of each double spread of an *EP-odd* paper are the authors' names on the left-hand page and the paper's title on the right-hand page. I will generally glance once at these two features for each paper I come to. When I move to the next paper, the pattern of author and title changes. It is not necessary to focus the eye on the title or author. I generally focus on the main contents of the page, noticing the title change with my peripheral vision.

This approach has significant advantages in the minimal hand and eye movements it involves. There is no movement of the left hand, and virtually no movement of the right hand. Despite this, the reader retains relatively good control over the process and can stop fairy easily on a required page. With regard to eye movement, the reader soon recognises the generic structure or layout of a page and, if interested in one part of that structure, such as general contents, author or page number, need hardly move the eyes at all. One feature may be kept in direct view, another one or two features may be in the reader's peripheral view. A reader may skim a paper several times, concentrating on a different feature or combination of features on each occasion.

The reader also has the option of concentrating on one side of a double page spread and allowing the other side to be in peripheral view. This can be done by tilting the journal to the left or right. Doing this can give readers a general overview of the paper without requiring them to concentrate on the contents of every page.

7.2 Using Features of Acrobat Exchange for Skimming

There are several features provided by Acrobat Exchange which might be potentially useful in replicating the browsing actions described above. They are

- the thumbnail feature,
- the browse and navigation buttons and
- the bookmark feature

To be of use as tools for skimming or browsing, the features described above need to be able to replicate those characteristics of manual skimming that maximise readability.

7.2.1 The Browse and Navigation Buttons

The browse and navigation buttons would appear obvious tools for skimming journal articles. However, although their use allows document pages to be displayed at a reasonable rate, the speed at which one would typically flick through an article or a journal issue cannot be replicated. This is a particular problem when pages incorporate high resolution images as these tend to be drawn more slowly.

One proposed solution to this problem would be to present the journal articles using low resolution images [1]. The resulting increase in speed of the browse and navigation buttons would improve their usefulness as tools for skimming. However, the low resolution might be inadequate for more concentrated study of the images involved.

A further difficulty concerns the degree of eye movement involved in the use of these buttons. Although relatively low, it does not replicate the near minimal eye movements required to skim a paper journal. The order in which Acrobat draws the elements of a page depends on the order in which the elements appear in the PDF file page description. In theory, this means that elements in a page of text may appear on the screen in any order. In practice, the PDF file is generally ordered such that the text appears in order of its position on the page, starting at the top of the screen. Thus, the user is unable to gain an instant overview of an entire page. Instead, they must wait until the whole page is drawn. The eye tends to follow each page as it is being drawn, rather than concentrating on the same section of each page as in manual skimming. Pages involving complicated figures further disrupt the pattern of eye movement as the order of appearance of the elements of the figure cannot be predicted by the user.

The hand movement involved in using the buttons to skim the articles is considerably greater than that in the manual process. The number of mouse clicks required to view even one article, at one click per page, can be tiring. If the user clicks the mouse button fast enough, they can jump two, three or even four pages at a time without the intermediate pages being shown; but this is particularly tiring for the hand. An ideal skimming feature might provide the electronic equivalent of letting the pages slip through the user's fingers. As this feature would be solely used for browsing, the use of low resolution versions of images would appear appropriate.

As most right-handed users (and many of us who are left-handed) use the mouse with the right hand, it might be appropriate to range the menu items, such as the browse and navigation buttons, along the right hand edge of the screen, rather than along the top edge. This would parallel more closely the actions taken when reading a paper manually. When I want to move forward to a particular page, I lift the right hand edge of the current page. In particular, I often lift the bottom corner of the right hand edge. A "Next Page" icon at the bottom right corner of the page display might therefore be useful. If a vertical display of icons was regarded as inappropriate because it interfered with displaying a paper at full screen width, the menu options might be ranged to the right hand end at the top of the screen, rather than starting at the left as currently. For those who use the mouse with their left hands, an option could be provided to customise this feature so that options were ranged to the top left, as currently, or along the left hand edge of the screen.

7.2.2 The Thumbnail Feature

Once thumbnail sketches are created, they can be presented at a faster rate than the pages of the main display. The sketches may be moved either using the vertical scroll bar or by clicking on the arrow heads at either end of this bar. This latter option can result in the sketches moving at a fairly fast rate. This is acceptable as there is little information to be gleaned from the thumbnail sketches, except in situations in which there are a large number of diagrams spread evenly through the paper.

Because the thumbnail sketches travel vertically down the page, again, it is not possible to replicate the eye movements involved in skimming a paper journal. If the sketches travelled horizontally across the page, the eyes would need to travel a shorter distance. However, the thumbnails would still not be very useful because the sketches are too small to convey much useful information. If the sketches are of a sufficient size to allow the user to see the details, very few will fit onto the screen at once. Again, the ideal feature might more closely replicate the manual operation of seeing just one page at a time but seeing the pages in rapid succession.

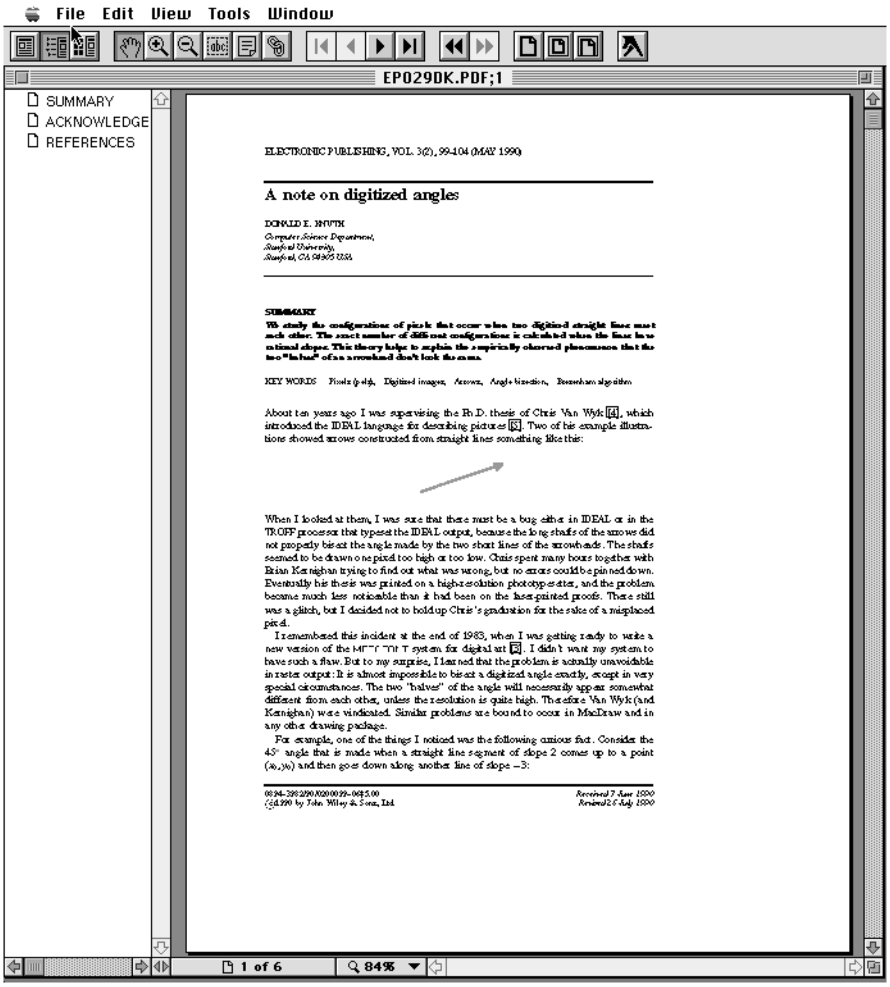


Figure 4: Example of uninformative bookmarks

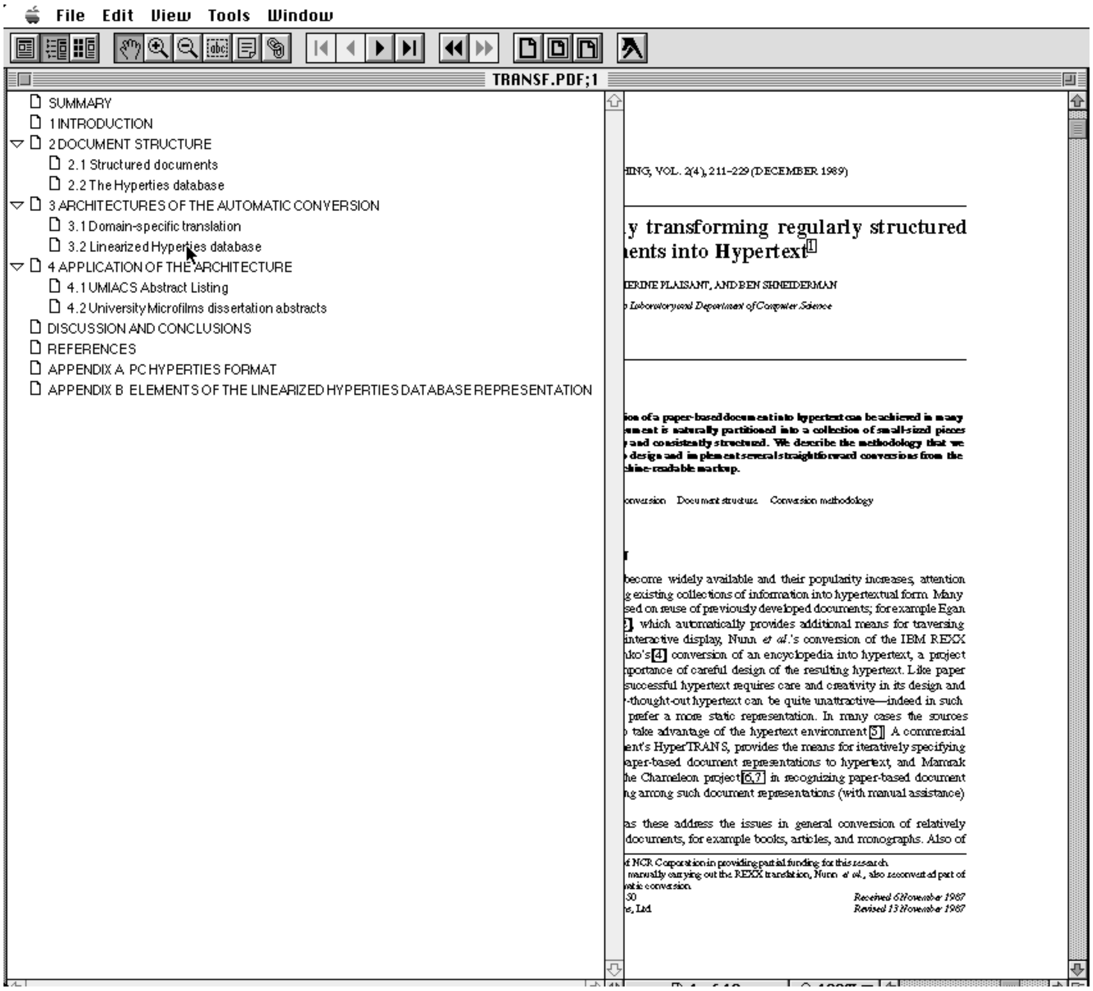


Figure 5: Top level bookmarks

7.2.3 The Bookmark Feature

As described in Section 4.5, the default setup of bookmarks in the CAJUN project is to provide a contents summary based on section headings and subheadings. In this case, the bookmark feature may be potentially useful for browsing papers or journals, depending on what information the author conveys via headings and sub-headings.

Main headings in journal articles are often very general. For example, Donald E. Knuth's "Note on Digitized Angles" [14] shown in Figure 4 has only three headings; Summary, Acknowledgments and References. There are no sub-headings.[4] In this case, the bookmark feature might well be useful in an indepth study of the article, but it would not be useful as a tool for initial browsing.

Other authors provide more information. Figure 5 illustrates the top level section headers of a paper [10]. The majority of section headings do not provide much clue as to the hypothesis or content of the paper, though Sections 2, 3 and 4 give some information about the general subject area. It is only when the bookmarks are expanded to show sub-headings, as in Figure 6 or when appendix titles are considered, that more specific content becomes clear. At this level, there is reference to the systems being discussed, for example, the Hyperties database and UMIACS abstract listing. Other authors provide yet more specific information in their main headings.

Thus, discretion on behalf of the author or editor concerning initial setup of bookmarks

[4] It should be noted that this article is technically described as a note rather than a paper.

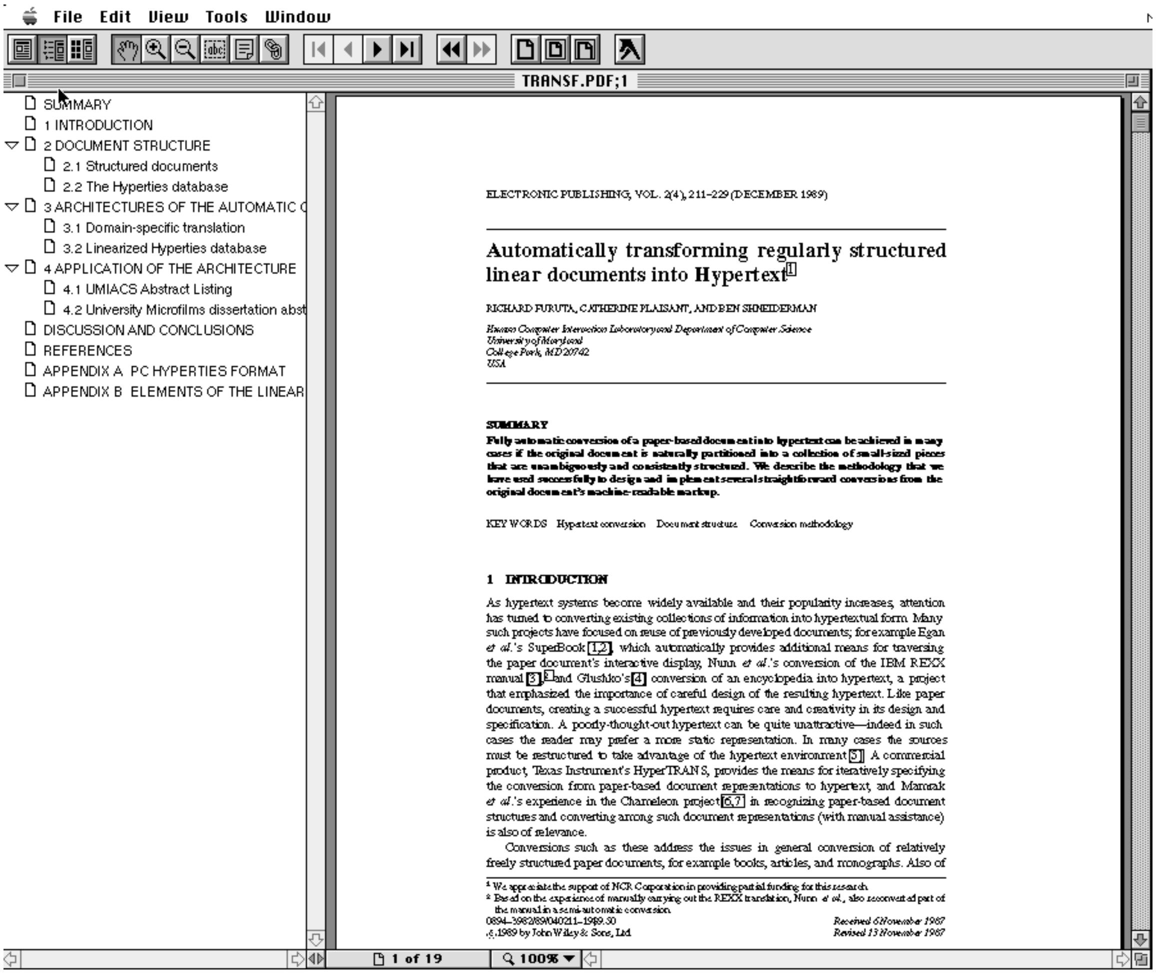


Figure 6: Multiple level bookmarks

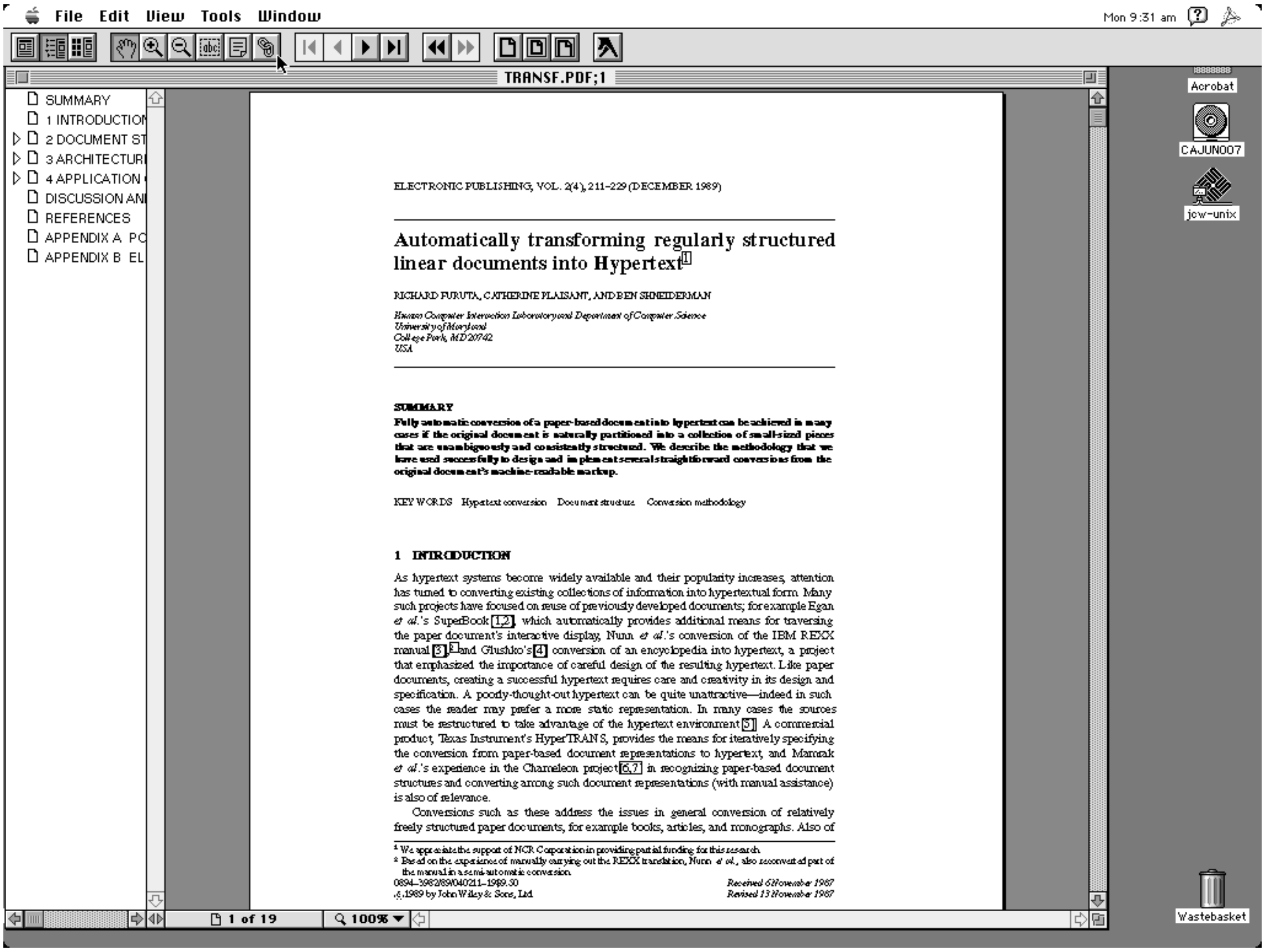


Figure 7: Illustration of potential problems with bookmark window width

appears necessary. Some lower level outlines might be significant and should possibly be visible when the paper is opened. Alternatively, it might be more appropriate to reconsider the default setting in which bookmarks automatically represent section headings and subheadings. However, a more flexible scheme in which bookmarks could, for example, describe the key ideas in particular sections or regions of an article, might reduce the scope for automating their setup.

With the proliferation of electronic journals and the standardisation of conventions concerning their layout, authors may become aware of the concerns described above and alter their style accordingly.

A further design decision concerns the default width provided for the presentation of bookmarks. Again, this will partly depend on the author's style; wordier headings will require a wider bookmark window. However, some defaults could improve general usefulness. For example, the phrase Appendix X could be excluded from bookmarks by default, thus avoiding the situation illustrated in Figure 7 where very little of the appendix title is visible.

One solution would be to introduce the concept of a separate window for bookmarks which could be brought to the front of the screen when required and moved to the back at other times. This would be simpler to manipulate than trying to expand or retract sections of the one window in which everything is currently displayed.

Where Acrobat bookmarks are used in the more traditional role of indicating pages of interest, or the point in the text which the reader has reached, a narrow window might be sufficient. In this case, the readers would introduce the bookmarks themselves and could choose titles of an appropriate length for the screen setup.

To conclude, thumbnails, as currently configured, have a limited role in skimming. Browsing and navigation buttons and bookmarks have a greater potential role in this area. Ideally, a specific skimming tool which more nearly replicates the manual process could be provided.

8 Studying a Paper

When I am interested in studying one particular journal paper, I often start by looking up its location in the journal's contents page. Once I have found its page number, I find the first page and flick through the article to get a feel for its size and structure, keeping track of the start of the article by keeping a finger at the first page. I may then read the summary. I continue by reading the headings and sub-headings and try to get a feel for the length and structure of text and diagrams under each heading. If, after this process, the article appears particularly interesting or relevant, I photocopy it and then personalise that copy. For me, personalising an article consists of some or all of the following actions.

1. Highlighting headings and subheadings,
2. Highlighting important or useful ideas and main concepts (often using a colour code),
3. Adding notes in margins,
4. Circling important words,
5. Underlining important sentences.

At the same time as performing the above actions, I read the article. I don't usually manage to read the entire paper in one sitting. When I decide to stop, I leave the photocopy open at the page I have reached or I make some mark on the copy to indicate where I stopped reading. In addition to marking the photocopy, I sometimes make separate notes, often grouped under the same headings as used in the article.

When I look back at the paper, my personalisations give it added value for me. They probably don't give the paper added value for anyone else and, if a colleague asks to look at the paper, I usually wish I'd remembered to take two photocopies and left one pristine.

8.1 Making Notes

The Note tool in Acrobat Exchange provides the potential basis for the types of personalisations described above. This tool could be extended in several ways.

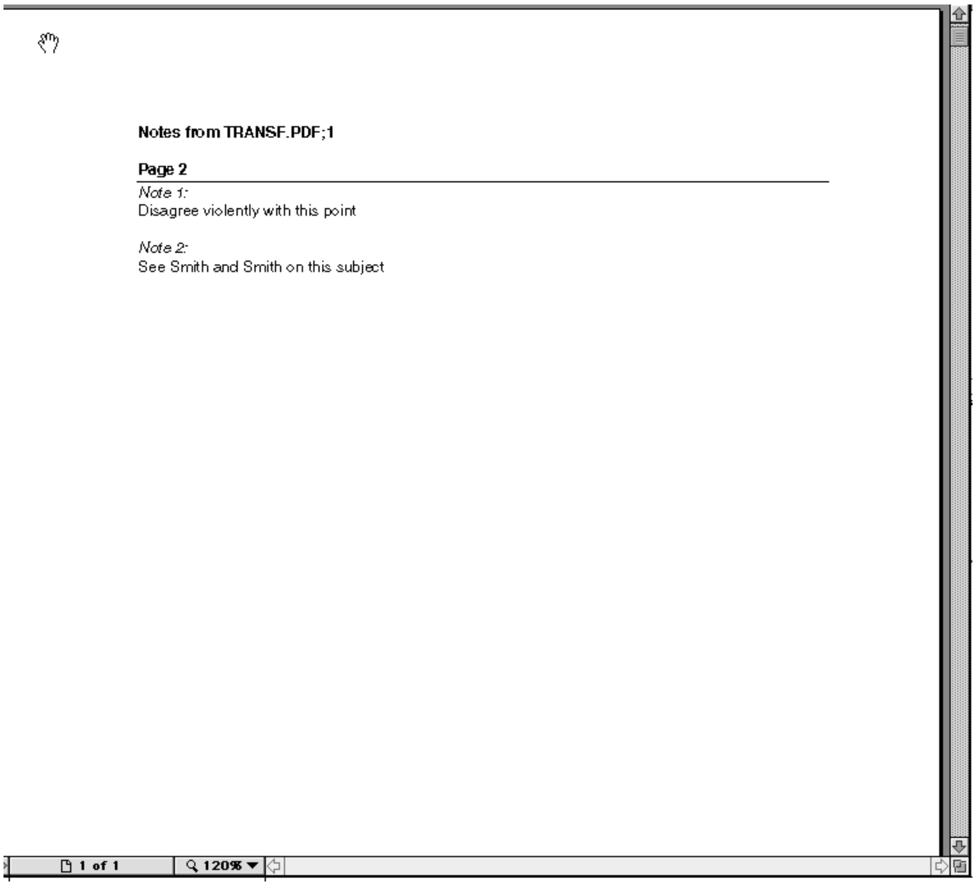


Figure 8: A Note File

As can be seen in Figure 8, a PDF file can be created to hold all the notes developed for a particular article. As currently implemented, there is no indication in this file, beyond the page number, as to where in a document a note appears. By incorporating some form of structure in the file, it could form the basis of a personal note system for the reader. If information from the bookmark feature could be incorporated, the notes made in the note windows could be displayed in the note file arranged under the appropriate headings and subheadings.

8.2 The Following Links Feature

As mentioned in Section 4.1, the Acrobat system allows the provision of hypertext links within documents. This has the potential of adding a great deal of value but also raises several issues concerning appropriate defaults and conventions. Ideally, such decisions should be based on the particular needs of journal users rather than simply adopting the conventions currently emerging for hypertext systems in general.

If hypertext does indeed bring about the "paradigm shift in text delivery and in human information processing" predicted by Carlson, then future authors may well be writing articles with hypertext in mind. They may even incorporate their own hypertext links. However, for the time-being at least, authors will continue to write articles geared, in the main, for submission to conventional paper-based journals. Thus, the journal articles involved may have been developed before the question of adding hypertext links becomes relevant. In this case, some of the central questions are

- Who will add links to the text?
- At what stage of article development will the links be added?
- Should defaults exist as to which features of an article should be linked?
- If so, what should these defaults be?

As Langford and Brown [15] point out, constructing effective hypertext is not simple. They note the conflicting temptations of either linking everything in sight or of providing only a very narrow author-defined path through the document. The former is very expensive, time-consuming and not always helpful to the reader. The latter may be inappropriately restricting to the reader. The time constraints involved in journal production may be even more stringent than in the development of a one-off hypertext system. Some form of link automation is foreseen in the CAJUN project [8].

As detailed in Section 4.1, the default chosen in the CAJUN project is to add links to most features in which there is a cross-reference to other parts of the same document. This includes links to figures, tables, equations and referenced sections, as well as to cited references. This can result in a large number of links; one sample paper included over eighty such links [8]. As a result of user feedback, articles are set up so that on opening them, the source of a link to a reference is made visible by highlighting it in blue (shown as boxed in the figures in this paper). As already mentioned a future release of Exchange should provide the facility to cross-reference outside a single document.

Section 6 referred to the necessity of journal structure consistency. Some form of house-style is even more vital when a hypertext facility is introduced. The challenge will be to marry this with a degree of author independence so that individual style is not lost.

When reading a paper-based article, I often keep one finger at the reference page and look up most of the references as they occur in the text. In Acrobat, following a link from a citation results in the entire reference page, on which the appropriate reference appears, being displayed, and the original page disappearing. Thus, the reader loses track of the text in which the reference occurs; they may return to it using a navigation button. This parallels the manual action in which, as the users moves to the reference page, a similar temporary loss of context occurs. An extension to the hypertext facility, allowing the particular reference to be presented alongside the link source, perhaps in a separate window, would add value to the electronic version.

The Acrobat system allows users to customise the magnification of the documents that they are looking at. Link destinations, be they bibliographic references, figures or sections of text, can be displayed at a variety of default magnifications. As a general rule, it would appear to be inappropriate to change the magnification setting initially chosen by the user.

This setting may have been chosen to take into account the user's eyesight and and the setting of the equipment. If a portion of the paper is presented at a different magnification, the user will have to refocus temporarily. This is an example of a more general rule; namely that a hypertext system should not "surprise the user" [5].

9 Conclusions

General conventions need to be developed for the electronic presentation of journals and their constituent articles. In some aspects, the introduction of online journals provides an opportunity to improve on the methods currently employed with their paper equivalents. An example of this is in the provision of hypertext links both between and within documents.

In other aspects, the current approach constitutes the best-known practice. In these cases, it would seem appropriate to attempt to replicate these methods in electronic form. In particular, it would be useful to replicate the browsing actions performed by users skimming paper copies. Various features of the Acrobat system are contenders in this area: the thumbnail and bookmark features and the browse and navigation buttons. None of these, however, provides the versatile functionality provided by paper copies. For more detailed study of articles, the Note feature appears a promising basis for the provision of customisation functions.

The concept of the metaphor used in many WYSIWYG[5] systems, is based on the assumption that new concepts can best be learnt if they are couched in terms of ones already understood. A desktop, for example, is considered to be more familiar to the majority of users than a command-line interface. But metaphors are topical; they can go out of date. At this point in the development of electronic journals, it might be appropriate to repeat the questions as to what concepts users find familiar and how this familiarity may change in the future. In this way, we can ensure that the features we retain in the transition from paper to electronic media are those which are based on the principles of good readability and not those based on notions of familiarity that may become outdated.

Acknowledgments

I am grateful to Heather Brown for many helpful discussions and for her comments on this paper. Thanks are also due to colleagues at the University of Nottingham, particularly Steve Probets and Phil Smith, for their help and advice.

References

[1] Cajun project: Initial impressions questionnaire. Dept of Computer Science, University of Nottingham, 1993.

[2] *Libraries and IT: Working Papers of the Information Technology Sub-committee if the HEFCs' Libraries Review*, University of Bath, UK, 1993.

[3] *Digital Libraries '94: Symposium on the Theory and Practice of Digital Libraries*, Texas A&M University College Station, TX, June 19-21 1994.

[4] D.F. Brailsford and R.J. Beach. Electronic publishing – a journal and its production. *The Computer Journal*, 32(6), 1989.

[5] P.J Brown. Assessing the quality of hypertext documents. In Rizk et al, editor, *Hypertext: Concepts, Systems and Applications*, pages 1–12. Cambridge University Press, 1990.

[6] P.A. Carlson. The rhetoric of hype. *Hypermedia*, 2(2):109–131, 1990.

[7] W.P. Dodd. Convergent publication, or the hybrid journal: paper plus telecommunications. *Electronic Publishing – Origination, Dissemination and Design*, 3(1):47–59, 1990.

[8] D. Evans, S. Probets, P. Sutton, S. Benford, D. Brailsford, and L. Harrison. The CAJUN Project First Year Report. Technical report, Dept of Computer Science, University of Nottingham, 1993.

[5] What You See Is What You Get

[9] J. Foster. Networked information: tools and training. The role of the librarian in bringing these to end-users. In *Libraries and IT: Working Papers of the Information technology Sub-Committee of the HEFCs' Libraries Review*. UKOLN: The Office for Library and Information Networking, University of Bath, UK, 1993.

[10] R. Furuta, C. Plaisant, and B. Shneiderman. Automatically transforming regularly structured linear documents into hypertext. *Electronic Publishing – Origination, Dissemination and Design*, 2.4:211–229, 1989.

[11] Adobe Systems Incorporated. *PostScript Language Reference Manual*. Addison Wesley, Reading, Massachusetts, 2nd edition, December 1990.

[12] Adobe Systems Incorporated. Adobe acrobat products and technology : An overview, November 1992.

[13] Adobe Systems Incorporated. *Portable Document Format Reference Manual*. Addison Wesley, Reading, Massachusetts, June 1993.

[14] D.E. Knuth. A note on digitized angles. *Electronic Publishing – Origination, Dissemination and Design*, 3.2:99–104, 1990.

[15] D. Langford and P Brown. Creating hypertext documents: Is it worth the effort? *ASLIB*, 45(4):91–95, April 1993.

[16] J. Maddox. Electronic journals are already here. *Nature*, 365, 21 October 1993.

[17] J. Nielsen. *Hypertext and Hypermedia*. Academic Press, 1990.

[18] Commission of the European Communities. Libraries program, 1993. Directorate General XIII.

[19] A. Okerson. The electronic journal: What, whence, and when? *The Public-Access Computer Systems Review*, 2(1):5–24, 1991.

[20] D. Pullinger, February 1994. Informal comments at SuperJournal Project Presentation, Imperial College, London, UK.

[21] T. Sheridan, J. Senders, N. Moray, J. Stoklosa, J. Guillaume, and D. Makepeace. Experimentation with a multi-disciplinary teleconference and electronic journal on mental workload. 1981.

[22] P.N. Smith, D. Brailsford, D. Evans, L. Harrison, S. Probets, and P. Sutton. Journal publishing with Acrobat: the CAJUN project. *Electronic Publishing – Origination, Dissemination and Design*, 6(4), 1994.

[23] P. Waddell. The Potential for Electronic Journals in UK Academia. In *Libraries and IT: Working Papers of the Information technology Sub-Committee of the HEFCs' Libraries Review*. UKOLN: The Office for Library and Information Networking, University of Bath, UK, 1993.