
Electronic reserves: the promise and challenge to increase accessibility

Kathy Konicek
Joy Hyzny and
Richard Allegra

The authors

Kathy Konicek is Instructional Technology Consultant in the Division of Information Technology (DoIT), Learning Technology & Distance Education (LTDE), University of Wisconsin-Madison, Madison, Wisconsin, USA.

E-mail: kkonicek@facstaff.wisc.edu

Joy Hyzny (jhyzny1@uic.edu) is Assistive Technology Specialist and **Richard Allegra** (rallegra@uic.edu) is Director, both at the Office of Disability Services, University of Illinois at Chicago, Chicago, Illinois, USA.

Keywords

Electronic resources, Information, Access control

Abstract

Electronic reserves help registered campus users who need anytime-access to documents. Electronic reserves comprise digital files, mostly HTML or PDF formats. In some circumstances the HTML or PDF file is "readable" to the sighted individual, but are sometimes either partially or completely unreadable to the visually impaired using assistive technology. Creating "accessible" PDF files poses more challenges than creating "accessible" HTML files. Several options are suggested to help solve this problem.

Electronic access

The Emerald Research Register for this journal is available at <http://www.emeraldinsight.com/researchregister>

The current issue and full text archive of this journal is available at

<http://www.emeraldinsight.com/0737-8831.htm>

Library Hi Tech
Volume 21 · Number 1 · 2003 · pp. 102-108
© MCB UP Limited · ISSN 0737-8831
DOI 10.1108/07378830310467445

Introduction

Materials placed in electronic reserves are convenient and relatively accessible for researchers, students and other users. Many people prefer electronic reserves because they can conveniently login to obtain information from home, school, or work on a 24-hour and seven-days-a-week basis. Libraries consider electronic reserves a viable service that addresses storage and access issues.

Electronic information has multiple advantages over hardcopy material. Typically, electronic information will increase the level of access because material can be obtained at a distance. For people with disabilities, electronic media can be especially vital. It offers access to people who might not be able to benefit from traditional printed materials.

Assistive technologies

Assistive technology is exactly what the name implies: technology that can assist a person in meeting a goal. There are a number of assistive technology products that are used to obtain access to computer-based information. These products can range from low-tech to high-tech and consist of hardware (a physical piece of equipment such as a scanner) or software (a program that is installed on a computer that will provide a specific feature or expand a specific function). Assistive technology products are designed to meet the needs of people with various disabilities, including physical, learning, cognitive, and sensory. Access considerations for electronic reserves are particularly important for individuals with sensory disabilities and, more specifically, people who are blind. This article discusses:

- assistive technology for users without vision;
- compatibility issues for this technology; and
- providing "accessible" electronic reserves.

People who are blind can access electronic material using a computer, providing it has the specifications required to operate the assistive technology software. Such software is called a "screen reader". A screen reader scans data displayed on the computer, and conveys the



information aurally by means of a synthesized voice, thereby empowering the person without vision to negotiate information on the computer through basic keyboard commands.

There are multiple screen readers on the market. The most popular screen readers for Windows-based applications are Window-Eyes by GW Micro, Inc. (<http://www.gwmicro.com/>) and Job Access With Speech (JAWS) by Freedom Scientific, Inc. (<http://www.freedomscientific.com/>). Each one can render information differently, and function will vary depending on other considerations such as browser compatibility. The length of time it takes for a page to load will depend on the connection method (e.g. modem or Ethernet connection).

Accessibility considerations continue to exist despite the development of assistive technologies. The personal computer in the early days of the Internet was originally a text-based medium for information. Now it contains graphical and multi-sensory media. A screen reader interprets textual information, but it is unable to determine graphical images. In the case of Web sites comprising hypertext markup language (HTML), a coded text tag will provide a brief textual description of the image.

Screen readers could not navigate within Adobe Reader until recently. Adobe Acrobat Reader 4.x released a plug-in to help screen readers read the PDF file. These PDF files needed to be non-image files or text-based files for the screen reader to read the information. The most common way to get a text-based PDF file is to have the digital (electronic) version of the document in Word, WordPerfect or other word processor and create a PDF document from this electronic file. Screen readers can recognize these characters and voice output the information. On the other hand, an image file is similar to taking a picture of the hard copy document with a camera. Essentially you have no recognizable characters and the screen reader cannot voice output this information other than telling the user it is image one, image two, etc.

Creating a PDF file from a scanned copy (placing a hardcopy document on a flatbed scanner) typically creates tagged image file format (TIFF) files or other image format files. Because TIFF files are image files, screen readers cannot read these. Adobe Acrobat

(<http://www.adobe.com>) produces a PDF file that amazingly retains the look of the original document. This “page scan” cannot be read by screen readers. However, if the TIFF files are scanned using optical character recognition (OCR) the text become readable. Adobe Capture performs an OCR scan, but when it can not decipher a letter, it imbeds images of unrecognized characters. These characters look like text, but are actually images and screen readers cannot read these “image” characters unless one edits these “image” characters to make them into text. Proofreading and editing such PDF files is extremely time consuming.

Copyright and PDF

Converting a document to a text based format (as described above and throughout this article) may infringe on copyright protection. It is evident that all libraries surveyed have knowledge of the copyright issue, and caution is used before proceeding. In light of this, libraries that make articles available online are placing only a very small percentage of items owned on the Web. Frequently, a login screen to prevent unauthorized users from obtaining electronic reserves is utilized.

There are limited statistics available that pertain to the use of electronic reserves and how they are developed. An informal phone survey of select university libraries revealed that most libraries simply create PDF files, although some also create HTML files of the same article. Nevertheless, only a small percentage of holdings are placed online. Copyright is the issue of primary concern. The libraries surveyed have different policies for proceeding with the implementation of information online. All try to obtain copyright permission from the publisher and will go forward only when permission is granted.

Most electronic reserves use either HTML or PDF (Adobe portable document format) file formats. There are fewer accessibility issues with HTML than with PDF reserves. The main issue to consider with HTML documents is coding to ensure accessibility of tables and graphics. There are additional ways of increasing the accessibility of HTML electronic

reserves by linking to categories within an article.

At this time, considerable potential exists for making electronic reserves accessible to people with disabilities. Being aware of the issues that prevent PDF files from being accessible will help us find a solution. When this is achieved, electronic reserves will meet the ultimate goal of providing universal access. Understanding the process and the issues faced in providing accessible PDF files will help librarians, disability services staff, faculty and students with disabilities work together towards a solution.

Neal Ewers, instrumentation specialist at the University of Wisconsin TRACE Center (Technology Research and Development Center) uses a screen reader and often has difficulty with PDF files. When tables are within a PDF file, the read order of the table is linearized (there is no easy way to move from column to column and from row to row). In his research, he has found that the author cannot access the “source code” of a table in the PDF file to correct the problem. It may look visually acceptable, but for those listening to a screen reader, the file may make no sense. In addition, Ewers sometimes cannot even access the file, especially when security settings (such as disabling copying) are applied to the PDF file. At this time, he prefers an HTML file or Word document with styles or even the hard copy processed through a scanning and reading program such as OPENBook by Freedom Scientific, Inc. (http://www.freedomscientific.com/fs_products/software_open.asp) which sometimes does a better job of a logical read order of text.

Making electronic material accessible

There are several ways and steps to make PDF files accessible and understandable to those with screen readers. And there are different steps to create an accessible PDF document, depending on the original format of the document (hardcopy, electronic copy or imaged based PDF file).

A move in the right direction is the Adobe Acrobat 5.0 “Make Accessible” plug-in. The Make Accessible feature allows the author of the PDF file to create a logical read order and

navigational system (as well as some other features mentioned later in this article). The University of Wisconsin-Madison has found that voice output was possible with JAWS version 3.7 and an accessible PDF (<http://wiscinfo.dotl.wisc.edu/ltde/access/>). However, depending on the PDF file, screen readers may still have difficulty with table read order and the bookmark list. In addition, not all screen readers on the market can access PDF.

Although Adobe is working closely with screen reader companies such as Freedom Scientific, Inc. and GW Micro, Inc. to overcome these obstacles, those who rely on screen readers may have to find other ways to access the PDF file.

If an electronic, or e-text, version of a given file is available then it can be converted to PDF format using Adobe Acrobat. This eliminates the need for an OCR program, but does not constitute an accessible PDF file. To render the file accessible it needs to be converted with the Adobe Make Accessible plug-in.

The “Make Accessible” plug-in

Regardless of whether you scanned the document or have the electronic version, the following process will create a more accessible PDF file. A screen reader needs underlying information on the structure of the document. When a sighted person scans a document they can identify headings and subheadings as a way to get to the information they want. By “tagging” a PDF document you are helping a visually impaired person get a feel for the layout of the document.

The Make Accessible plug-in creates a “tagged” Adobe PDF file from an “untagged” PDF file. A “tagged” PDF has the following characteristics: page content contained in a logical read order, word boundaries explicitly identified, font encoding mapped to standard font encoding and a structure tree comprising a standard set of tags. A sighted person can see the bold or larger font size to help understand the structure of the document. The screen reader needs this underlying structural information to allow the reader to maneuver through headings and subheadings within the PDF file.

The Make Accessible plug-in works only on a PC and needs to be downloaded from Adobe, if it was not shipped with your version of Adobe Acrobat 5.0. This plug-in allows you to convert text based PDF files made from electronic documents to a more accessible format for those using screen readers. Once installed, this plug-in appears under a menu item in Adobe Acrobat 5.0[1]. If the PDF file was created from a version other than Adobe Acrobat 5.0 then it is best to take the original file, use Adobe 5.0 to convert the file, and then apply the Make Accessible plug-in.

When using Adobe's security settings, keep in mind that some of these security options will render the file inaccessible to those using screen readers. In Adobe Acrobat 4.0 when the "copying text and graphics is prohibited" option is chosen, the file is then inaccessible. In Adobe Acrobat 5.0 one can disallow copying and pasting while still making the file accessible. Adobe Acrobat's "How to Create Accessible Adobe PDF Files" documentation contains these directions (<http://access.adobe.com/onlinetools.html>).

OCR

The major accessibility problem for PDF documents is when no (text based) electronic version is available. One possible solution, so that screen readers can read the document, is to convert the PDF "Image Only" file to a text based file. Paper copies or page scans are converted to text files with OCR software. Screen readers can then identify the text, and provide voice output. OCR software accuracy varies, depending on the quality of the document to be scanned and other factors. If the hardcopy is of poor quality (smudges, copy from a copy) it may require "clean up" by the software before OCR is performed. In addition, the complexity of the document and the quality of the OCR program will determine the amount of editing required after completing the OCR scan. This editing process can be time consuming and costly since it requires personnel.

First time OCR users may be discouraged by the number of errors. However, if faced with keying to re-create documents versus OCR,

then OCR may be more cost-effective, depending on the given document. In the past, OCR lacked the ability to reproduce the layout of complex documents. The McKinley OCR lab sampled different documents, including four magazine style pages. Omnipage Pro 11 OCRed them at 93.41 percent to 97.18 percent accuracy, and retained their layout with unheard of economy of file size (McKinley, 2001). What is sufficient accuracy? One student who uses a screen reader commented to the author that he would gladly take documents with 95 percent accuracy!

OCR programs

An article published in *PC World-Hong Kong* compared OCR programs (Chan, 1999). According to a 1999 report, OCR programs vary in accuracy from 80 to 90 percent. However, since 1999 the accuracy of OCR programs has increased, in some cases with accuracy rates as high as 99 percent. The accuracy rate is affected by the quality, layout, and in some cases font type of the original document. OminiPage Pro Version 11 combines the strengths of Caere (Omnipage) and Scansoft (Textbridge X) (McKinley, 1997, 2001). Among the new features are better page reproduction, the ability to scan over 100 languages, higher accuracy of recognition, new PDF conversion features, and voice output for PDF files.

A more recent article in the September 2001 issue of *PC Magazine* compares OmniPage Pro 11 and FineReader 5.0 (Poor, 2001)[2,3]. Both support over 100 languages and provide accurate text conversion. Choosing one product over the other will depend on the price and the features needed from an OCR program. FineReader may recognize page layout better, but for purposes of making an accessible file available for those with screen readers, layout of this sort is not as important as text accuracy.

Christine DeMerchant, manager of the Digital Media Unit at the Canadian National Institute for the Blind Library, finds OmniPage Pro 11 to be extremely accurate, often finding fewer than one error per print page. In particular she finds recognition of numbers to be superior. Most errors occur in punctuation

and in recognition of italicized text. She also states that the preferred alternative is a text file instead of a PDF file for the visually challenged. A lot can go wrong technically, even for the computer savvy individual with Adobe Acrobat Reader and screen readers to read the PDF file.

Another OCR program worth noting is AutoScan 32 by VisionShape (<http://www.visionshape.com/>) AutoScan 32 (which uses the FineReader OCR engine) shows promising speed and accuracy when used with their high speed Millennium DS scanner (60-135 pages per minute). In addition, features such as OCR-to-HTML conversion and OCR-to-PDF conversion were not only easy to set , but could be setup for batch processing. Many of the other OCR programs can also do batch processing, and some programs are adding a scan-to-HTML conversion.

Xerox's DocuShare has a feature which allows users to copy documents on a Xerox copier, which creates a PDF file that can be sent directly to the DocuShare Web space for students to access (www.xerox.com). These and other programs may be worth taking a look at. However, choose an OCR program carefully. For example, AutoScan32 costs \$995 or higher (depending on features), since it was designed for high volume production. OmniPage Pro 12 is \$569. FineReader Office version is \$399, and Adobe Capture 3.0 ranges from \$394-\$1,600 (depending on the volume). Each institution needs to determine the volume of its workload to select an OCR program. (Educational software prices quoted during Spring 2002.)

From PDF to screen reader format

Kurzweil 1000 is software that not only converts imaged based PDF files to text (OCR), but also voice outputs the text like a screen reader. The Kurzweil product converts PDF to TIFF and then performs OCR. The accuracy of the OCR is unknown for the type of documents that are typically put into electronic reserves, so this issue still needs to be investigated. Kurzweil then voice outputs the text, so this software would have the potential to run on students' machines, instead of having electronic reserve staff convert the PDF document to a readable format for screen readers. Kurzweil has a unique feature in its ability to read tables. It

reads column and row headers in a different voice than the data within the table. While the product would not require changes in the scanning process used by most electronic reserves staff (though a clean copy that is correctly aligned helps), it could be cost-prohibitive.

OCR programs are not perfect and the amount of proofreading it takes to create an accessible PDF file may be beyond the scope of electronic reserves staffing. Would 99 percent accuracy be sufficient? If the document contains a crucial text error, how will the student know and how would the correct information be conveyed? As Klein (2000) states in "Digitizing reserves", "Our requirements are simple, set up a production system that isn't going to tank our institution." What does this "production system" look like for higher education institutions? Will an OCR-based production system solve this problem? Perhaps checking out software to students who need the text read aloud to them is in fact a viable solution.

File size

An issue that libraries often face is having enough storage space for digital documents. PDF files can become quite large. OCR documents can sometimes decrease the size of a file, but if the file contains both text and images, the file size may actually increase. File size is a major concern for libraries with limited server space. PDF Writer and Distillers both perform file compression with Distiller doing the better job. To obtain the best compression, Klein (2000) suggests trying different settings in the "Job Options" dialog box of Distiller. Distiller can take a 2MB PDF file down to 500-750k with proper compression. Setting compression will add to staff time or the time needed to set up an automated system to create a small PDF file.

Service options and choices

The issues of providing an "accessible" PDF document for those using screen readers does not involve a one-step solution. Even after scanning documents using an OCR program, to make a truly accessible file one should use

Adobe's Make Accessible plug-in. Other service options range from including OCR in the scanning process, using Adobe's online site conversion tools, having students obtain image-to-text conversion software (e.g. Kurzweil), and increasing faculty awareness.

Batch processing

Scanning using batch processing can be accomplished with an OCR software package such as OmniPage Pro, Autoscan 32 by Visionshape, Textbridge Pro, FineReader, Xerox DocuShare FlowPort, Adobe Capture. They will convert non-text to text-based documents. If the solution is to convert all documents to a readable format, then research on OCR accuracy needs to be performed on a variety of documents, so that institutions can make an informed decision on the software, hardware and the process needed to efficiently convert their hardcopy documents to e-text.

We need continued progress in making PDF files accessible with help from Adobe and OCR companies. The future of OCR looks promising as the recently released new versions of OmniPage (version 12) and Fine Reader (version 6) both claim to have improved accuracy. However, when converting to PDF format, OCR programs do not appear to run the Adobe Make Accessible plug-in automatically. If this could be built in, it would eliminate another step for those wishing to create a truly "accessible" PDF document.

Online sites

Adobe online conversion tools are available (http://access.adobe.com/access_email.html). Adobe offers a variety of solutions from sending the text based PDF to a service to have it converted to a more accessible format, to allowing the user to convert up to 3 TIFF PDF files (under 50 pages) per day without a fee, if you have Adobe Acrobat 5.0.

However, in our testing of the service, many of the documents required extensive editing of the "image" text to true text so that it was readable by a screen reader. This would probably not be a viable solution for electronic reserves, due to the large conversion workload.

Kurzweil software for students

An interim solution could be to purchase OCR software for students who need access to voice output of PDF files. For example, Kurzweil Scan 1000 v6.0 costs \$995 and has a site license of ten copies for \$2,895 (<http://www.kurzweiledu.com/>). The ten copies could be loaned to students and returned to the university when the student leaves or graduates. Margaret Londergan, manager of Adaptive Technology Center at Indiana University, Bloomington has set up a program in conjunction with campus libraries whereby students are not charged for the software (Kurzweil 3000), but need to renew the license each semester. University of Wisconsin Madison also has set up a similar program and is currently reviewing the accuracy of the documents converted using Kurzweil 1000.

Faculty awareness

The typical electronic reserve probably deals with mostly hard copies, and in some cases over 90 percent of the requests coming from faculty are for conversions from hard copies. Obtaining an electronic version may be a great obstacle for the immediate future, and leaves unresolved the issue of who will request the electronic version: faculty or electronic reserve staff.

In one case a blind student received a PDF document that was not accessible. He searched the Web for the required article and was able to find the publisher's version in a readable PDF format.

Faculty awareness is the first step in understanding the need for electronic versions when creating PDF files on the Web. The steps include trying to obtain the electronic file; if that is unavailable, then requesting a digital copy from the publisher whenever possible. Encourage faculty to convert hand written documents to digital formats before converting to PDF.

Conclusion

PDF files, even when "made accessible" by Adobe's Accessibility plug-in still pose a challenge for some screen reader programs. Section 508 of the Federal Rehabilitation Act

uses the term “effective communication”, implying that information given to a non-disabled person should be available at the same time in a format that it is accessible to those with a disability.

Electronic versions of documents are becoming the norm. Getting everyone involved in the production of e-text to understand the issues being raised is a step toward producing electronic reserves that are accessible to all.

Notes

- 1 Download the Make Accessible Plug-In, available at: www.adobe.com/products/acrobat/main.html (Select downloads and select Make Accessible Plug-In).

- 2 <http://www.caere.com>, OmniPage Pro 11 or 12.
- 3 <http://www.abbyy.com>, FineReader 5.0 or 6.0.

References

- Chan, S. (1999), “OCR software eases text input chores”, *PCWorld Online*, available at: www.pcworld.com.hk/comparison/comp0799.htm
- Klein, L. (2000), “Digitizing reserves”, *Library Journal Net Connect*, Vol. 125 No. 17, Fall, pp 45-7.
- McKinley, T. (1997), *From Paper to Web: How to Make Information Instantly Accessible*, Adobe Press, San Jose, CA.
- McKinley, T. (2001), “2001 OCR lab tests: OmniPage Pro 11.0 incorporates Textbridge X – best accuracy ever”, *Intelligent Imaging*, available at: <http://imagebiz.com/capture.htm>
- Poor, A. (2001), “OCR puts pages in your PC”, *PC Magazine*, Vol. 20 No. 15, pp. 50-2.