READ IT TO ME!

Frank L. Walker and George R. Thoma
National Library of Medicine

Keywords: DocMorph Server, Speech Synthesis, Image Processing, World Wide Web, TIFF, PDF, OCR, DocView, NLM, Ariel, Internet

Abstract: New technology and software such as Ariel and DocView have made it possible for libraries to distribute information to patrons over the Internet. Ariel, a product of the Research Libraries Group, converts paper-based documents to monochrome bitmapped images, and delivers them over the Internet, while the National Library of Medicine's DocView enables library patrons to receive, display and manage documents received from Ariel systems. While libraries and their patrons are beginning to reap the benefits of this new technology, both technological and social barriers still exist that lead to difficulties in the use of library document information. For example, the formats of files received through the Internet may not be usable on the recipient's computer. In addition, disadvantaged groups of people such as the blind and visually impaired often have difficulties in reading either printed or electronic information. To investigate solutions to some of these barriers, the Communications Engineering Branch of the Lister Hill National Center for Biomedical Communications, an R&D division of NLM, has developed a web site called the DocMorph Server. This is part of an ongoing R&D program in document imaging that has spanned many aspects of electronic document conversion and preservation, Internet document transmission and document usage. The DocMorph Server web site allows users to upload document image files via the Internet for conversion to alternative formats (e.g., TIFF to PDF), thereby enabling wider delivery and easier usage of library document information. A recent development in the DocMorph Server is speech synthesis to permit users to convert scanned images of the printed word into the spoken word. From any place on the Internet, a user can upload scanned images of a book, for example, and have DocMorph return a web page that allows the user's computer to read the material out loud. This reading facility enables certain groups of people to read material that would otherwise be difficult or impossible to read. This paper describes the architecture of the prototype DocMorph Server and the design of its computer-assisted reading facility.

1. Introduction

Over the past two decades document delivery by libraries and information service providers has been evolving from use of the postal service to delivery via the Internet. With the widespread use of the Ariel[TM] system developed and distributed by Research Libraries Group, Ariel has enabled several thousand libraries to do interlibrary loan electronically via the Internet.[1,2] In addition, with software such as the National Library of Medicine's DocView, library patrons in the late 1990's began benefiting from Internet document delivery.[3,4] DocView, software that runs on all Windows[TM] operating systems, enables a library patron's desktop computer to receive documents sent by a library's

Ariel system. After an operator uses Ariel's scanning process to produce a file of bitmapped images, Ariel sends the file over the Internet using either File Transfer Protocol (FTP)[5] or Multipurpose Mime Email Extensions (MIME) email.[6] DocView is capable of displaying monochrome bitmapped images in either the Group on Electronic Document Interchange[7] (GEDI) file format used by Ariel systems, or in the Tagged Image File Format[8] (TIFF).

While libraries and their patrons are beginning to reap the benefits of this new technology, both technological and social barriers still exist that lead to difficulties in the use of electronic library document information. For example, file format incompatibility often presents a problem for some recipients. In addition, electronic documents may pose a challenge to some groups of disadvantaged people such as the blind or visually impaired. To research solutions to such potential problems, the Communications Engineering Branch of the Lister Hill National Center for Biomedical Communications, an R&D division of NLM, has developed a web site called the DocMorph Server.[9] This is part of an ongoing R&D program in document imaging that has spanned many aspects of electronic document conversion and preservation, Internet document transmission and document usage.

2. The DocMorph Server

The DocMorph Server is an experimental web server whose services are publicly available. It is located at http://docmorph.nlm.nih.gov/docmorph on the World Wide Web, and serves two purposes:

1. It provides document image processing capabilities to the biomedical library community and to other visitors via the World Wide Web. This unique web site allows users to process electronic library documents in different ways, often without requiring additional software in their computers.
2. It allows in-house staff to research document imaging techniques and algorithms that could find application in future versions of the DocView software and other systems being developed at NLM. For example, the consequences of introducing modifications to in-house algorithms would be known within hours or days, allowing fast creation of useful and reliable software.

The three chief areas in which the DocMorph Server offers solutions to problems encountered in electronic document delivery are file format incompatibilities, document editing, and access to electronic literature by the visually challenged. The first area, file format incompatibilities, arises because document delivery librarians often encounter a variety of users and computers in their patron populations, especially on college campuses where it is common to find different types of computers running Windows, UNIX, LINUX, OS/2 and Macintosh operating systems. While the Ariel system used by many libraries delivers files in GEDI format, a superset of TIFF, it is not possible to find a single distributor of free TIFF viewers for all these platforms. A better solution for document delivery to a diverse patron population would be Adobe's Portable Document Format (PDF), since the Acrobat Reader that handles PDF is freely available for all these computing platforms. DocMorph provides an opportunity for librarians to deliver to a

diverse population because it allows them to convert their GEDI files to PDF. Likewise, a patron who does not have a TIFF viewer for his or her computer, but who can readily acquire Acrobat Reader, may use DocMorph to convert library documents from TIFF to PDF.

DocMorph also offers users a simple method of editing TIFF files that may not be available through their own TIFF viewers such as DocView. The ability to extract all pages from a TIFF file and to create a new multipage TIFF file allows users a method of inserting pages into, and removing pages from, a TIFF file. While DocView does not allow document editing, these two DocMorph functions permit existing DocView users to have a method of editing TIFF files without acquiring new software.

The third area where DocMorph offers a solution to problems encountered in electronic document delivery is in providing improved access to electronic documents by the visually impaired or blind individual. While there are a number of software tools available for those with vision problems, such as screen readers, magnifiers and optical character recognition software (OCR), none of these tools handle GEDI files. DocMorph offers the user a new option for reading certain electronic documents, particularly those in the GEDI and TIFF formats commonly delivered by libraries via the Internet.

The DocMorph Server provides four document conversion functions:
- *Concatenate one or more TIFF files to create a PDF file.* This function concatenates one or more GEDI files or monochrome (black and white) image TIFF files to create a single PDF file. This helps both librarians and patrons in cases where PDF is preferred over TIFF.
- *Concatenate TIFF files to create one multipage TIFF file.* This function concatenates single page or multipage TIFF image files to form one multipage TIFF file that contains all images. This is useful for creating new single file documents. For instance, single file pages can be concatenated to form a chapter, and chapters can be concatenated to create a book.
- *Split a multipage TIFF file into individual image files.* This function is useful for document editing, especially when combined with the previous function. By using the two functions together, a user may replace, delete or insert pages in a multipage TIFF file.
- *Computer-assisted reading.* By changing TIFF or text files to synthesized speech, this function allows the user's computer to read out loud the material the user sends to DocMorph. This function is potentially useful to enable people who are visually challenged to gain easier access to electronic literature. It has been estimated that in 1990 there were 550,000 Americans who were blind in both eyes and an additional 7.6 million who had vision impairments.[10] Blindness is defined as best-corrected visual acuity of 20/200 or worse in the better eye; visual impairment as best-corrected visual acuity of worse than 20/40 but better than 20/200 in the better eye.

The DocMorph Server was launched in May 1999. By the end of the first nine months of use, over 900 users registered to use the DocMorph Server. Of these, more than 500 actually used the system over this period. They submitted 5,200 jobs to DocMorph, containing more than 50,000 images or approximately 4.7 gigabytes of data. Of these, 4,800 jobs ran successfully. Most of the remaining 400 could not be processed, since their file formats (e.g., GIF and JPEG) are not currently supported by DocMorph. In some cases users submitted TIFF files that were defective, usually with corrupted TIFF headers. Of the 4,800 successful jobs, 3% were concatenation of TIFF files, 6% were splitting TIFF files, 10% were computer-assisted reading, and 81% were conversions of TIFF to PDF. Because the computer-assisted reading function was introduced into DocMorph six months after the web site was first launched, it is expected that its usage will increase with time.

3. DocMorph Server Architecture

The DocMorph Server is designed as an application server with a three-tier architecture. As illustrated in Figure 1, it consists of three desktop computers, each running Windows NT Server with the Microsoft Internet Information Server (IIS) running as the web server. There is sufficient room for expansion in the current system as usage increases, since the system architecture permits operation of up to ten computers. Each computer in the system is selected to handle specific conversion jobs. As shown in Figure 1, the main DocMorph computer handles job types 1, 2 and 3, corresponding to the TIFF to PDF conversion, TIFF concatenation, and TIFF splitting functions. The two other computers, named DocMorph1 and DocMorph2, both handle job type 4, which corresponds to the computer-assisted reading function. Users remotely located on the Internet may use their web browsers to upload document image files to be processed. The Job Switch on the main DocMorph computer senses the type of job requested by the user, and channels the request to the appropriate computer. The Job Switch contains a load-balancing algorithm that routes each user's job among the computers so that jobs are evenly distributed in the system, and no one computer receives an inordinate amount of work. At each computer the job to be run on that computer is entered into a Microsoft Access database. On a First In First Out (FIFO) basis, the Conversion Server takes each job in turn, does the conversion and returns the result to the user over the Internet via an HTML web page that it dynamically creates. Two hours after a user has downloaded converted files, the Conversion Server removes the converted files from the computer to free up disk space. The database is used to ensure proper job selection and timing, and it also keeps track of user statistics and feedback. While the main DocMorph computer contains a database called the Long Term Database, the other computers contain a Short Term Database. The difference between these databases is that the Long Term Database contains information about activity on all the DocMorph computers, whereas each Short Term Database contains job activity information relevant only to that specific computer. As each computer other than the main DocMorph processor finishes processing a job, it updates not only the Short Term Database, but also the Long Term Database on DocMorph. The advantage of consolidating all system activity on one centralized database is that it is easy to monitor status of the entire system.
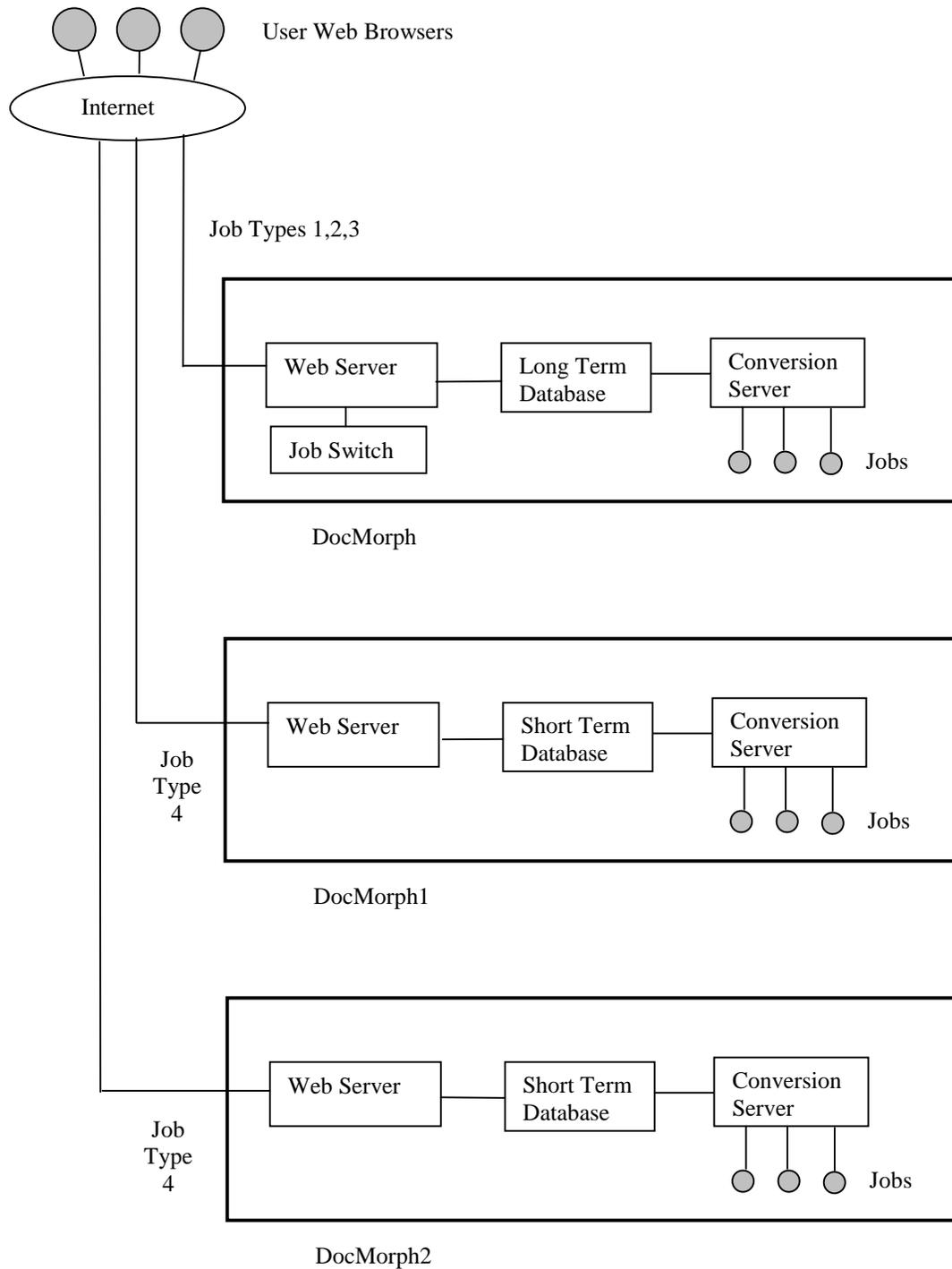
Figure 1. DocMorph Server Architecture

The software created for the DocMorph Server was written in several languages, including C, Basic, assembly language and JavaScript, with the large majority being in C++. Visual Basic was used to create front-end software for managing the dynamic creation of HTML web pages, and for handling user file uploads. While C++ was used to create the Conversion Server, some portions for image compression and expansion were created in C and assembly language. The user interface for the computer-assisted reading function uses JavaScript, which is dynamically created by the Conversion Server.

As mentioned, the ability to allocate specific job types to individual computers allows the system to operate in a balanced manner, without overloading any one computer. Depending on the conversion algorithm, a computer may require more time to run a job. As listed in Table 1, the average time to run a job for computer-assisted reading requires about twice the time required for other jobs.

Type of Document Conversion	Average Time for Conversion (seconds)
TIFF to PDF	11
Concatenate TIFF Files	9
Split TIFF File	11
Computer-Assisted Reading	23

Table 1. Average Conversion Times

OCR conversion is the cause for the large amount of time required for computer-assisted reading. This function accepts either text files or monochrome TIFF files. For text files, no OCR is required, and DocMorph quickly reformats the information and returns to the user a web page containing JavaScript code that allows the user's computer to read the information out loud. Processing of TIFF images takes longer, since on a 500 MHz computer the average processing time for OCR averages about 10 seconds for a typical biomedical journal page. With a typical journal article averaging 10 pages, the processing time could conceivably amount to 100 seconds. The average time of 23 seconds required by DocMorph reflects the fact that most job submissions to date are text files, rather than image files. Text files take an average of about 10 seconds to process, regardless of the file size.

4. Computer-Assisted Reading Design

DocMorph's computer-assisted reading function is the only one that requires a specific browser and computing platform. While the other three functions may be used with most common browsers and computers, the computer-assisted reading capability requires Microsoft Internet Explorer version 5 or higher, and a Windows platform such as Windows 95, 98, NT or 2000. Moreover, the user should have a sound card installed in the computer, and a method to hear the synthesized speech, such as speakers or earphones. Prior to using the function, the user should download the DocMorph speech software, which includes an ActiveX^{TM} control for speech synthesis that is distributed as part of the Microsoft speech software development toolkit. The installation of the

DocMorph speech software is a one-time event that does not need to be repeated. The web page that DocMorph produces contains a reference to the ActiveX control, which invokes the speech synthesizer on the user's computer to read the material submitted to DocMorph.

After DocMorph first receives a file from the user, it sends the user a web page that informs the user of job status. The web page updates itself every 20 seconds by querying DocMorph to determine whether the job has been completed. This query invokes DocMorph to check the system for job completion. DocMorph will return a web page to the user to indicate either a recheck after another 20 seconds, or that the user may begin reading the document using the speech synthesizer.

In the next step of the process, DocMorph produces an audio document in the form of an HTML web page that contains instructions for the synthesized speech ActiveX control. DocMorph automatically produces this web page, plus JavaScript code for providing the user interface. Figure 2 illustrates the procedural flow for processing a user document.
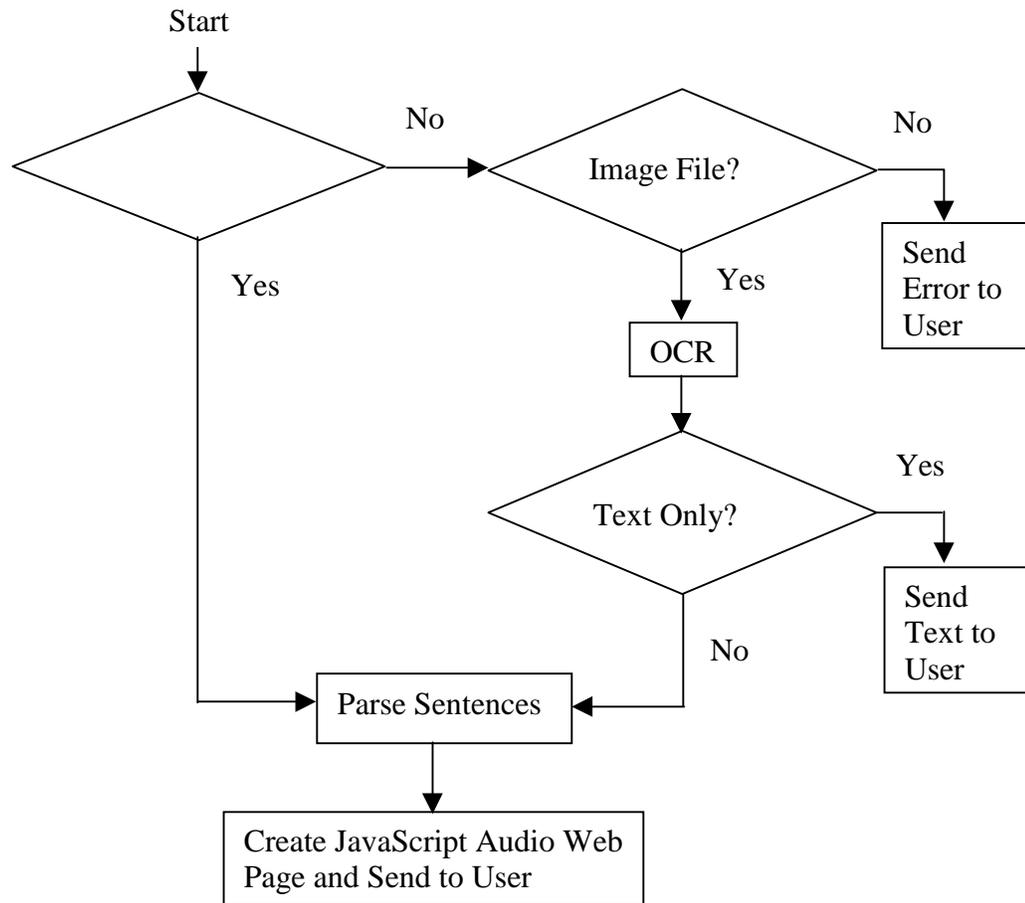

Figure 2. Procedural Flow for Conversion to Synthesized Speech

DocMorph first determines whether the file is a text file. If so, it parses the sentences and produces the audio web page. If the file contains no text, DocMorph then checks to see if the file is a GEDI or TIFF image file. If it is not, DocMorph returns an error message to the user. If the file is a legitimate image file that can be processed, DocMorph activates OCR software to convert the file to ASCII text. Then it parses the sentences and produces the audio web page.

If the user had exercised the option to specify that text should be produced rather than speech, DocMorph stops here and returns the ASCII text to the user. The advantage of receiving ASCII text is that it offers the user an opportunity to correct OCR mistakes. There are two potential error sources in the computer-assisted reading function: OCR conversion and speech synthesis. While the user is given a chance to correct the OCR errors, the speech synthesis software does not offer a method for correcting problems. After the user corrects the OCR errors, he or she can upload the text file and allow DocMorph to change this to synthesized speech.

The final step is for the user to download the audio document web page containing the text and JavaScript code. This invokes the speech synthesizer software on the user's computer and allows the user to "read" the document by listening to it. Table 2 lists the user interface functions that DocMorph offers for controlling the audio. As a bare minimum, the user only needs to use the space bar key to start or stop reading.

Keyboard Key	Function
Space Bar	Start and Stop
Page Down	Next Page
Page Up	Previous Page
J	Next Sentence
H	Previous Sentence
K	Repeat Last Sentence
F	Search for Text
R	Repeat Last Search
G	Go to a Page
C	Change Color
S	Change Size
Right Arrow	Read Faster
Left Arrow	Read Slower
Escape	Help
Browser Back Button	Process More Files

Table 2. User Interface Functions for Reading the Audio Document

Additional functionality includes the ability to page up or down, jump to a specific page, stop after each sentence, and to search for audio text. The user interface provided by the JavaScript code displays one sentence at a time in a window on the user's web browser. While the sentence is displayed, the speech synthesizer speaks the sentence. At the

completion of the sentence the software introduces a short delay, then it displays and speaks the next sentence. Features that are of particular interest to those with vision problems include the ability to change the color of screen text. Some individuals can see white text on a black background better than black text on a white background. In addition, users may change the size of the text and the reading speed.

Figure 3 contains a fragment of the DocMorph's JavaScript code to illustrate how it produces speech synthesis on the user's computer. At the top portion of the web page is an object named "ActiveVoice", which is a reference to the ActiveX software that controls the speech synthesizer. There are three sentences on the example page, and they are stored in an array of three elements. The function ReadSentence() controls the actual speaking, and it does this through the ActiveVoice.Speak() function by passing it a pointer to the sentence to be spoken.
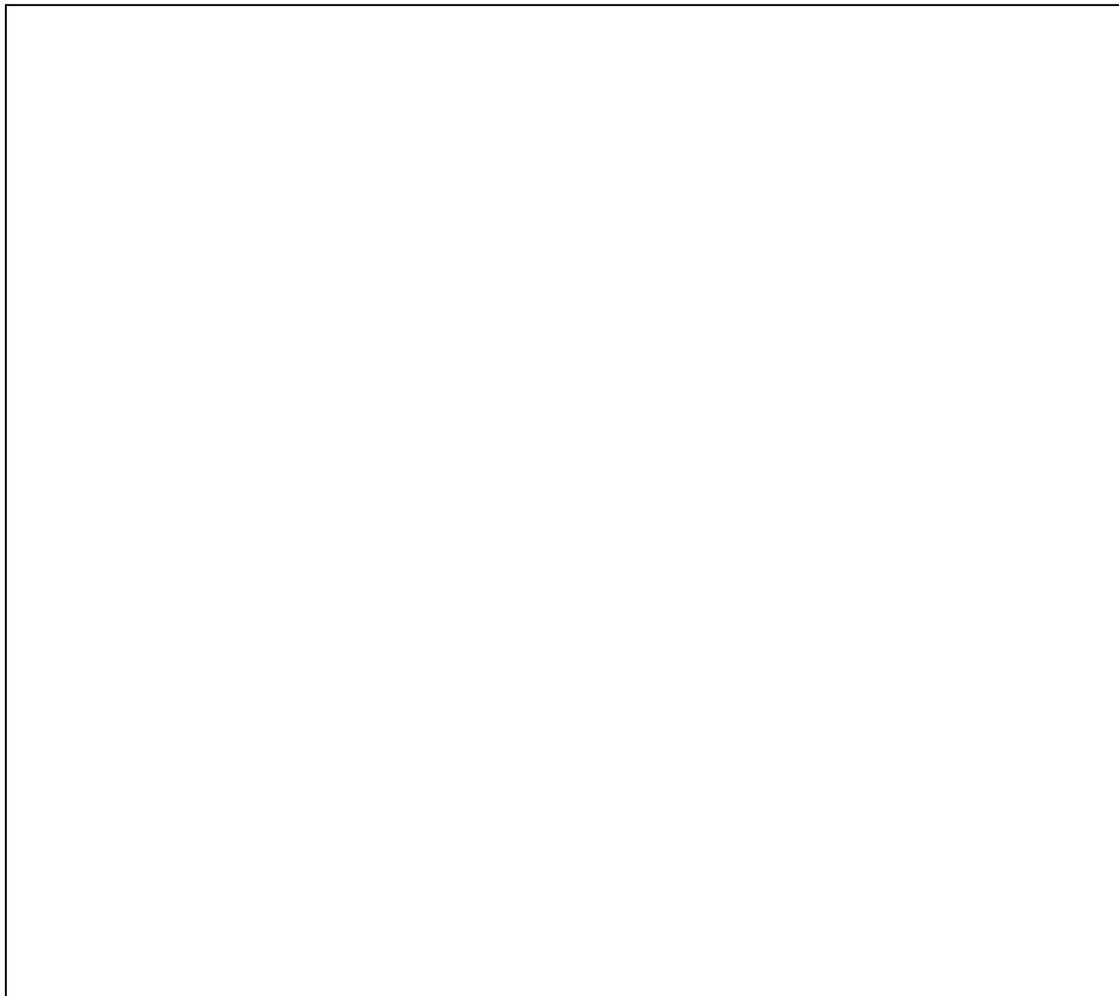

Figure 3. JavaScript Code Fragment

Once the user has downloaded the audio document from DocMorph, it may be saved on hard disk for future use. This allows repeated use of the audio document without the

need for redundant processing via DocMorph. In this manner the user may build an audio document collection on hard disk. Note that this capability is also useful for creating audio content on an Intranet. For instance, a library could create an audio-enabled portion of a website containing library information that could be accessible by library patrons.

5. Next Step: The Reading Room

DocMorph's computer-assisted reading facility is being enhanced to facilitate use by additional groups of potential users. It could prove to be of value to those without vision problems. For instance, non-readers of English may use DocMorph's computer-assisted reading facility to read documents if they can understand spoken English. Potential users in this category include foreigners, illiterate adults and children with certain types of reading difficulties. This function also allows a novel method for proofreading electronic documents; some individuals can recognize errors more easily through listening than through seeing. With this function DocMorph would allow authors to use the system as an aide in proofreading papers or books they may be writing. DocMorph could then be a useful resource not only to libraries delivering document literature and to the patrons who read it, but also to the creators of the literature. Once this new facility, called The Reading Room, is completed, it will provide unique user interfaces for each potential type of user. Other upcoming changes will include web pages that are friendly to screen readers, since the layout of the current web site's use of tables is not particularly suitable for screen readers used by many blind users. For those with difficulties reading English, the system may also include an improved user interface that includes word highlighting, so that each word is highlighted on the screen as it is spoken. These upgrades in DocMorph design will be evaluated through a user survey provided on the web site. Results of the user survey, which allows users to submit comments on their experiences with DocMorph, are stored in the system's Long Term Database for future analysis.

6. Summary

The National Library of Medicine's prototype DocMorph Server is based on a multi-computer architecture, and provides functions for document image processing. Publicly accessible through the World Wide Web, DocMorph allows users to upload image files for processing, and returns results that users may store on their computers. Current functionality includes file conversion such as TIFF to PDF, file editing, and conversion of scanned images of printed literature to synthesized speech. This paper has described the system architecture and the design method for implementing the conversion to synthesized speech. While the DocMorph Server is designed to provide a useful service to the public, it is also an image processing test platform that its designers are using to evaluate algorithms and techniques for possible inclusion in other image processing software products being developed at NLM.

7. References

1. Berger, M.A., "Ariel Document Delivery and the Small Academic Library," College & Undergraduate Libraries, Vol. 3(2). The Haworth Press, 1996; 49-56.
2. The World Wide Web address for Research Libraries Group is located at this URL: http://www.rlg.org.
3. Walker FL, Thoma GR, "DocView: Providing Access to Printed Literature through the Internet," Proceedings IOLS'95. Medford NJ: Learned Information, 1995; 165-173.
4. Walker FL, Thoma GR, "Internet Document Access and Delivery," Proceedings IOLS'96. Medford NJ: Learned Information, 1996; 107-116.
5. Postel, J. and Reynolds, J. File Transfer Protocol, Request for Comments #959, October 1985, available at URL http://sunsite.auc.dk/RFC/ .
6. Borenstein, N. and Freed, N. MIME (Multipurpose Internet Mail Extensions), Request for Comments #1341, June 1992, available at URL http://sunsite.auc.dk/RFC/
7. Information on the Group on Electronic Document Interchange (GEDI) format is available at URL http://lib-www.uia.ac.be/MAN/T02/t51.html.
8. TIFF Revision 6.0, Aldus Corporation, June 3, 1992.
9. Walker, FL and Thoma, GR, "Web-based document image processing," Proceedings of IS&T/SPIE Conference on Internet Imaging, San Jose, California, January 2000, 268-277.
10. "Trends in the Health of Older Americans: United States, 1994," National Center for Health Statistics, Center for Disease Control and Prevention, April 1995. Available at this URL: http://www.cdc.gov/nchswww/data/sr3_30.pdf .