

A SOAP-Based Tool for User Feedback and Analysis

Frank L. Walker (walker@nlm.nih.gov)
and George R. Thoma (thoma@nlm.nih.gov)
National Library of Medicine

Keywords: MyMorph, DocMorph, SOAP, Web Service, World Wide Web, TIFF, PDF, DocView, NLM, Internet

Abstract: Simple Object Access Protocol (SOAP) is a new technology utilizing extensible markup language (XML) sent over the Internet via the Hypertext Transfer Protocol (HTTP). SOAP was used to implement a tool to elicit user feedback on a prototype web service developed at the Lister Hill National Center for Biomedical Communications, an R&D division of the National Library of Medicine (NLM). The service, called MyMorph, allows an end user to convert files in more than fifty different formats to the Portable Document Format (PDF). The MyMorph service uses client software running on the user's computer to send files via SOAP to a computer system at NLM called DocMorph, which converts them to PDF and returns the results to the user's computer. MyMorph can support document delivery, as well as file migration for electronic document preservation. To help determine its usefulness, the MyMorph software included an optional user survey that allowed users to provide feedback to the developers via SOAP. This paper describes how the SOAP-based survey was implemented, and analyzes the survey results.

1. Introduction

In May 1999 the National Library of Medicine's Lister Hill National Center for Biomedical Communications launched the DocMorph web site as part of an ongoing investigation into the issues of delivering, processing and using electronic library information. DocMorph, located on the web at <http://docmorph.nlm.nih.gov/docmorph>, enables remote users with web browsers to upload files for processing in any of five ways.¹ First, it can convert files in any of more than fifty format types to the Portable Document Format (PDFTM). These file types include black and white images, grayscale and color images, and word processing files. Second, it can convert any of these file types to TIFF² images. To support TIFF editing, DocMorph includes a third function for splitting a multipage TIFF file into single TIFF images. DocMorph's fourth function is for extracting text from any of the file types it processes: in 1999 it became the first publicly available web site to offer image-to-text conversion via optical character recognition. Finally, as a research tool for improving accessibility to library information, DocMorph has a function for converting files to synthesized speech.³ In its first 3 years of operation, more than 5,500 people registered to use DocMorph, and submitted to the system more than 67,000 jobs consisting of 65 GB of data, or about 700,000 document pages. About 75 percent of the jobs submitted to DocMorph were for conversion to PDF.

While DocMorph provides useful functionality to its user community, it also serves as an R&D tool for our design team to investigate the utility, reliability and speed of image and information processing algorithms. For instance, its algorithms have been steadily refined to accommodate the varieties of file formats submitted by users. While studies have shown that DocMorph takes an average of less than 20 seconds to process each submitted job, the time spent by the user can be significant. Depending on the size of the file submitted and the speed of the user's Internet connection, a typical job could take several minutes for uploading to DocMorph, processing, and downloading. Because the web browser interface to DocMorph requires the user's attention while jobs are processed, this can affect user productivity. To address the user interface problem, in June 2002 DocMorph's design team released for beta testing a new software program called MyMorph that runs under Windows-based operating systems. It allows the user to easily select large numbers of files for conversion to PDF. MyMorph automatically handles the uploading of files to DocMorph, waiting for results, and then downloading the PDF files to the local disk drive. This automation allows user interaction to be minimal, taking as little as ten seconds of a user's time for conversion of a batch of files.⁴

MyMorph has several potential uses. First, it serves as a model for bulk file migration for the purpose of electronic document preservation. MyMorph makes it very easy to convert thousands of electronic files from formats that may be decreasingly used (e.g., TIFF) to a format that is becoming increasingly more popular because it can be used on all the major computing platforms (e.g., PDF). In addition to its role in file migration, MyMorph is also useful for supporting document distribution. It is being used daily by document delivery librarians who need to convert scanned images to a format that is portable and easily distributed. In addition, individuals will often use it on a one-time basis to convert a research paper, for example, from Postscript to PDF.

2. The MyMorph Beta Test

From its release for beta testing in June through December 2002, more than 950 users downloaded the MyMorph client software from the DocMorph web site, installed it on their computers and registered the product. Table 1 shows the number of jobs submitted monthly by MyMorph users to DocMorph.

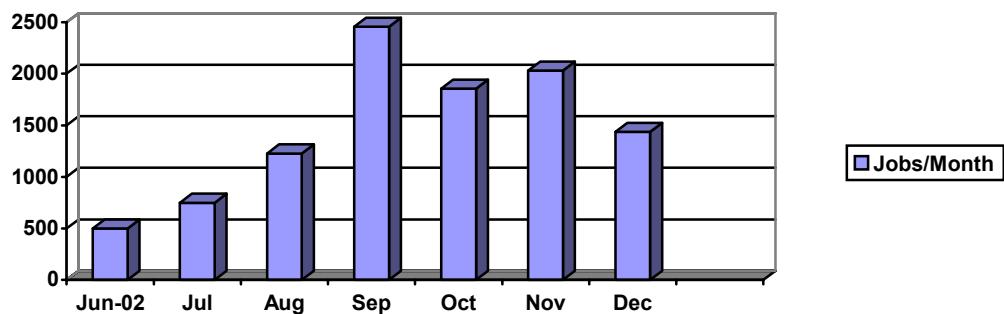


Table 1. MyMorph Usage

One consequence of introducing MyMorph was that fewer users submitted jobs to DocMorph through their browsers. Prior to MyMorph's release, about 75 percent of jobs submitted to DocMorph were for conversion of files to PDF. Table 2 shows the relative changes in system usage for browser-submitted PDF conversions versus MyMorph-enabled PDF conversions. After three months into the beta test, more PDF conversions were being done through MyMorph than through web browsers.

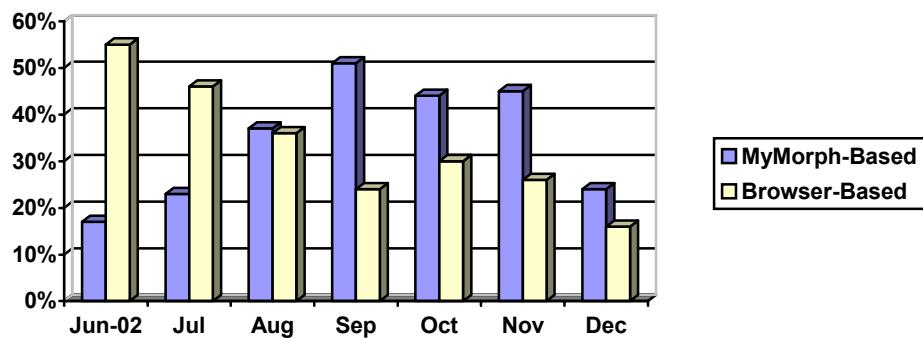


Table 2. Percentage of Total DocMorph Conversions

The MyMorph beta test provided data to our design team on the reliability and ease of use of the software, through problems that users reported. During this 6-month period users reported several problems, most of which were addressed through new releases of the product. As of January 2003, MyMorph, in its sixth beta release, was a more stable software program that was easier to install and use than the initial release. The beta test was also an opportunity to gather user opinions on the software to help guide future work in this area. Two methods were used for gathering user feedback during the beta test. First, users were given an email address to which they could report problems, especially those that were serious. During the first two months of beta testing, about 8% of users encountered serious problems preventing use of the software. For these users, email was the only but effective means for providing feedback. The second technique for gathering user feedback employed a user survey tool built directly into MyMorph. Thirty days after the software was installed on the user's computer, MyMorph prompted the user to answer the survey questions. About ten percent of MyMorph users opted to answer this optional survey, which consisted of fourteen questions, and took about ten minutes to answer, after which MyMorph sent the answers via the Internet to DocMorph, where they were gathered for analysis.

3. The Tool Design

There are a number of techniques for conducting user surveys. Some are conducted in person or through postal mail, while others are done electronically, such as telephone, fax, email or through a web site. The technique chosen for MyMorph's user survey was similar to one used in the DocView project several years ago at NLM. DocView is software we developed that is a tool for end users to receive documents sent over the Internet by Ariel systems.^{5,6} During the DocView beta test, users were prompted by an electronic survey on their computer 45 days after the software was installed.⁷ After the

user completed the survey, DocView forwarded the answers via File Transfer Protocol (FTP) over the Internet to an FTP server at NLM, where they were collected for analysis. Instead of using FTP for transmitting survey results, MyMorph used Simple Object Access Protocol (SOAP). SOAP was a logical choice since it was already built into MyMorph for communications with DocMorph. It is an Internet protocol that allows a pair of computers, typically a client and a server, to communicate with each other using Extensible Markup Language (XML) commonly sent via Hypertext Transfer Protocol (HTTP). The advantage that HTTP has over FTP as a delivery method is that HTTP normally passes through firewalls, whereas FTP may not. If either client or server is protected by a firewall, the client is far more likely to achieve a successful connection with the server using SOAP than FTP. Hence, SOAP was judged to be the preferred method for delivering the survey data.

The MyMorph design team used Microsoft's SOAP toolkit version 2 to implement the communications between MyMorph and DocMorph. For this application, SOAP consists of an XML message exchanged via HTTP between the MyMorph client and the DocMorph web site. A SOAP message, as shown in Figure 1, consists of three parts:

- The Envelope: a top-level container for the message
- The Header: containing added features for the SOAP message, and
- The Body: containing information for the recipient of the message.

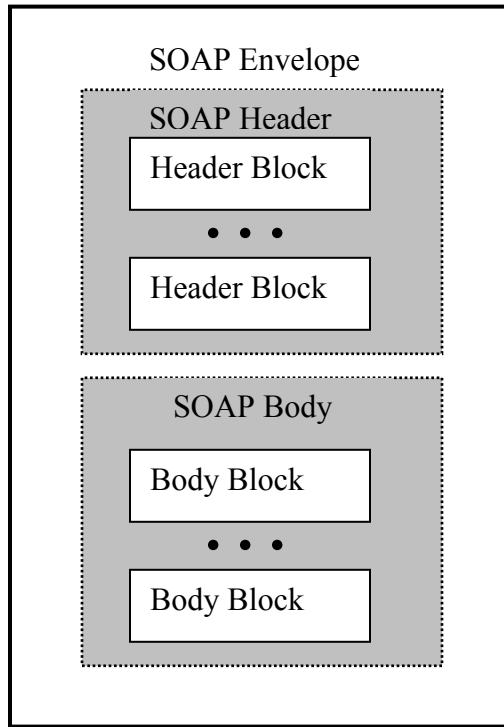


Figure 1. SOAP Message

A set of five functions was designed for controlling the SOAP communications between the MyMorph client and DocMorph. These include four basic functions for submitting the user registration, determining the current version of MyMorph software, obtaining permission to submit a job to DocMorph, and a function for submitting the job. The fifth

function, SendSurvey, is run only once – after the user has answered the survey questions. The information in the SOAP message sent from the MyMorph client software to DocMorph contains three elements:

UserID – This is a 32-character field used to uniquely identify each received survey.

Answer – This is a set of fourteen 16-bit integers containing the answers to the fourteen survey questions.

Comments – This is a field containing the user's general comments.

After the user completes the survey, MyMorph sends to DocMorph a SOAP message consisting of these three information elements. Once DocMorph receives the information, it responds by returning an acknowledgement handshake to the MyMorph client.

The following is a typical SOAP message sent from MyMorph to DocMorph for the SendSurvey function:

```
<xml version="1.0" encoding="UTF-8" standalone="no"?>
<SOAPENV:EnvelopeSOAPENV:encodingStyle="http://schemas.xmlsoap.org/soap/encoding/">
<ns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope"/>
<SOAP-ENV:Body>
  <SOAPSDK1:SendSurvey xmlns:SOAPSDK1="http://tempuri.org/message/">
    <UserID>74385072483376763968361100416094</UserID>
    <Answers xmlns:SOAPSDK2="http://www.w3.org/2001/XMLSchema">
      <SOAPSDK3="http://schemas.xmlsoap.org/soap/encoding/">
        <SOAPSDK3:arrayType="SOAPSDK2:short[14]">
          <SOAPSDK4="http://www.w3.org/2001/XMLSchema-instance">
            <SOAPSDK4:type="SOAPSDK3:Array"><SOAPSDK3:short>2177</SOAPSDK3:short>
            <SOAPSDK3:short>0</SOAPSDK3:short><SOAPSDK3:short>5</SOAPSDK3:short>
            <SOAPSDK3:short>0</SOAPSDK3:short><SOAPSDK3:short>2</SOAPSDK3:short>
            <SOAPSDK3:short>2</SOAPSDK3:short><SOAPSDK3:short>0</SOAPSDK3:short>
            <SOAPSDK3:short>0</SOAPSDK3:short><SOAPSDK3:short>0</SOAPSDK3:short>
            <SOAPSDK3:short>0</SOAPSDK3:short><SOAPSDK3:short>0</SOAPSDK3:short>
            <SOAPSDK3:short>1</SOAPSDK3:short><SOAPSDK3:short>0</SOAPSDK3:short>
            <SOAPSDK3:short>2</SOAPSDK3:short></Answers><Comments>MyComments.
        </Comments>
    </SOAPSDK1:SendSurvey>
  </SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

The survey questions were grouped into five general sections:

1. *Job Function*. This section determined the general occupation of the user.
2. *Your Internet Connection*. This section determined the user's Internet connection type and its speed.
3. *Ease of Use*. This section consisted of three questions to determine whether the user found MyMorph easy to use.
4. *Impressions of MyMorph*. This consisted of four questions designed to elicit user opinions about the software.

5. *Future Enhancements.* This consisted of four questions to determine what types of enhancements should be made to MyMorph, plus a comments section for the user to enter general comments on the software.

4. User Feedback Analysis

Once DocMorph received the completed survey from a MyMorph client, it stored each in a database for conveniently summarizing and analyzing the results. During the first 6½ months of beta testing, user surveys were received from 81 MyMorph users, and the results are presented here. The first section in the survey, Job Function, determined the general occupation of the user as shown in Table 3. For users whose jobs spanned multiple occupations, the survey permitted the user to select more than one. This accounts for the total percentages exceeding 100%. Most users were employed in libraries (38%) or education (36%). Large numbers were employed in health (30%) and research (30%), and nearly one-fourth was employed in document delivery. Obviously these are not mutually exclusive categories.

Occupation	Percentage of Users Responding
Library	38%
Education	36
Health	30
Research	30
Document Delivery	23
Administrative	22
Technical	22
Student	10
Government	9
Industry	9
Other not given here	6
Publisher	3
Legal	3
Secretary	2

Table 3. User Occupations

The second area of the user survey, Your Internet Connection, determined the type of connection the user had in addition to the speed. The results show that 14% of the users had dialup connections (using a standard modem), and 86% had direct connections (provided by local area network, DSL or cable modem). Table 4 lists typical connection speeds for each user. It shows that nearly one-third did not know their connection speed, while more than one-third reported speeds ranging from 128 Kbs to T1 (1.544 Mbs).

Connection Speed	Percentage of Users
0 to 19 Kbs	0
19.2 to 33 Kbs	17%
33 to 56 Kbs	2
56 to 128 Kbs	7
128 to T1	36
Faster than T1	7
Unknown	30

Table 4. Internet Connection Speeds

The three remaining parts of the survey consist of statements, for which the user could select an answer from a list of choices. The Ease of Use section consists of three statements. Table 5 summarizes the responses. It shows that a large majority of users felt that MyMorph was both easy to learn to use and that its messages to the user were easy to understand. One statement was designed to determine whether MyMorph was easier to use than browser-based DocMorph. About 27 percent of the responders said this was not applicable; in other words, they had never used DocMorph. For the remaining users who had used both DocMorph and MyMorph, 85 percent either strongly agreed or tended to agree that MyMorph was easier to use than DocMorph.

Statement	Strongly agree	Tend to agree	Hard to Decide	Tend to disagree	Strongly disagree	Not applicable
MyMorph was easy to learn to use	83%	12%	4%	1%	0	-
MyMorph messages were clear in their meaning	80	14	4	2	0	-
The software was easier to use than DocMorph	49	12	10	1	0	27

Table 5. Ease of Use

The next survey section gathered feedback on user impressions of MyMorph. As shown in Table 6, nearly all users, 96 percent, felt positive about the usefulness of the software. Most users (90 percent) also felt that the software was reliable. Large numbers of users

also felt the time taken to process the files they submitted was fast. Finally, 88 percent said they would recommend the software to colleagues.

Statement	Strongly agree	Tend to agree	Hard to Decide	Tend to disagree	Strongly disagree	Not applicable
The software was useful to me	90%	6%	4%	0%	0%	-
The software was reliable	89	1	7	1	0	-
The time taken to process my files was fast	81	10	4	2	2	-
I would recommend this software to my colleagues	86	2	10	0	1	-

Table 6. Impressions of MyMorph

The Future Enhancements section of the survey, shown in Figure 7, elicited opinions on what users considered most important in future versions of the MyMorph software. The MyMorph design team listed four areas for potential enhancement. Users were allowed to enter additional opinions in a text comments field at the end of this section. While users were lukewarm to one of the suggested enhancements, they were more enthusiastic about the remaining three. Only 61 percent of the users agreed that converting files at user-specified times of the day would be desirable. A larger number, 82 percent, were positive about adding a folder monitoring capability. This feature would allow MyMorph to monitor a user-specified folder on the local disk drive, and convert files to PDF as soon as they appear in the folder. This feature could be useful in a document scanning operation where large numbers of images require conversion to PDF. A slightly larger number of users, 85 percent, felt that adding a new delivery capability would be useful. While only 22 percent of the users reported having jobs in document delivery, it is interesting to note that a much larger number recommended combining conversion with delivery. This implies that many users need to deliver PDF files after creation, other than those professionally engaged in document delivery. This could be because of the nature of PDF: it is a file format that is easily distributable. Finally, about 86 percent of users felt positive about adding a scanning capability. This implies that their current scanning

program delivers images that require conversion to PDF as a second step; combining the two functions of scanning and PDF conversion into one program could make the entire process easier or faster.

Statement	Strongly agree	Tend to agree	Hard to Decide	Tend to disagree	Strongly disagree	Not applicable
Add a folder monitoring capability	67%	15%	12%	4%	2%	-
Convert files at user specified times of day	49	12	28	7	2	-
Add a new delivery capability	75	10	12	1	1	-
Add a new scanning capability	77	9	11	2	1	-

Table 7. Future Enhancements

Remarks in the general comments section of the survey showed that most users felt the MyMorph software was highly worthwhile. Several users wanted the additional capability of combining multiple files to create a single PDF file. This was a feature in the web-based DocMorph not designed into MyMorph. One user recommended a feature to include text-searchable PDF. Several users reported problems with converting color word processing documents to PDF; we found that DocMorph was inadvertently eliminating color during the file conversion. In parallel with the beta test the MyMorph design team redesigned the software to permit Postscript, word processing, and spread sheet files to be converted into text-searchable PDF files. Also, the capability of correctly handling color files was built into DocMorph as a direct result of user feedback obtained through the survey.

7. Summary

This paper has described the application of a new technology, SOAP, for implementing a tool for eliciting user feedback. During the beta test of a SOAP-based client program for bulk conversion of files, user feedback was collected by both email and a SOAP-based survey. While email feedback allowed users with serious problems to report them immediately, the built-in survey allowed users with fewer problems to provide in-depth feedback after using the system for a period. The survey results showed that a large

percentage of the users who participated in the survey felt the MyMorph software was fast, reliable, useful, and easier to use than browser-based DocMorph. The results have allowed the design team to correct problems and plan future enhancements of the MyMorph software.

8. References

1. Walker FL, Thoma, GR. Web-based document image processing. Proceedings of IS&T/SPIE Conference on Internet Imaging, San Jose, California, January 2000, 268-277.
2. TIFF Revision 6.0, Aldus Corporation, June 3, 1992.
3. Walker FL, Thoma GR. Read It To Me! Proceedings of the Twenty First National Online Meeting, May 16-18, 2000. Medford NJ: Information Today, 2000; 473-483.
4. Walker FL, Thoma GR. A SOAP-Enabled System for an Online Library Service. Proceedings of InfoToday 2002. Medford N.J.: Information Today, 2002, 320-329.
5. Walker FL, Thoma GR. DocView: Providing Access to Printed Literature through the Internet. Proceedings IOLS'95. Medford NJ: Learned Information, 1995; 165-173.
6. Berger MA. Ariel Document Delivery and the Small Academic Library. College & Undergraduate Libraries, Vol. 3(2). The Haworth Press, 1996; 49-56.
7. Walker FL, Thoma GR. Internet Document Delivery: An End User Survey. Proceedings IOLS '97. Medford N.J: Information Today, 1997; 145-153