

A Secure Internet Service for Delivering Documents for the Blind

Benoit Guillon¹, Dominique Burger¹, and Bruno Marmol²

¹ Université Pierre et Marie Curie B23, INSERM U483,
75252 Paris Cedex, France
{benoit.guillon,dominique.burger}@snv.jussieu.fr,
<http://www.snv.jussieu.fr/inova/>

² INRIA Rhône-Alpes,
38334 Saint Ismier, France
bruno.marmol@inrialpes.fr

Abstract. The access to written information is essential for the inclusion of individuals in modern societies. At school, at work it is an important success factor. At home it is source of pleasure and cultural development. In this paper we describe a service that has been developed to improve the cooperation between the different actors involved in producing an distributing books in alternate formats for visually impaired persons.

1 Introduction

Different techniques can be used to produce documents for visually impaired persons : Large print, Braille, audio recording, digital audio, electronic publishing. In any case the format of the original book has to be modified in an alternate accessible format. The adaptation of documents involves a co-operation between three types of actors (figure 1)

- original publishers which own the Intellectual Property Rights (IPR) of the document. In many countries their agreement is mandatory for any adaptation of the original document, even if nor the content or the meaning are modified.
- adaptation centres which have the necessary expertise in producing alternate accessible formats. These centres ask permissions to original publishers. In some cases they obtain the electronic files of the documents in order to process them on computers. Finally, adapted documents are distributed to end users
- In some cases specialised printing centres use electronic files produced by adaptation centres to print the versions in Braille or large print [1].

In spite of efforts made by these organisations, the needs for adapted documents are not fulfilled. Most of the books that are published every year have no chance to be read by visually impaired people. The reasons are numerous :

- Very often adaptation centres and specialised printing centres are small sized organisation with a limited production;
- Original publishers give permissions too slowly - if not reluctantly.
- When provided by publishers, files of books are often in electronic formats that cannot be processed easily;
- Coordination is not sufficient. It is often observed that the same book is adapted simultaneously by different organisations.

The Internet and electronic publishing provide an opportunity to create a cooperative framework where :

- Adaptation and printing centres could cooperate;
- The adaptation process would be more efficient;
- Resources could be shared;
- IPR would be better guaranteed.

The masterpiece of this framework is a secured Internet server, Helene, which has been developed jointly by the INRIA and the University Paris 6 in France. It has been tested out in 2001 and 2002 with users in real situations and several aspects have been enhanced accordingly. The server Helene is currently run by the BrailleNet association.

This paper describes the features of this server.

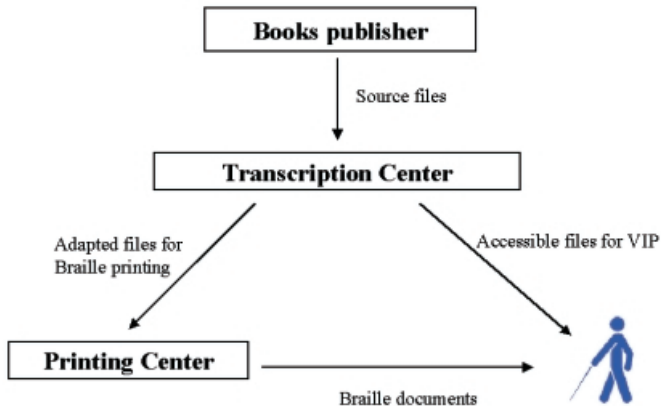


Fig. 1. Adapting books for visually impaired users.

2 Architecture

The figure 2 illustrates the main functions of the Helene server which are 1) to store files securely and 2) to deliver them to authenticated users over the

Internet. In this section we briefly describe the different parts of the Helene server.

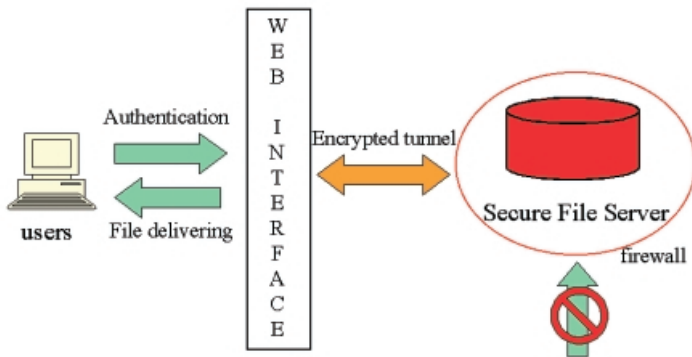


Fig. 2. General Architecture of the Helene Server

The Helene server is composed of two UNIX servers:

- A Document Server is used to store the files under copyright. It is insulated from the Web by a firewall based on the ipchains software provided by the GNU/Linux 2.2.x kernel. The integrity of the server is periodically checked by the tripwire software.
- A Public Web Server provides the interface between end-users and the Document Server. It provides a public catalogue describing the documents. Requests can be formulated according to several criteria. Bibliographical data are described in XML files, indexed in a MySQL database. Perl and CGI (Common Gateway Interface) have been used to implement the functionality of this catalogue, including : administration facilities (monitoring access to files), advanced search, a newsletter, ...

This Public Web Server also manages HTTP authentication of users who request secured files. Authentication is made on basis of a login and a password encapsulated into HTTP requests sent from the user to the Web Interface. The user interface is made of HTML pages in accordance with the WAI accessibility Guidelines [2].

The two servers uses a protocol HTTPS (HTTP over SSL [5]) to communicate in order to reduce the vulnerability in the exchanges. The SSL is a protocol providing a practical, application-layer, widely applicable connection-oriented mechanism for Internet client/server communications security. Note that this standard is widely used in online banking and online payment on many commercial Website.

3 Access Rights

Different types of files can be stored on the Document Server with different levels of security:

- Files of books in the Public Domain. No security is needed for their delivery. They are considered as having the lowest security (level 0).
- Files prepared for Braille print. These files have been processed for Braille editing so that information concerning the original layout, images, has been lost. Moreover the coding of the document uses specific Braille codes. Such files cannot be used to recover the original book format. They need a rather low level of security (level 1).
- Source files provided by the publishers. These files have to be secured at the highest level (level2).

Different group of users have been considered, each being allowed to achieve different operations:

- Administrators have full access to all documents (level 2) and administration front-end.
- Adaptation centres need to access source files provided by publishers in order to adapt them (level 2).
- Specialised printing centres access only files pre-processed for special print (level 1)
- Single end-user can access documents at the level 0.

4 Secured Documents Delivering

In order to respect the IPR of documents, files have to be sent confidentially to authenticated users only. In our implementation of Helene, this implies different steps :

- Authorised users are registered. They receive a digital certificate signed by the Helene Server, acting as certification authority, and including a private key. A public key corresponding to this user is also stored on the server.
- Users are authenticated. When a user sends a request to the the Document Server via the web Interface, after checking the password, the server verifies that the login of the user corresponds to a registered certificate.
- Before to be sent documents are encrypted. The Helene server uses asymmetric encryption, based on a pair of keys (public and private). The encryption is performed using the public key of the authenticated user. The encrypted document is provided to the user.
- The authenticated user opens the document. To open the document the digital certificate, including the private key, is needed. A password is also necessary to use this certificate.

E-mail constitute a simple method for delivering electronic data to targeted users. This is the reason why S/MIME [6] has been chosen to implement the whole process (S/MIME is a normalised extension of MIME), using openssl [4], an open source toolkit containing most of the existing cryptographic methods:

- public/private key generation, encryption, decryption, signature, ...
- X.509 certificate management,
- experimental SSL/TLS,
- S/MIME,
- digest generation (MD5, SHA-1, ...),
- S/MIME e-mail generation and verification, ...

The e-mail generated by openssl can be decrypted, verified and parsed by the most widely used software, like Microsoft Outlook, Netscape, Mozilla Mail, ...

5 Interoperability

Such a server could exist alone, but rapidly needs have appeared to interconnect it with other existing servers. For instance, the Helene server had to receive answer to requests coming from a server running a national catalogue of adapted books managed by the INJA (Institut National des Jeunes Aveugles). The Web Interface is a suitable place for running such dialogues between servers. It can also be used to send a request to several databases and to formulate an unified answer.

Two different solutions have been explored and implemented in the Helene Server, to manage this type of interconnection:

- Z39.50 [3] is a protocol for text-searching and retrieval, deployed in application domains as libraries. Originally an American (ANSI/NISO) standard, it is now ratified as international standard ISO 23950.
A Z39.50 client sends queries to a Z39.50 server and recovers matching records according to a protocol. It offers platform-independent interoperability, continuous segmented transmission of large result sets (unlike HTTP), item sorting, and a large set of other services. Z39.50 can be easily implemented over an existing program using specialized API available in most programming languages.
- XML via CGI. CGI scripts that are widely used to generate HTML contents can also return XML files when they receive a query. This method offers more flexibility than Z39.50 and does not require the installation of a standalone server. But it is limited by the HTTP protocol when retrieving large result sets and requires that the two parts agreed on a particular DTD. The returned XML file can be parsed using standard API available in most programming languages.

6 Conclusion and Perspectives

The Helene server has been opened in september 2001. Agreements have been signed with 22 publishing companies for the provision of more than 600 titles. Currently around 20 transcription and specialised printing centres have been certified and have used the Helene Server with success. This demonstrate the operability of the solution we are proposing and its the technical validity. The server Helene appears to be a suitable answer in the context of countries like France where legal restrictions on producing adapted books are severe. Nevertheless, the number of titles currently available should increase considerably to make the Helene server a real service. This objective could be reach if publishers would provide contents systematically. Discussions have started with their national representatives to obtain an extension of the agreements we have with a few to a national frame agreement.

Acknowledgements. The development of the Helene server could not have been achieved successfully without the support of the French Ministry of Culture, the Institut National des Jeunes Aveugles, the Fédération des Aveugles de France and the companies Hewlett-Packard and Alcatel.

References

1. BUGER D. et al. (1999)
A solution to the copyright barrier to access to information, CSUN 1999
http://www.snv.jussieu.fr/inova/publi/csun_99.htm
2. WAI: Web Accessibility Initiative
<http://www.w3.org/WAI/>
3. Z39.50 protocol
<http://www.loc.gov/z3950/agency/>
4. OpenSSL Project
<http://www.openssl.org>
5. SSL (Secure Socket Layers)
<http://www.netscape.com/security/techbriefs/ssl.html>
6. S/MIME (Secure / Multipurpose Internet Mail Extension)
<http://www.imc.org/smime-pgpmime.html>