Title	Digital archives using XML description and application to historical researches
Author	Shibayama, Mamoru / Yoshii, Yoshikuni / Goto, Makoto / Raghavan, Venkatesh
Citation	大阪市立大学学術情報総合センター紀要. Vol. 4, p.1-9.
Issue Date	2003-03
ISSN	1345-4145
Type	Departmental Bulletin Paper
Textversion	Publisher
Publisher	大阪市立大学学術情報総合センター
Description	

Placed on: 大阪市立大学学術機関リポジトリ

Placed on: Osaka City University Repository

Digital Archives using XML Description and Application to Historical Researches

Mamoru SHIBAYAMA[†], Yoshikuni YOSHII[†],
Makoto GOTO[‡], and Venkatesh Raghavan[†]

Abstract: This paper describes an optimal content delivery in the digital archives for 4 documents ("Japanese Economic History", "Fushimiya Zenbei", "Osaka Official Notices", and "Mori Collection") using the XML(eXtensible Markup Language) description in Edo era in Osaka. Further, we discuss various aspects relating to the development a "virtual library" which employs an interactive environment for exchanging opinions and comments in the research activities. A "virtual bookshelf" feature with walk through is also available for document retrieval. Several user-interfaces related to the "virtual library" are presented and the simultaneous retrieval for these documents based on Z39.50 standard is also discussed.

Further, this paper describes attempts to apply the XML/XSLT (eXtensible Style-sheet Language Transformations) for describing the process of restoration in "Shoso-in Monjo" which handed down from the Shoso-in of Todaiji Temple from the 8th century. This document exhibits an unique dissimilarity between their logical structure and physical layout due to transcriptions on both sides of paper and also "rearrangement" in 19th century. The database system using XML/XSLT allows users to examine a procedure of restoration of the "Shoso-in Monjo" and a virtual recreation of the original document called "Tanzaku" as a MS-Word document.

XML記述によるデジタルアーカイブと歴史研究への応用

柴山　守[†]、吉井良邦[†]、後藤　真[‡]、ベンカテッシュ・ラガワン[†]

近世大坂の4文書（『日本経済史資料』、『伏見屋善兵衛文書』、『大坂町触』、『森文庫』近世関連)のデジタルアーカイブはXML(eXtensible Markup Language)記述で構築し、電子図書館機能の「近世資料室」として一般に提供している。この「近世資料室」では、利用者間でWebを介して、研究活動における研究メモやコメントが交換できる環境、バーチャル書架による検索、ウオークスルーを「バーチャル図書館」の機能として実現した。また、情報検索Z39.50標準を適用して4文書の統合検索を実現し、XMLのリンク機能で検索結果から「バーチャル図書館」の機能が利用できる。

また、8世紀から東大寺の正倉院に伝来する正倉院文書の復原過程にXML/XSLT(eXtensible Style-sheet Language Transformations)を適用した事例について述べる。本文書は、19世紀初頭からの「整理」作業のために「帳簿」の形態が失われ、物理構造と論理構造の差異という特徴を持つ。利用者は、このXML/XSLTに基づくデータベースにより、復原過程での手続きをXSLTで整合性を調べ、論理構造を適宜表示すると共に、「短冊」と呼ぶ物理構造を表現する文書が復原できる。

1. Digital Archives and Integrated Retrieval Based on Z39.50 Protocol

Recently, the digitalization material such as historical document and the valuable books is being carried out by many organizations aiming to develop electronic libraries and museums. Such digitalization efforts often fail to adequately address important issues concerning document preservation and digital content delivery. Moreover, even in the case of providing access via WWW, these digital presentations provide merely a photographic record of the original materials.

In this paper, we describe the process for digitalization of 4 documents including "The Document of Japanese Economic History" taking into consideration aspects such as document preservation and optimal content delivery. Further, we discuss various aspects relating to the development a "Virtual library" which would provide an interactive environment for exchanging opinions and comments about these documents. A walk through and "virtual bookshelf" feature are newly implemented for the document retrieval. Several user-interfaces related to the "virtual library" are presented and an information retrieval based on Z39.50 standard using OpenText5.1 is also discussed.

[†]Media Center, Osaka City University
大阪市立大学　学術情報総合センター

[‡]Graduate School of Literature, Osaka City University
大阪市立大学大学院　文学研究科

1.1 "Digital Archives of Japanese Historical Documents"

"Digital Archives of Japanese Historical Documents"(DAJHD) is composed of the following databases.

Table 1. Content of DAJHD

[1] "The Document of Japanese Economic History" image database [Jecoh] , 627 booklets and about 20,000 images
[2] "The Fushimiya Zenbei Document" image database [Fushimi] 1259 documents and 1,986 images
[3] "The Document of Osaka Official Notices" in the Edo Period [Ofure], 10,708 notices
[4] "The Microfilm Archives of the Mori Collection" image database [Mori], 2,259 books and 297,177 blips.

First of all, the digitalization in "Jecoh" is described. The process from the digitalization of the 0th material (generation of the 1st material) to be provided as the archives can be thought by 2 stages below. First, the digital images are made and obtained through the scanning of films, it means the digitization process, after taking picture to the color microfilm for all of documents. It is, of course, based on the coexisting ideas of <archives> and <distribution> in the information resources. It must be considered that how to storage/distribute huge documents and materials under the circumstances such as the digital record media and the input/output devices are short-lived. Next, a building of the image or full text database, and the providing scheme have to be considered.

The characteristics of virtual library DAJHD is as follows;

(1) All the catalog and the full text databases have described based on the XML standard, and the retrieval system is provided by OpenText5.1.

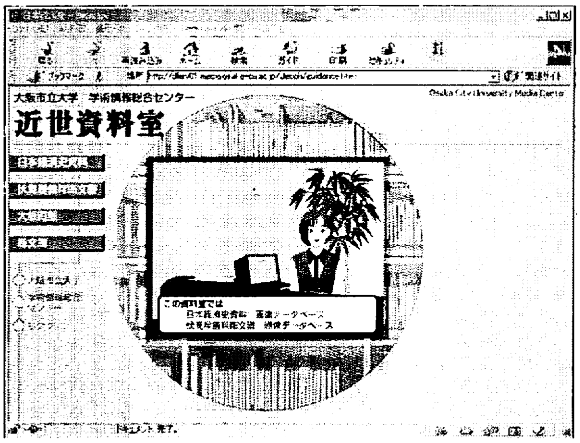

大阪市立大学 学術情報総合センター
Osaka City University Media Center
近世資料室

Fig.1 Reference Counter

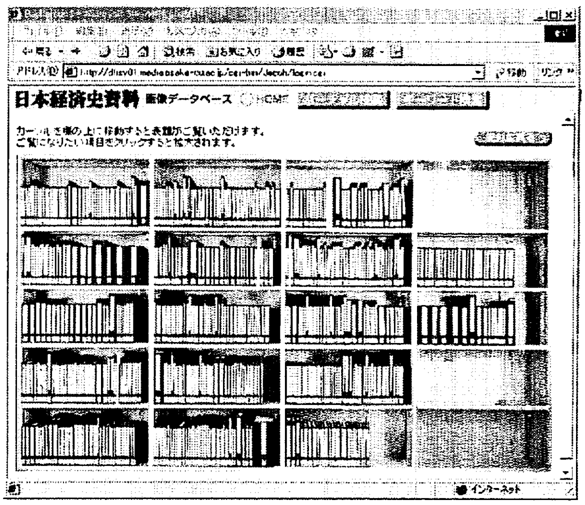


Fig.2 Virtual Bookshelf(1)

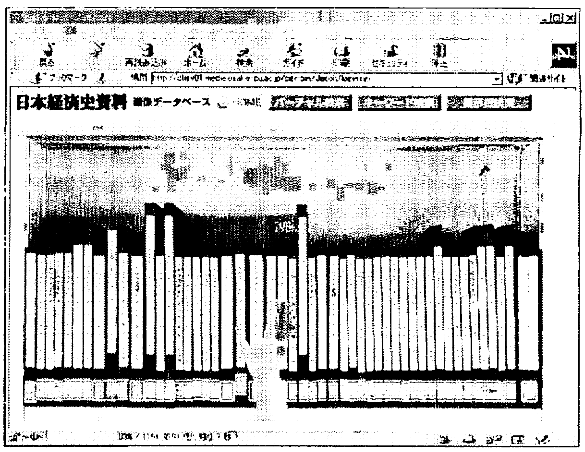


Fig.3 Virtual Bookshelf(2)

(2) Guide by walk through and the document retrieval through the virtual bookshelf. The "Reference counter" in DAJHD is shown in Figure 1. Figure 2 and 3 are taking out of material from

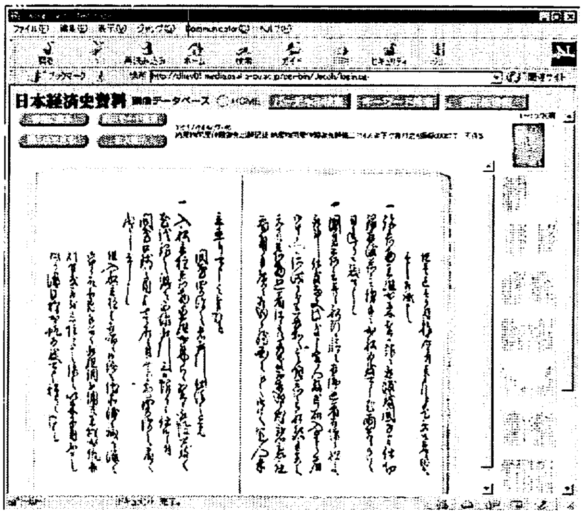
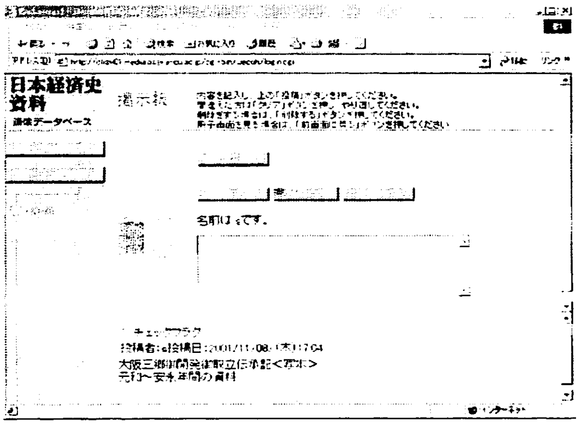


Fig.4 Retrieval Result in [JECOH]

"Virtual bookshelf". The content of the catalog is displayed on "Virtual bookshelf" at a mouse position. The user can be found out the content of material inside in a box at once based on the catalog display function. Figure 4 shows an example of retrieval result.


日本経済史資料

画像データベース

掲示板

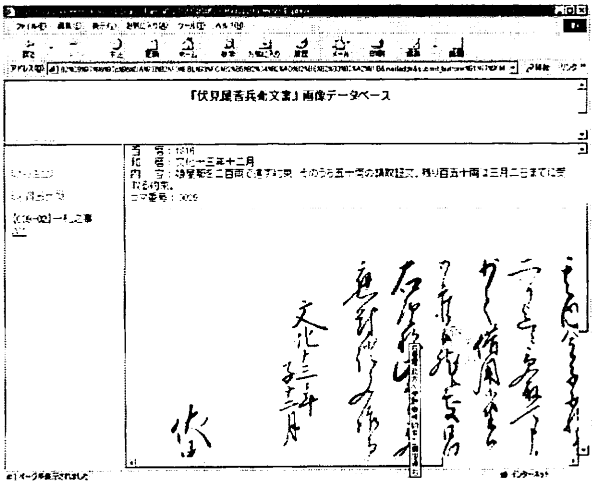
名前は　です。

大阪三郷御闕発御取立伝承記<写本>

元和～安永年間の資料

Fig.5 "Research Memo" Function

(3) A bulletin board function "Research Memo", as shown in Fig.5, which allows users to exchange the memorandum in research activities has equipped to be presented and be opened to the public users (Jecoh).

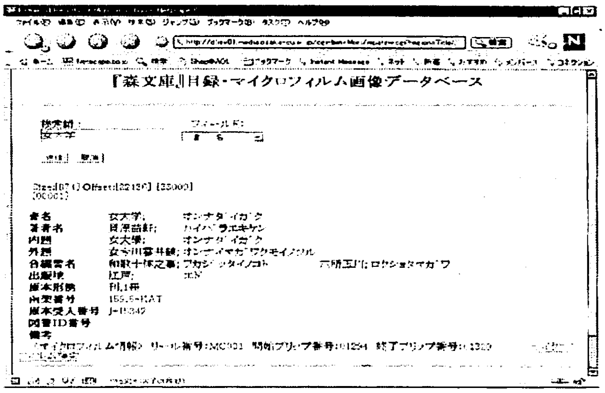

『伏見屋善兵衛文書』画像データベース

年　号：文化十三年十二月

Fig.6 Automatic Transcription in [Fushimi]

(4) A display of automatic transcription at a point of mouse cursor for calligraphic writing on the document. In [Fushimi], the transcription in a sentence where the mouse is currently pointing can be displayed as shown in Figure 6.

(5) New attempts in the user interface, for instance, page reference functions based on the thumbnail display which makes the original position as the upper right corner in the historical document(See Fig.4).


『森文庫』目録・マイクロフィルム画像データベース

書名　女大学;　オンナダイガク

著者名　貝原益軒;　カイバラエキケン

内題　女大学;　オンナダイガク

出版地　江戸;

版本形態　刊,1冊

内架番号　159.5-KAT

図書ID番号

備考

Fig.7 "Mori Collection" Microfilm Retrieval

Besides, (1)the information retrieval standard Z39.50 has introduced and the simultaneous retrieval in various databases has implemented, (2)the mutual referencing function in the several Official Notices [Ofure] using Link function of XML, and (3) the material of [Mori] (a part related with the early modern age) is recorded in about 300,000 blips of microfilm. The catalog can be retrieved through a page of Web browser(See Fig.7), then a blip in retrieval result is dynamically taken out of the stocker (mechanical storehouse) on the fly, then scanning for it done at once(See Fig.8). (4)All of catalog in the database has written by XML standard. The retrieval system has implemented by the engine PAT50 in "OpenText5.1".

Mutual reference and links between the records in the [Ofure] may be added dynamically after extracting the specific keyword related with the reference request from the result of retrieval including XML link description.

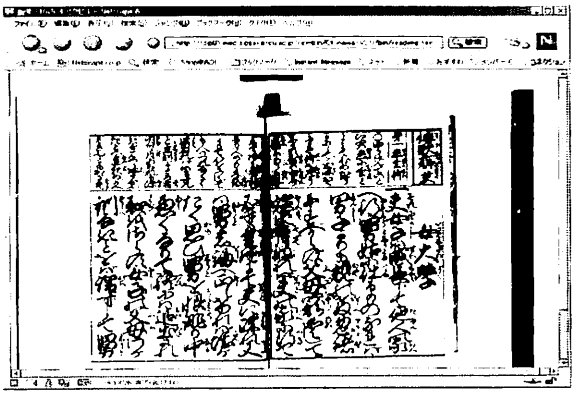


Fig.8 Result of Microfilm Retrieval

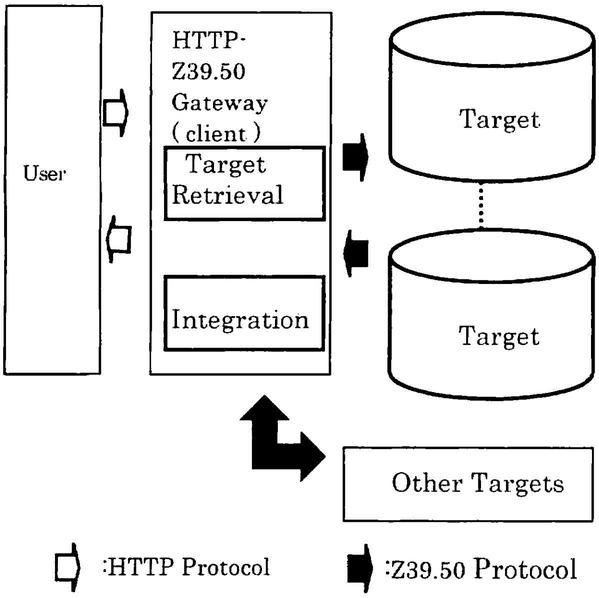

User
HTTP-Z39.50 Gateway (client)
Target Retrieval
Integration
Target
Target
Other Targets
:HTTP Protocol
:Z39.50 Protocol

Fig.9 HTTP-Z39.50 Gateway

1.2 Integrated retrieval based on the Z39.50 standard

The information retrieval Z39.50 standard which has been used in the bibliography information retrieval was newly applied into the historical documents.

The client, server, and retrieval session in Z39.50 are defined as the origin, target , and association respectively. The origin, hereafter, is called as the client in this paper. Moreover, in order to carry out the information retrieval based on Z39.50 protocol from HTTP, the function HTTP-Z39.50 Gateway which converts the protocol from HTTP into the client can be provided. Figure 9 shows this outline. On the image server which employs DAJHD in the digital library, the Z39.50 target, the Z39.50 client,

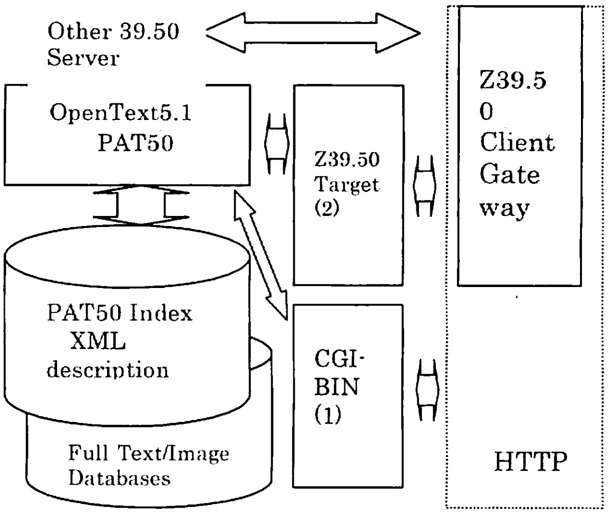

Other 39.50 Server
OpenText5.1 PAT50
Z39.50 Target (2)
Z39.5 0 Client Gate way
PAT50 Index XML description
Full Text/Image Databases
CGI-BIN (1)
HTTP

Fig.10 Z39.50 & OpenText5 Configuration

```
<?xml version="1.0" encoding="Shift-JIS" ?>
<!-- $文書番号$.#0 順序番号.#1 標題.#2 標題の読み.#3 差出人.#4 受取人.#5 西暦.
    #6 和暦.#7 内容.#8 フィルム番号.#9 コマ番号.#10 登録日付] -->
<FushimiyaDocument>
<Document>
<Dno>001-00</Dno><Sno>0001</Sno><Title>預り中銀子之事</Title>
<Reading>アズカリモウスギンスノコト</Reading>
<Sender>嵐三八, 妻　まつ</Sender><Receiver>伏見屋善兵衛</Receiver>
<Year>1796</Year><JapaneseYear>寛政八年五月</JapaneseYear>
<Comment>銀 五百目の借用証文（金一両六十一匁替）。利息（額は不記）をつけ返済。</Comment>
<FilmNo>000001</FilmNo><Blip>0003</Blip><Registration>950515</Registration>
<ImageFile>f0001.jpg</ImageFile><Link>/Fushimi/html/phtml/pf0001.html</Link>
</Document>
```

Fig.11 XML Description in [Fushimi]

```
# Attribute bib-1 Type:Use Value:1
<Use>                          Item name in bib-1
<1016>Document</1016>          #Any
<12>Dno</12>                   #Local number
<53>Sno</53>                   #Number local call
<4>Title</4>                   #Title
<35>Reading</35>               #Title parallel
<1003>Sender</1003>            #Author
<1002>Receiver</1002>          #Name
<31>Year</31>                  #Date of Publication
<30>JapaneseYear</30>          #Date
<63>Comment</63>               #Note
<1011>Registration</1011>      #Date/Time added to db
</Use>
```

Fig.12 Fushimi's Tag Mapping

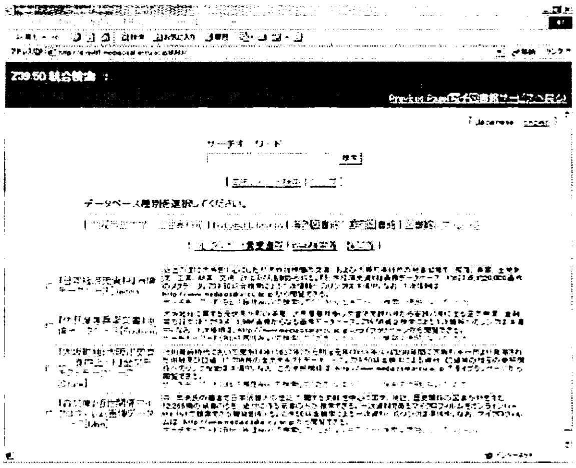

Z39.50 統合検索 :
Japanese
サーチキーワード
データベース種別を選択してください。

Fig. 13 Retrieval based on Z39.50

and HTTP-Z39.50 gateway are worked simultaneously. Figure 10 shows a system configuration where Z39.50 has been implemented by GlobalFinder. Next, it is necessary to consider the attribute set Attribute Sets in Z39.50 standard, its abbreviation is ATR, as the retrieval function. The GlobalFinder treats only bib-1 set, and it is based on the bibliography information. Therefore, in the database shown as in Table 1, it is not easy to map each items into the bib-1 attributes such as Fushimi and Ofure. Figure 12 in the tags as shown in Fig.11 shows a case of mapping between tags in [Fushimi] and bib-1 attribute. Figure 13 shows the retrieval page for specifying a keyword. The retrieval result of catalog links the first information(original resource) using an anchor tag except the Ofure target.

2. Description for restoring process of "Shoso-in Monjo"

The "Shoso-in Monjo" is a generic name of a set of documents that were handed down in the Shoso-in of Todaiji Temple from the 8th century. The "Shoso-in Monjo" has been remarkably deformed and ruined due to transcriptions on both sides of the paper and also "Rearrangement" in the 19th Century. The "Shoso-in Monjo" exhibits an unique dissimilarity between their logical structure and physical layout. This research attempts to apply the XML/XSLT for restoring "Shoso-in Monjo" and develops techniques that could contribute towards a better understanding and preservation of similar historical documents.

The authors had adopted the XML/XSLT technique for restoring remarkably deformed and ruined document in the "Shoso-in Monjo", and developed a support system which examines the procedures for formulating a rule-based logic. The "Shoso-in Monjo" structure is based on related plural material that can be converted into XML expressions using the structured support tool. This XML document can be displayed on a Web browser using the XSL which restores the original logical form of the document while the physical structure is restored with a MS-Word macro as "Tanzaku."

2.1 Shoso-in document

The Shoso-in document shows all the documents which have been stored in Shoso-in in a broad sense among the Shoso-in treasures which are managed by the Imperial Household Agency Office of the Shoso-in Treasure House. In a narrow sense, this shows a part of 667 volumes and 5 books of Shoso-in ancient documents, and a part of "Mokkan".

The paper materials used for these documents had been reused by the reason why a large amount of wastes of official documents. In general, the document(book) before and after reusing the wastes in this paper are defined as 1st and 2nd sides of sheet respectively. During Tempo years (1833-1836), the "arrangement" which the official parts of document had been extracted and edit new version focused on the 1st side of sheet had been carried out and continuously been compiled in Meiji era.

2-2. Shoso-in document "Restoration"

The necessity of the restoration on the 2nd side of sheet depends on the reason described as follows; in the both arrangements in Tempo and Meiji era, the appropriate connecting and/or

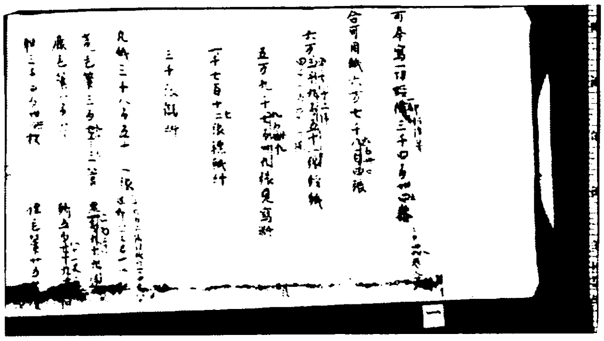
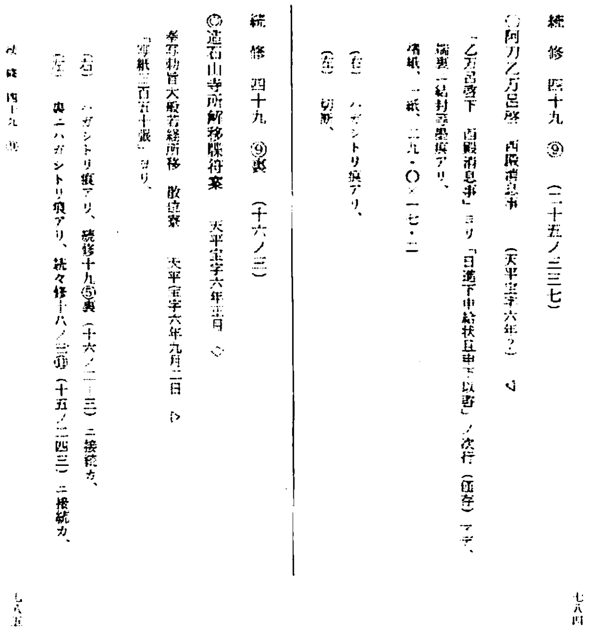


Fig.14 Photogravure Microfilms(D1)



続修　四十九　⑨　（二十五ノ二三七）

○阿刀乙万呂啓　西殿消息事　（天平宝字六年？）　◁

「乙万呂啓下　西殿消息事」ヨリ「日進下中給状且申如件、以啓」ノ次行（㈣存）マデ、

端裏ニ紀封墨痕アリ、

楮紙、一紙、二九・〇×一七・二

（右）　ハガシトリ痕アリ、

（左）　切断、

七八四

続修　四十九　⑨裏　（十六ノ三）

◎造石山寺所解移牒符案　天平宝字六年三月　◇

奉写勅旨大般若経所移　散位寮　天平宝字六年九月二日　▷

「写紙三百五十張」ヨリ、

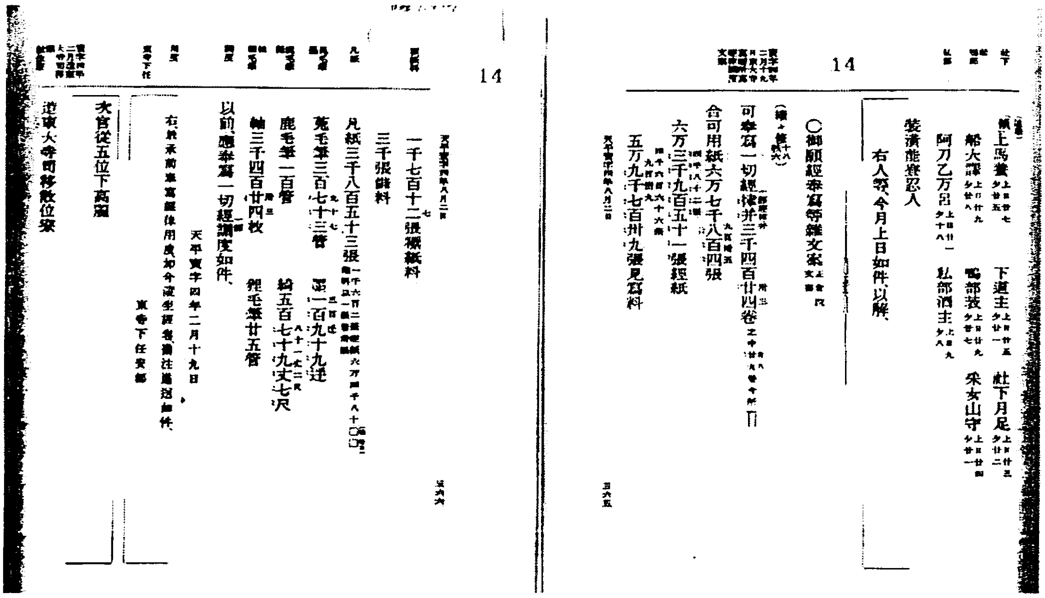
（右）　ハガシトリ痕アリ、続修十九⑤裏（十六ノ二三三）ニ接続カ、

（左）　裏ニハガシトリ痕アリ、続々修十八ノ三⑪（十五ノ二四三）ニ接続カ、

続修　四十九　⑨

七八五

Fig.15 "Shoso-in Document Catalog"(D2)



上馬養　下道主　杜下月足
船大譯　鴨部枝　采女山守
阿刀乙万呂　私部酒主
装潢能登忍人
右人等、今月上日如件、以解、

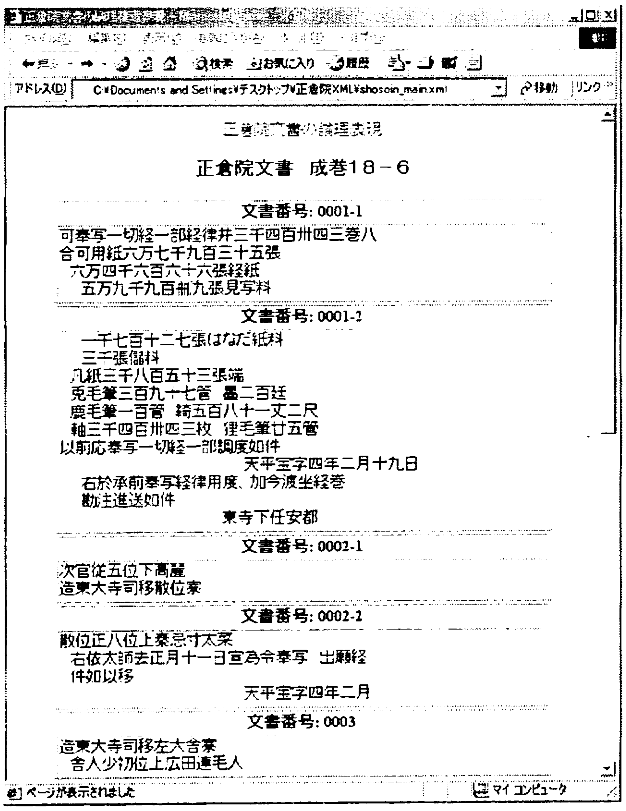
○御願經奉寫等雜文案
可奉寫一切經律并三千四百廿四卷
合可用紙六万七千八百四張
六万三千九百五十一張經紙
五万九千七百卅九張見寫料

一千七百十二張標紙料
三千張餝料
凡紙三千八百五十三張
兎毛筆三百七十三管　墨一百九十九挺
鹿毛筆一百管　綺五百七十九丈七尺
軸三千四百廿四枚　狸毛筆廿五管
以前、應奉寫一切經調度如件、
天平寶字四年二月十九日
右、於承前奉寫經律用度、加今渡坐經卷、勘注進送如件、
東寺下任安都

次官從五位下高麗
造東大寺司移散位寮

Fig.16 "Dainihon Komonjo"(D3)

binding between sheets had been lost because of being collected only excellent documents by making an arbitrary selection. There also are the parts where the relation of the surface and back sides of sheet had confused because it "arrangement" was to be paid attention to the 1st side of sheet, too. Therefore, it lost forms in document should be restored to an original form with focusing on the 2nd side of sheet.



←戻る
検索
お気に入り
履歴
アドレス(D)
C:¥Documents and Settings¥デスクトップ¥正倉院XML¥shosoin_main.xml
移動
リンク

正倉院文書 成巻18－6

文書番号: 0001-1

可奉写一切経一部経律并三千四百卅四三巻八
合可用紙六万七千九百三十五張
六万四千六百六十六張経紙
五万九千九百卌九張見写料

文書番号: 0001-2

一千七百十二七張はなだ紙料
三千張儲料
凡紙三千八百五十三張端
兎毛筆三百九十七管　墨二百廷
鹿毛筆一百管　綺五百八十一丈二尺
軸三千四百卅四三枚　狸毛筆廿五管
以前応奉写一切経一部調度如件
天平宝字四年二月十九日
右於承前奉写経律用度、加今渡坐経巻
勘注進送如件
東寺下任安都

文書番号: 0002-1

次官従五位下高麗
造東大寺司移散位寮

文書番号: 0002-2

散位正八位上秦忌寸太undefined
右依太師去正月十一日宣為令奉写　出願経
件如以移
天平宝字四年二月

文書番号: 0003

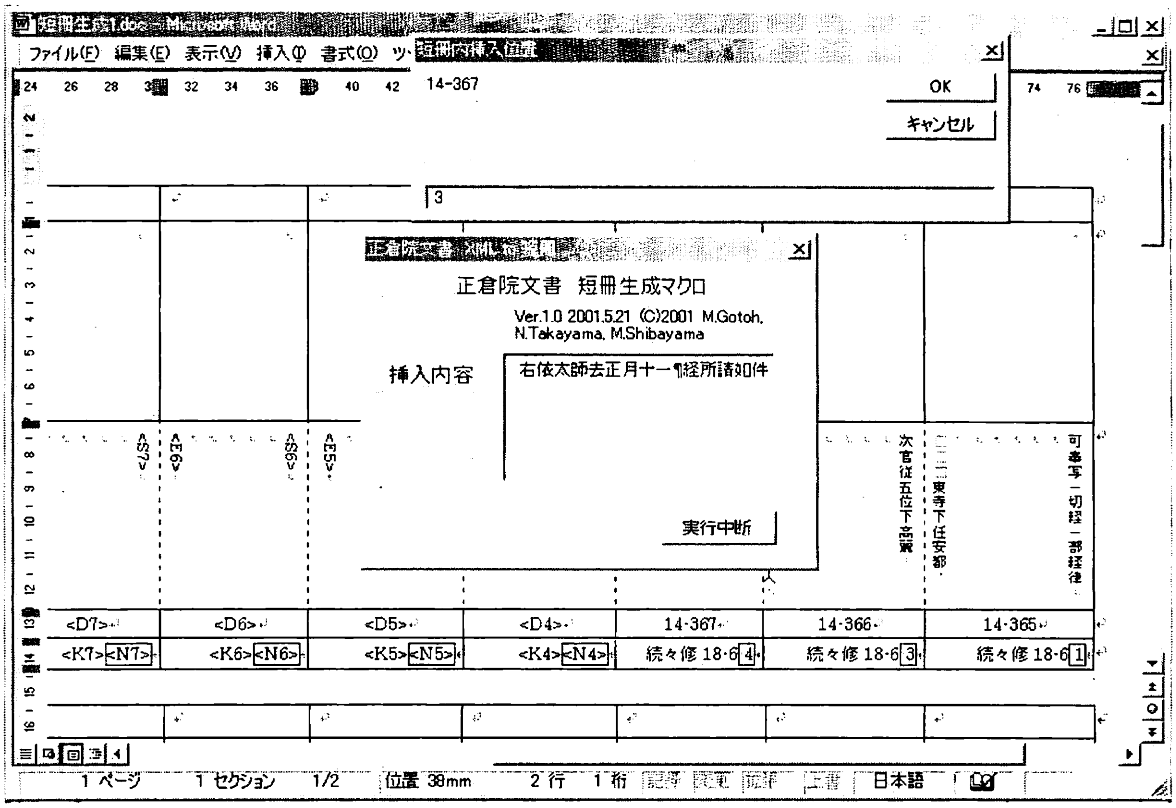
造東大寺司移左大舎寮
舎人少初位上広田連毛人

Fig.17 Logical Expression on Web

The "Tanzaku" can be confirmed by glancing at relevant part for connecting and binding between sheets induced from an restoration idea, the relation of inside and/or outside in both 1st and 2nd sides of sheet, layout in sheet, and a part of the compilation of "Dainihon Komonjyo" (D3). In order to create "Tanzaku", various materials such as photogravure microfilms (D1), "Shoso-in document Catalog" (D2), and D3 are used.

2-3. Examination of restoration in Shoso-in document using XML

Here, the structure as a unit which makes a sheet up is called a physical structure. The information on this structure can be obtained from the D1(see Fig.19) as shown in Figure 14 and D2(see Fig.20) as shown in Figure 15. The structure which is straightened the appropriate form as the document is called a logical structure regardless of the connection of sheets, and the unit of the structure is described in D3 as shown in Figure 16. The both D1 and D2 as the order of arrangement differs from D3 in volumes, page numbers, and the logical structure. Therefore, the way of expression using XML becomes one of problems with the volume and page of D3 against D1 and D2. In case of creating "Tanzaku", the system which employs a feature of being reconfirmed by glancing at the following things has to be equipped: the both first and end lines are shown and an indication which shows the characters position in the sheet paper is needed, the both order and shape which show the order of connection between sheets, and the relation to the position in D3.



短冊生成1.doc - Microsoft Word
ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツ
短冊内挿入位置
OK
キャンセル
14-367
<D7> <D6> <D5> <D4> 14-367 14-366 14-365
<K7><N7> <K6><N6> <K5><N5> <K4><N4> 続々修 18-6 4 続々修 18-6 3 続々修 18-6 1
1 ページ 1 セクション 1/2 位置 38mm 2 行 1 桁 日本語
挿入内容
実行中断

正倉院文書 短冊生成マクロ

Ver.1.0 2001.5.21 (C)2001 M.Gotoh, N.Takayama, M.Shibayama

右依太師去正月十一¶経所請如件

Fig.18 "Tanzaku" Creation Process using MS Word

2-4. Description of restoration process using XML/XSLT

First of all, the Shoso-in document researcher inputs each historical materials on a certain rule, and creates the Shoso-in document data base. D1, D2, and D3 which configure this database are each RDB, XML/RDB, and the XML forms respectively. Structural conversion from the RDB form to XML is an already-known technology. After converting it, a transformation which creates an integrated new "Structure" for whole data from each XML form in each data has to be needed. A rule used for this transformation is called a structure rule. This rule also implies a knowledge base in restoration process of the Shoso-in document as below.

The structure rule is described with XSLT. Figure 19 shows XML document which is the intermediate expression generated based on this structure rule. In this XML document, the description following a tag <正倉院文書> (Shoso-in document) constitutes basically the connected sheets as a unit of sheet with the physical structure (it is shown as a part enclosed in the short dashed line in Figure 19). However, the expression of the logical structure is described behind the unit of the physical structure based on the logical ID (It is shown in a part enclosed with the solid line of Figure 21). The expression of the logical structure assumed to be display by Web a browser, and the physical structure is restored on the "Tanzaku".

2-4-1. Expression of logical structure

In order to display the document with the logical structure, a style sheet using XSL has to be made firstly, and then an output function which employs the display the XML document according to the logical structure has implemented. The style sheet displays the Shoso-in document with XML expression onto Web browser after constituting a logical structure from the intermediate document. Figure 17 shows its result. In Gaiji expression, the TrueType font of "Konjyaku Mojikyo" is used.

2-4-2. Tanzaku(Strip of paper)

In "Tanzaku" which expresses the physical structure, "Tanzaku" generation macro has developed by using the macro-function on the word processor MS-Word (Word2000 edition). This macro reads the XML document which is the intermediate expression as shown in Figure 19, and inserts the tags necessary for creating "Tanzaku" such as the sheet, volume and page numbers in "Dainihon Komonjo", the first line, the end line, and connecting information into Word document. The work of the "Tanzaku" making which has been very taken time up now can be easily processed by using the macro (see Figure 18). By developing these functions, the implemented system contributes not only to accurately proceed the restoration work but also to reduce the working time and to efficiently advance the Shoso-in researches.

3. Conclusion

An optimal content delivery in the digital archives for 4 documents using the XML description in Edo era has been discussed. And, Several user-interfaces, including a "virtual bookshelf" feature with walk through, related to the "virtual library" has presented and the simultaneous retrieval for these documents based on Z39.50 standard also has described.

Also, an application of the XML/XSLT for describing the process of restoration in "Shoso-in Monjo" has introduced.

References

[1]Mamoru Shibayama, Yoshikuni Yoshii, Kazutoshi Fujikawa, and Venkatesh Raghavan: Historical Document Databases and Integrated Retrieval System, Bulletin of Media Center, Osaka City University, Vol.3, pp.41-49, 2001

[2] Makoto Goto, Norihisa Takayama, and Mamoru Shibayama: "An examination of structurizing Shoso-in document and restoration using XML," SIG Report of IPSJ, 2001- CH-51, Vol.2001, No.67, pp.31-38, 2001

[3] Khun Yee Fung: "XSLT: working with XML and HTML", Addison-Wesley, 2001

[4] Towao Sakaehara: "Shakyo and Dairi in the Nara era (Nara-jidai no Shakyo to Dairi)", 2000

[5] Yumi Yamashita: "Development of Shoso-in Monjo research and approach to historical materials, (Shoso-in Monjo Kenkyu no Tenkai to Shiryo heno Apurohchi)", 'Rekishi Kagaku', No. 156, 1999

[6] Shuichi Yasuzawa: "ENCODED ARCHIVAL DESCRIPTION EAD: Adaptation on SGML-XML", SIG Report of IPSJ, 2001-CH-51, Vol. 2001, No. 67, pp.17-24, 2001.

[7] Toshiharu Gotoh: "The possibility of historical materials description using XML -- On the second edition of ISAD(G) (International standard: a general principle of record document description) and the database", materials from a meeting of the Humanities and Social Sciences Section of the Japan Society of Information and Knowledge, and 16th "Workshop on Computer Application in Historical Studies ", 2001.

[8] Makoto Goto, Norihisa Takayama, and Mamoru Shibayama: "Structuring and Restoration of "Syoso-in Monjyo" using the XML", SIG Report of IPSJ, 2001-CH-51, Vol. 2001, No. 67, pp.31-38, 2001

[9] Sugimoto Kazuki: "Investigation of the original Shoso-in Monjo", 'Kodai Monjo-ron', 1999

[10] Eiichi Ishigami: "Result of compilation in Shoso-in Monjo and viewpoint of reexamination in ancient document theory", 'Kodai Monjo-ron', 1999

[11] Sachihiro Yamada: "XML sample collection", 'Syuhwa systems', 2001

[12] Microsoft Web Workshop, Reference 2001.08.20

[13] Mojikyo Kenkyukai: "Pasokon Yu Yu Kanji Jyutsu", 'Kinokuniya Shoten', 2001

[14] Takashi Yoshida: "Ritsuryo Jidai no Koeki", 'Ritsuryo Kokka to Kodai no Shakai', 1983

[15] Towao Sakaehara: "Shoso-in Monjo Bibliographic Catalog", 'Shoso-in Monjo Kenkyu', No. 1, 2, and 3, 1999

```
<?xml version="1.0" encoding="Shift_JIS" ?>
<?xml-stylesheet                        type="text/xsl"
href="makexml.xsl" ?>
<output>
<record><ID>1</ID>
<識別番号></識別番号>
<所属>続々修 18-6</所属>
<紙番号>1</紙番号>
<大日本古文書巻>14</大日本古文書巻>
<大日本古文書頁>365</大日本古文書頁>
<先頭行>可奉写一切経</先頭行>
<最終行>東寺下任安都</最終行>
<備考></備考>
</record><record><ID>2</ID>
<識別番号></識別番号>
<所属>続々修 18-6</所属>
<紙番号>2</紙番号>
<大日本古文書巻>14</大日本古文書巻>
<大日本古文書頁>366</大日本古文書頁>
<先頭行>次官従五位下高麗</先頭行>
<最終行>舎人少初位上広田連毛人</最終行>
<備考></備考>
</record>
      以下、省略
</record>
</output>
```

Fig.19 XML expression of photogravure data: D1

```
<?xml version="1.0" encoding="Shift_JIS" ?>
<catalog>
<data><ID>1</ID>
<識別番号></識別番号>
<所属>続々修 18-6</所属>
<大日本古文書巻>14</大日本古文書巻>
<大日本古文書開始頁>365</大日本古文書開始頁>
<大日本古文書終了頁>366</大日本古文書終了頁>
<史料名></史料名>
<年代></年代>
<料紙情報左><切断 /><白紙 /></料紙情報左>
<料紙情報右></料紙情報右>
<料紙情報他></料紙情報他>
<接続情報左></接続情報左>
<接続情報右></接続情報右>
</data>
      以下、省略
</catalog>
```

Fig.20 XML expression of
Shoso-in Monjo Catalog: D2

```
<?xml version="1.0" encoding="Shift_JIS"?>
<正倉院文書>
<成巻><所属>続々修 18-6</所属>
<成巻本文>
<成巻題目>御願奉写等雑文案</成巻題目>
<紙>
<大日本古文書本文><大日本古文書巻頁>14-365</大日本古文書巻頁>
<論理文書 id="1">
<line_s>可奉写一切経<del>律并</del><ins>一部経律并</ins>三千四百卌四<ins>三</ins><a>巻
</a><wari>之中卌九</wari><ins>八</ins><wari>巻今所</wari></line_s><br />
<line>合可用紙六万七千<del>八百四</del><ins>九百三十五</ins>張</line><br />
<line>  六万<del>三千九百五十一張</del><ins><del>四千八十二張</del><ins>四千六百六十六張
</ins></ins>経紙</line>
<line>    五万九千<del>七百卌九</del><ins>九百<gt set="mojikyo" name="002712" />九</ins>
張見写料</line>
</論理文書>
</大日本古文書本文>
<大日本古文書本文><大日本古文書巻頁>14-366</大日本古文書巻頁>
<論理文書 id="1">
<line>    一千七百十二<ins>七</ins>張<f>はなだ</f>紙料</line>
<line>    三千張儲料</line>
<line>  凡紙三千八百五十三張<wari>一千六百二張経師六万四千八十張</wari><ka>端</ka><wari>
継料以一張着</wari><gt set="mojikyo" name="002712" /><wari>張</wari></line>
<line>  兎毛筆三百<del>七十三</del><ins>九十七</ins>管  墨<del>一百九十九廷</del><ins>二
百廷</ins></line>
<line>  鹿毛筆一百管  綺五百<del>七十九丈七尺</del><ins>八十一丈二尺</ins></line>
<line>  軸三千四百卌四<ins>三</ins>枚  狸毛筆廿五管</line>
<line>以前応奉写一切経<ins>一部</ins>調度如件</line>
<line>                                天平宝字四年二月十九日</line>
<line>    右於承前奉写経律用度、加今渡坐経巻</line>
<line>    勘注進送如件</line>
<line_e>                              東寺下任安都</line_e>
</論理文書>
</大日本古文書本文></紙>
<切断  /><白紙 />
<紙>
<大日本古文書本文><大日本古文書巻頁>14-366</大日本古文書巻頁>
<論理文書 id="2">
<line_s>次官従五位下高麗</line_s>
<line>造東大寺司移散位寮</line>
</論理文書>
</大日本古文書本文>
<大日本古文書本文><大日本古文書巻頁>14-367</大日本古文書巻頁>
<論理文書 id="2">
<line>散位正八位上秦忌寸太菜</line>
<line>  右依太師去正月十一日宣為令奉写  出願経</line>
<line>  件如以移</line>
<line>                                天平宝字四年二月</line>
</論理文書>
<論理文書 id="3">
<line>造東大寺司移左大舎寮</line>
<line_e>  舎人少初位上広田連毛人</line_e>
</論理文書>
</大日本古文書本文></紙>
<紙>
<大日本古文書本文><大日本古文書巻頁>14-367</大日本古文書巻頁>
<論理文書 id="3">
<line_s>    右依太師去正月十一日宣為令奉写</line_s>
<line_e>    経所請如件</line_e>
</論理文書>
</大日本古文書本文></紙>
<紙>
                        以下、略
</成巻>
</正倉院文書>
```

Fig.21 Intermediate Expression Based on XML Description