人文科学とコンピュータ　36－4
（１９９７．１１．１５）

インターネット上の古文書画像データベースシステムの作成

岡部 建次　永田 大　広瀬 順晧
駿河台大学 文化情報学部

概要

インターネット上で古文書画像データベースを公開する。原文書コピーの画像データベースを、インターネット上で公開すれば利用は促進される。明治政治史料の個人文書では、多くのデータは書簡の形をしている。書簡は数ページから成り立っているため、１書簡を１Web とし、書簡を構成する画像を貼り付ければよい。こうして多量の１書簡１Web の Web ファイル（HTML ファイル）から成り立つデータベースをインターネット上に作った。本稿は、インターネット上で利用できる古文書画像データベースとデータベース作成のノウハウの報告である。

Archives image database system on the Internet

Kenji Okabe　Dai Nagata　Yoshihiro Hirose
The Faculty of Cultural Information, Surugadai University

Abstract

We open an Archives image database system to the public through the Internet. The open through the Internet promotes the access by the users. Most private documents of Meiji political archives take the form of letters. The letter consists of several pages. We assigned one letter to one Web, and each page of the letter is pasted to each web as an image. Then the image database which consists of many Web pages is prepared on the Internet server computer. We report the way of producing Web database.

1．はじめに

原文書コピーの画像データベースを、インターネットで誰もがどこからでも手軽にアクセスできるようにすれば利用は促進される。インターネットホームページでは、Web に写真が貼り付けてあるものが多い。そこで我々は写真の代わりに古文書画像データを貼り付ければ、インターネット上で古文書画像データベースを公開できると考えた。古文書のうち、我々が対象としている明治政治史料の個人文書では、多くのデータは書簡の形をしている。書簡は数ページから成り立っている。従って１書簡を１Web とし、書簡を構成する各ページ画像を貼り付ければよい。こうして多量の１書簡１Web の Web ファイル（HTML ファイル）から成り立つデータベースをサーバー上に作った。データベースとデータベース作成のノウハウを報告する。

2． 問題点

2．１ ＣＤ－ＲＯＭ版画像データベースの場合

通常画像データベースはＣＤ－ＲＯＭで提供される。ＣＤ－ＲＯＭ版画像データベースは高画質な画像データを提供できる反面、更新ができない。インターネット経由の場合、最新の画像データを容易に提供できる。

2．２ Ｗｅｂデータベース作成上の問題点

・画像貼付の方法

Web データベースでは、１書簡に相当する HTML ファイルに書簡の各ページ画像をリンクする。これは<IMG>タグを用いて、図１のような HTML 形式のファイルを作成する。図２にこの HTML ファイルをブラウザで表示した実際の画像を示す。

・ 大量の原文書画像

数枚(数ページ)から成る１書簡は図３のようにブラウザに表示される。これは図４のような HTML ファイルを作ることで可能になる。個人文書データベースにおいては、画像ファイルは膨大な数になる。例えば、谷干城関係文書画像は約４０００枚、書簡件数で約９００に及ぶ。このような大量の画像ファイルをリンクした書簡 Web（谷干城関係文書の例では９００）を手作業で１件ずつ作成するには、大変な時間と労力のかかる。

3．解決法

大量の画像ファイル(各ページ)を書簡ごとにリンクする作業は、プログラムによって自動的に行う方法を考案した。

4．プログラムによる HTML 化の方法

4．１ 原文書画像のファイル名と文書(書簡)の関係

原文書画像のファイル名は、文書(書簡)番号（５桁）＋（文書(書簡)内での）枝番号（３桁）の計８桁である。この８桁のファイル名に拡張子が３桁付与された形式になっている。このファイル名から、HTML 形式によるリンクファイル(テキストファイル)をプログラムで作成する。

表１のように、原文書画像が存在するものとする。先頭の k は黒田清隆関連の文書であることを示す。次の４桁は、一つの書簡を表す数字である。一つの書簡は、複数の原文書画像から構成される。次の３桁は、一つの書簡内の枝番号となる。つまり、表１の例では、k0001001.gif から k0001007.gif までが一つの書簡であり、k0002001.gif から k0002003.gif までが、次の書簡を意味する。書簡によって、

構成する原文書画像の数はさまざまである。１つの画像のみ書簡もあれば、100 を超える画像から構成されるものもある。

表１ 原文書画像のファイル名（サンプルであり実際のものとは、異なる）

k0001001.gif	k0001002.gif	k0001003.gif	k0001004.gif	k0001005.gif
k0001006.gif	k0001007.gif	k0002001.gif	k0002002.gif	k0002003.gif

４．２ 「１書簡－１Ｗｅｂページ」によるリンク

１つの HTML ファイルから１つの書簡分の画像をリンクするようにする。つまり、「１書簡－１Web ページ」である。「１書簡－１Web ページ」にすることによって、文書目録データベースで検索した書簡をすばやく見ることができる。

「１書簡－１Web ページ」であるから、同一の書簡の原文書画像が一つの HTML ファイルとしてプログラムで出力される。出力される HTML ファイルのファイル名は、書簡の番号（原文書画像の先頭５桁）を付加した。表２・表３のように、HTML ファイルから１つの書簡分の画像がリンクされる。

表２ k0001.htm からリンクされる原文書画像（書簡 0001 の画像）

k0001001.gif	k0001002.gif	k0001003.gif	k0001004.gif	k0001005.gif
k0001006.gif	k0001007.gif			

表３ k0002.htm からリンクされる原文書画像（書簡 0002 の画像）

k0002001.gif	k0002002.gif	k0002003.gif

すなわちプログラムで文書１件ごとに画像ファイル名を読んで、解析して、図４のテキストファイル（HTML ファイル）を作成していく。図３に実際の画像を示す。

４．３ 「一定の画像数－１Ｗｅｂページ」にする必要性

「１書簡－１Web ページ」によるリンクを行うと、書簡によっては、100 を超える画像から構成されるものもある。このような書簡をすべて一つの HTML ファイルにしてしまうと、画像ファイルを読み込むだけで非常に時間がかかる。100 以上の画像が並んでしまっては、かえって手紙が読みにくい。そこで、一定の画像数を超えた場合、１書簡の中でも別のファイルを分けることにした。例えば、表１の原文書画像の書簡 0001 において、５つ以上の原文書画像から構成される書簡を複数の Web ページに分けた場合、表４・表５のようになる。プログラムを利用するので、何画像で別の HTML ファイルに分けるかの設定は、容易に変更できる。

表４ k0001.htm からリンクされる原文書画像（書簡 0001 の前半の画像）

k0001001.gif	k0001002.gif	k0001003.gif	k0001004.gif	k0001005.gif

表５ k0001_2.htm からリンクされる原文書画像（書簡 0001 の後半の画像）

k0001006.gif	k0001007.gif

出力される HTML ファイルは図５・図６のような形式である。

5．おわりに

1 Web が 1 文書になっているためインターネットで容易に利用できる。各書簡内の各ページ(1 画像)は 1 Web に収まる大きさで、画面で読める解像度、プリントして読める鮮明さを維持している。

既に谷干城文書の CD-ROM の公開配布をおこなっており、インターネット経由による web データベースの作成中である。本研究は私学振興財団学術研究奨励金、駿河台大学学内共同研究補助金、科学研究費（データベース促進）の助成を受けて行った。

図1 HTML による画像のリンク

```
<HTML>
<HEAD>
<TITLE>k0001</TITLE>
</HEAD>
<BODY>
k0001001.gif<BR><IMG SRC=k0001001.gif><P>
</BODY>
</HTML>
```

図2 HTML によるリンクの画像

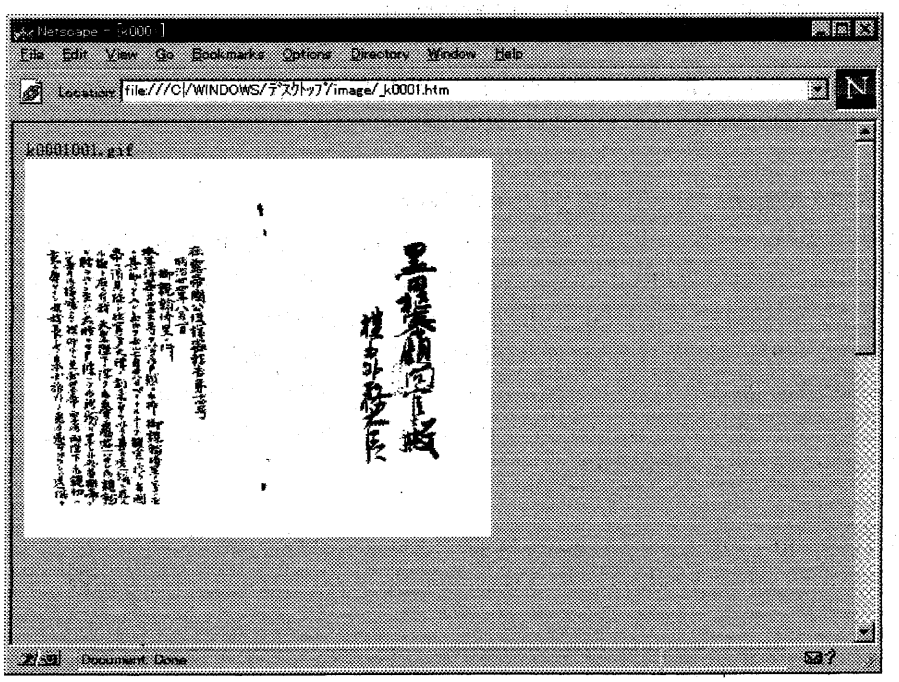

図3　「1書簡－1Webページ」によるリンクの画像

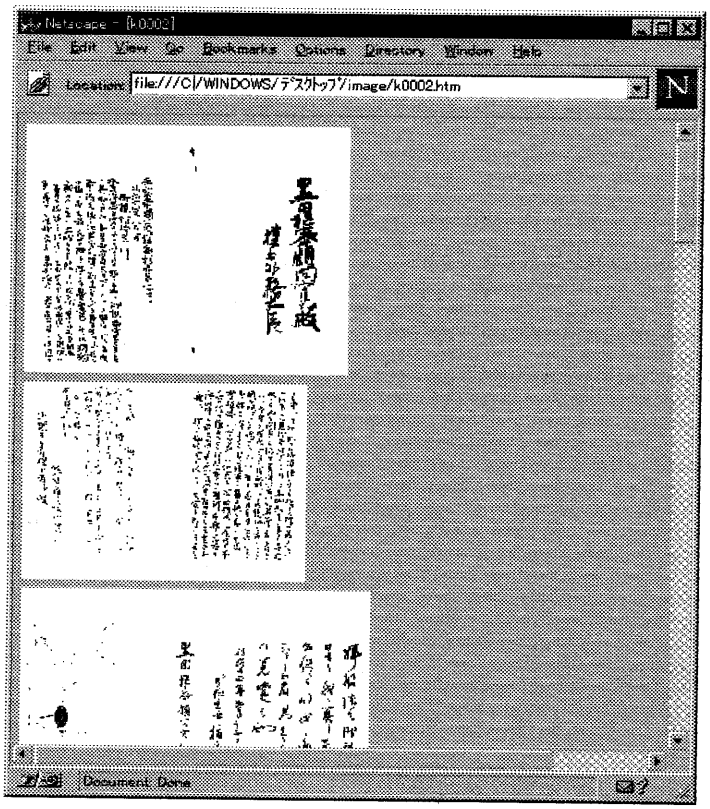

図4　「1書簡－1Webページ」によるリンクのHTMLファイル（ファイル名 k0001.htm）

```
<HTML>
<HEAD>
<TITLE>k0001</TITLE>
</HEAD>
<BODY>
k0001001.gif<BR><IMG SRC=k0001001.gif><P>
k0001002.gif<BR><IMG SRC=k0001002.gif><P>
k0001003.gif<BR><IMG SRC=k0001003.gif><P>
k0001004.gif<BR><IMG SRC=k0001004.gif><P>
k0001005.gif<BR><IMG SRC=k0001005.gif><P>
k0001006.gif<BR><IMG SRC=k0001006.gif><P>
k0001007.gif<BR><IMG SRC=k0001007.gif><P>
</BODY>
</HTML>
```

図５　「一定の画像数－１Webページ」によるリンクのHTMLファイル（ファイル名 k0001.htm）

```
<HTML>
<HEAD>
<TITLE>k0001</TITLE>
</HEAD>
<BODY>
k0001001.gif<BR><IMG SRC=k0001001.gif><P>
k0001002.gif<BR><IMG SRC=k0001002.gif><P>
k0001003.gif<BR><IMG SRC=k0001003.gif><P>
k0001004.gif<BR><IMG SRC=k0001004.gif><P>
k0001005.gif<BR><IMG SRC=k0001005.gif><P>
</BODY>
</HTML>
```

図６　「一定の画像数－１Webページ」によるリンクのHTMLファイル（ファイル名 k0001_2.htm）

```
<HTML>
<HEAD>
<TITLE>k0001</TITLE>
</HEAD>
<BODY>
k0001006.gif<BR><IMG SRC=k0001006.gif><P>
k0001007.gif<BR><IMG SRC=k0001007.gif><P>
</BODY>
</HTML>
```