

データベース構築論2

19章「画像ファイルのデータベース化」

中島康彦

§19. 1 印刷物の小型化

マイクロフィルムリーダー

▶ 光学的に拡大表示/印刷



§19. 2 印刷物の電子化

マイクロキャナシステム

- ▶ マイクロフィルム画像を画像ファイルへ
- 電子ファイリングシステム(EFS)
- ▶ 印刷物を光磁気ディスク上の画像ファイルへ



§19. 2 印刷物の電子化(続き)

PC+スキャナ

- ▶ 印刷物を汎用記録メディアへ



§19. 3 印刷物のテキスト化

画像ファイルをテキストファイルへ変換

- ▶ OCR(Optical Character Reader)によりテキストファイルへ変換

キーワード検索

- ▶ 予め入力してある正しいキーワード(文書名)を用いる

全文検索

- ▶ 認識誤りを含むテキストファイルを用いる

§19. 4 利用可能なデータ

画像ファイル名と文書名, 日付の対応

- ▶ EUC

index

00000000.tif&通番&文書名&日付

00000001.tif&通番&文書名&日付

00000002.tif&通番&文書名&日付

画像ファイル

- ▶ マルチページTIFF

00000000.tif

00000001.tif

00000002.tif

画像から抽出した, 認識誤りを含むテキストファイル

- ▶ EUC

00000000

00000001

00000002

§19. 5 WEBにまとめる(test4.archive/)

前章のシステム(test3.archive/)に対して

mailshow.cgi	...	不要
howto.html	...	同じ
index.cgi	...	同じ
index.shtml	...	同じ
jcode.pl	...	同じ
search.cgi	...	同じ
search.shtml	...	同じ
mailindex	...	変更
mailsearch	...	変更

§19. 6 mailindex

```
#!/usr/local/bin/perl
$Filename = $ARGV[0];

if ($Filename !~ /^\.d+$/) {
    exit 0;
}

open(FILE0, "index");
while (<FILE0>) {
    chop;
    if (/^$Filename\.tif&(.*?)&(.*?)&(.*?)&.*$/i) {
        $Date    = $3;
        $Subject = $2;
        last;
    }
}
close FILE0;

$Date    = "Unknown" unless ($Date);
$Subject = "None"     unless ($Subject);
printf "<a href=\"%s.tif\" target=frame3>%-12.12s %-40.40s </a>", $Filename,
$Date, $Subject;
printf "\n";

exit 0;
```

§19. 7 mailsearch

```
#!/usr/local/bin/perl

require 'jcode.pl';

sub init_form {
:中略
}

$QUERY_STRING = $ARGV[0];
$Filename = $ARGV[1];
shift;

if ($Filename !~ /^\\d+$/) {
    exit 0;
}
```

§19. 7 mailsearch(続き)

```
&init_form('euc');
$string = $form{'string'};
$found = 0;

open(FILE0, "index");
while (<FILE0>) {
    chop;
    if (/^$Filename\\.tif&(.*?)&(.*?)&(.*?)&.*$/i) {
        $Subject = $2;
        $Date = $3;
        if ($Subject =~ /$string/i) {
            $found = 1;
        }
        last;
    }
}
close FILE0;

$Date = "Unknown" unless ($Date);
$Subject = "None" unless ($Subject);
```

§19. 7 mailsearch(続き)

```
open(FILE0, "$Filename");
while (<FILE0>) {
    chop;
    if (/.$String/i) {
        $found = 1;
        last;
    }
}
close FILE0;

if ($found == 1) {
    printf "<a href=\"%s.tif\" target=frame3>%-12.12s %-40.40s </a>",
$Filename, $Date, $Subject;
    printf "\n";
}

exit 0;
```

§19. 8 実行結果

The screenshot shows a web browser window with the following elements:

- Address Bar:** `http://localhost/test4.archive`
- Search Bar:** "本文検索" with "go" and "clear" buttons.
- Search Results:**
 - 93/05/17 創刊の辞
 - 93/05/17 貧富問題 (1)
 - 93/05/17 でう みつど・ひゆーむノ経済学説 (1)
 - 93/05/17 地代ノ性質ニ就テ
 - 93/05/17 地方財政ノ調整
 - 93/05/17 減債基金ト鉄道資金
 - 93/05/17 独逸ノ自治制ニ就キテ
 - 93/05/17 戦争ト社会問題
- PDF Viewer:** Displays a document titled "京都市立大学法科大学大 経済論叢 第一卷 第一号". The document content includes a list of articles:
 - 貧富問題 (1)
 - でう みつど・ひゆーむノ経済学説 (1)
 - 地代ノ性質ニ就テ
 - 地方財政ノ調整
 - 減債基金ト鉄道資金
 - 独逸ノ自治制ニ就キテ
 - 戦争ト社会問題
 - 津村博士ノ国
 - 独逸ノ新傾向
 - 獨逸ノ外國政策
 - 獨逸ノ地方ノ現勢
 - 著名ノ婦人ニ關スル
 - 獨逸ノ教育ト米國
 - 獨逸ノ人口増進政策
 - 日本債權者對獨逸ノ
 - 和田博士ノ經濟學
 - Richard Meier 著スル

§19. 9 経済論叢データベース

http://i.econ.kyoto-u.ac.jp/test4.archive/

The screenshot shows a web browser window with a table of contents and a PDF viewer. The table of contents lists various economic topics and their authors, all dated 1915. The PDF viewer shows the cover of the book 'Economic Essays' (経済論叢) published by the University of Kyoto.

Author	Title	Year	Page
	創刊の辞	1915	1.1
田島錦治	貧富問題(1)	1915	1.1
福田徳三	ていつど:ひゆーむノ経済学説(1)	1915	1.1
戸田海市	地代ノ性質ニ就テ	1915	1.1
神戸正雄	地方財政ノ調整	1915	1.1
小川郷太郎	減債基金ト鉄道資金	1915	1.1
財宮隆毅	独逸ノ自治制ニ就キテ	1915	1.1
米田庄太郎	戦争ト社会問題	1915	1.1
河上良彦	津村博士ノ経済学	1915	1.1

The PDF viewer shows the cover of the book 'Economic Essays' (経済論叢) published by the University of Kyoto. The cover features the title in large characters and a list of authors and topics.

今日はここまで