

CA on CD Tagged File の Tomoko Text File への変換

原田 茂治¹

Conversion of CA on CD Tagged File to Tomoko Text File

HARADA, Shigeharu¹

CA on CD の検索結果の一出力形態である tagged-formatted file を、不定長データベースソフトである「知子の情報」¹⁾に連続登録可能なファイル形式(知子テキストファイル, Tomoko text file)に変換する簡単なプログラムを作成した。これによって、ほとんど人手をかけることなく、検索結果をパーソナルデータベース化することができる。

1. 緒言

化学徒にとって最大かつ最良の2次情報誌である Chemical Abstracts の重要性はもはや言うまでもない。その印刷冊子体マニュアル検索の「味わい」は別として、この巨大なデータベースには機械検索が適切であることも指摘するまでもないことである。商用データベースである Dialog の CA Search ファイルの検索結果を、MS-DOS 時代の「知子の情報 Ver. 3」を使って、全文パーソナルデータベース化する方法を既に報告して、²⁾利用してきた。その後、著者は CA on CD を利用するようになったので、その出力の一形式である tagged-formatted file を Tomoko text file に変換する簡易プログラムを、Quick Basic 4.5³⁾(NEC MS-DOS 用)と Visual Basic 6.0⁴⁾で作成した。いずれも Basic 入門者が練習に作成した程度のものであるが、利用価値があると思われるので公開することにした。

2. ファイル変換

CA on CD の tagged format は List 1 に示す構造をもっている。一方、知子テキストファイルには、List 2 に示す構造をもたせなくてはならない。そこで具体的には List 3 に記載したスタイルのファイルを生成させることにした。このために Quick Basic 4.5 を用いて変換プログラムを作成した(List 4. cascd1.bas)。これをコンパイルしたものが cascd1.exe である。MS Windows 上では、変換したい tagged-formatted file をアイコン化した cascd1.exe にドラッグすれば、ファイル基

¹ 静岡県立大学短期大学部一般教育等 〒422-8021 静岡市小鹿 2-2-1
E-mai: haradas@bambi.t.u-shizuoka-ken.ac.jp

部が同名の.txt ファイルを同一パスに生成する。ところで cascd1.exe は Quick Basic 4.5 が動作する機種 ,NEC PC-9801 または PC-9821 の MS-DOS 2 以上か MS Windows 上で動作する。今どきそのような環境は珍しいであろうから ,一般的な MS Windows(開発環境は MS Windows98)上で動作するように Visual Basic 6.0 で作成したプログラムも掲載しておく。この場合には ,拡張子が tag あるいは txt であるファイルを読み込み ,「知子の情報 Ver. 8」に連続登録するデフォルト拡張子 ren をもつファイルを生成する。Visual Basic のソースファイルが source.lzh ,コンパイルした実行型ファイルが ,CAtoTomoko.exe ,ランタイムライブラリおよびセットアッププログラムを同梱したものが fullset.lzh である。これらはファイル名をクリックすることによってダウンロードできる。なお ,著作権の関係上 ,CAonCD の検索結果ファイルは添付されていない。稚拙なプログラムではあるが ,著者と同じ目的をもっている研究者の手間を省くことができれば幸である。ソースの手直し等の指摘を受けことができれば ,感謝に堪えない。

3 . 知子の情報への連続登録

著者が使用中の「知子の情報 Ver. 8」(最新版は Ver. 9)では ,「ツール」 「知子テキストの読み込み」 「一気に登録」で ,検索結果のすべてがタイトルおよびキーワード付きで登録される。

「知子の情報」の検索には ,高速検索 ,普通検索 ,および全文検索があるが ,高速検索は登録されたキーワードの先頭 4 バイトのみの検索になるので ,漏れが生じる。普通検索では ,本プログラムによって自動的に登録される「CA 抄録番号 ,論文等の標題 ,著者名 ,誌名 ,巻数(号数) ,ページ-ページ ,使用言語 ,発行年 ,出版社 ,号末の Keyword Index に記載の Keyword 」から検索される。巻末索引 (General Subject Index) の検索語や抄録等も対象に検索する方が望ましいので ,全文検索をお薦めする。

List 1. An Example of Tagged-Formatted Files Exported from CA on CD

Copyright (c) 2001 American Chemical Society All Rights Reserved.

AN Abstract Number (CA の巻数:抄録番号)
 TI 論文等の標題
 AU 著者名
 CS 所属機関名

CS2 住所
SO 誌名, 巻数(号数), ページ-ページ (使用言語) 発行年 出版社.
JT 誌名
VOL 巻数
ISS 号数
PAGE 最初のページ-最後のページ
LANG 使用言語
PY 発行年
COD CODEN
ISSN ISSN
(中略)
DT ドキュメントタイプ
SC CA Section Code
SCXNUM
AB 抄録全文
KW 号末の Keyword Index に記載の Keyword
IN 巻末の General Subject Index に記載の索引語

Copyright (c) 2001 American Chemical Society All Rights Reserved.

(次のデータ)

List 2. An Example of Structure Required for Tomoko Text File

#S (レコードの開始を宣言)
#Wnn (一行の桁数を指定。nn は 5 から 255 までの数字)
#T タイトル (表題を設定)(省略可)
#K キーワード (索引を指定, 少なくとも一つの索引文字列が必要)
#E (レコードの終了を宣言)
#S (次のレコードの開始を宣言)
.
.
#E (次のレコードの終了を宣言)
ファイルの最後に改行のみの行を挿入

List 3. Style of Tomoko Text File Converted from Tagged-Formatted file

```
#S
#K Abstract Number
#T 論文等の標題
#K 著者名
#K 誌名, 巻数(号数), ページ-ページ (使用言語) 発行年 出版社.
#K 発行年
#K 号末の Keyword Index に記載の Keyword
AN から始まって IN に終わるレコードのすべて
#E
```

List 4. List of cascd1.bas

```
REM CA on CD の Tagged-Formatted File を 知子の情報の 連続読み込みテキ
ストファイルに変換
IFNAME$ = COMMAND$
OFNAME$ = IFNAME$
MID$(OFNAME$, INSTR(IFNAME$, ".") + 1) = "TXT"
REM INPUT "入力ファイル名"; IFNAME$ 'DAISIZEN.LOG など
REM INPUT "出力ファイル名"; OFNAME$ 'DAISIZEN.TXT など
DIM A$(500), B$(50), C$(500)
OPEN IFNAME$ FOR INPUT AS #1
OPEN OFNAME$ FOR OUTPUT AS #2
Start$ = "Copyright (c)": End1$ = "Copyright (c)"
flag = 0
DO WHILE flag = 0 AND NOT EOF(1)
    DO WHILE NOT EOF(1)
        LINE INPUT #1, BUN$
        IF INSTR(BUN$, Start$) = 1 THEN
            flag = 1
            EXIT DO
        END IF
    LOOP
```

```

DO WHILE flag = 1
  I = 1
  DO WHILE NOT EOF(1)
    LINE INPUT #1, BUN$
    IF INSTR(BUN$, End1$) = 1 THEN
      EXIT DO
    END IF
    IF EOF(1) THEN
      flag = 0
      EXIT DO
    END IF
    IF LTRIM$(BUN$) <> "" THEN
      A$(I) = BUN$
      I = I + 1
    END IF
  LOOP
  N = I - 1
  J = 2: B$(1) = "#S"
  FOR I = 1 TO N
    IF INSTR(A$(I), "TI  ") = 1 THEN
      B$(J) = "#T" + MID$(A$(I), 6)
      J = J + 1
    END IF
    IF INSTR(A$(I), "AN  ") = 1 THEN
      B$(J) = "#K" + MID$(A$(I), 6)
      J = J + 1
    END IF
    IF INSTR(A$(I), "AU  ") = 1 THEN
      B$(J) = "#K" + MID$(A$(I), 6)
      J = J + 1
    END IF
    IF INSTR(A$(I), "SO  ") = 1 THEN
      B$(J) = "#K" + MID$(A$(I), 6)
      J = J + 1
    END IF
    IF INSTR(A$(I), "PY  ") = 1 THEN

```

```
        B$(J) = "#K" + MID$(A$(I), 6)
        J = J + 1
    END IF
    IF INSTR(A$(I), "KW  ") = 1 THEN
        B$(J) = "#K" + MID$(A$(I), 6)
        J = J + 1
    END IF
NEXT I
K = J - 1
FOR I = 1 TO K
    C$(I) = B$(I)
NEXT I
FOR I = K + 1 TO K + N
    C$(I) = A$(I - K)
NEXT I
C$(K + N + 1) = "#E"
FOR I = 1 TO K + N + 1
    PRINT #2, C$(I)
NEXT I
LOOP
LOOP
PRINT #2, ""
CLOSE
ERASE A$, B$, C$
END
```

References

- 1) 「知子の情報」, (株)テグレット技術開発, 本報では Ver. 8 を使用した。
- 2) 原田茂治, 静岡県立大学短期大学部研究紀要, (4), 173 (1991).
- 3) 「Microsoft QuickBASIC Ver. 4.5」, Microsoft.
- 4) 「Microsoft Visual Basic 6.0」, Microsoft.

(2002年2月4日受理)