

テキスト・データベース「ゲート・ファイル」の公開

樋口, 忠治
九州大学言語文化部

<https://doi.org/10.15017/1468177>

出版情報：九州大学大型計算機センター広報. 21 (3/4), pp.167-176, 1988-07-25. 九州大学大型計算機センター
バージョン：
権利関係：



テキスト・データベース「ゲーテ・ファイル」の公開

樋口 忠治*

1. はじめに

テキスト・データベース「トーマス・マン・ファイル」に続いて、「ゲーテ・ファイル」が今回九州大学大型計算機センターの公用データベースとして一般に公開されることになった。J.W.v.Goethe (1749-1832) のハンプルク版全集(全14巻)に基づくものであるが、今回提供されるのは第6巻から第14巻までのすべての散文テキストである。利用にあたっては、参考文献1及び2を参照のこと。

データベースとしての取り扱いはずべて「トーマス・マン・ファイル」と同様である。ただし、「トーマス・マン・ファイル」のプレフィックスが“A70152B”であるのに対して、「ゲーテ・ファイル」のそれは“A70152C”である点が異なっている。従って、利用者はSIGMAシステムに入ったら、まずプレフィックスの選択をしなければならない。Th. Mann ファイルならば、“A70152B”, Goethe ファイルならば“A70152C”を指定することになる。MEMO ファイルは共通である。両者の相違点を只ひとつあげると、今回のゲーテ・ファイルにおいては、分離した前綴りに“+”記号をつけてはいない。

2. ファイル名について

ファイル名の一括投入が行い易いように、すべてのファイル名に通し番号をつけることにした。この方法の採用により、すべてのファイルについて検索を行う場合は、ファイル名ファイルを別に作らなくても、“S.G*”を入力することによりディレクトリの順序通りに逐次ファイル名が投入されることになる。単一のファイルだけについて検索を行う場合には若干面倒でも“S.GB06”のようにファイル名の指定をしなければならない。ファイル名一覧はプレフィックスを指定した後であれば、“DIR S”あるいは“DDIR S”で見ることができる。ファイル名において“GB”はBd.6~Bd.12, “GC”はBd.13~Bd.14に対応する。

プレフィックスを指定する前にファイルの一覧表示を行った例を示し、各巻に含まれる作品の一覧を掲げる。ファイル一覧は、DDIRECTORY コマンドで表示したものである。OPTION に“A”を指定しているので、ファイル名とその属性(長さ=文字数, 作成日付時刻など)のほか、コメントが表示されている。以下の例において下線部分は利用者による入力を表し、記号◀は利用者による入力復改のみであることを表す。また左の番号は説明に対応するものである。

昭和63年5月31日受理

*九州大学言語文化部

```

READY
SIGMA
SIGMA)DDIR
FILE:=S.'A70152C.*
PASSNUMBER:=R5X8K&W%M%B#
OPTION:=A
  FILENAME                ID ALIAS      SIZE      DATE TIME
GB06                      543    0      1206349 88:04:04 13:39
    BD. 6 ( 7-513)
GB07                      107Y   0      2595909 88:04:04 13:39
    BD. 7 ( 9-610) BD. 8 ( 7-516)
GB09                      2210   0      1851319 88:04:04 13:40
    BD. 9 ( 7-598) BD.10 ( 7-187)
GB10                      302Q   0      1679732 88:04:04 13:41
    BD.10 (188-547)
GB11                      376C   0       727280 88:04:04 13:41
    BD.11 ( 9-556)
GB12                      4074   0       795736 88:04:04 13:41
    BD.12 ( 9-364)
GC13                      4420   0      1097175 88:04:04 13:42
    BD.13 ( 7-523)
GC14                      4902   0       586775 88:04:04 13:42
    BD.14 ( 7-269)
MAXIMEN                   516T   0       326584 88:04:04 13:43
    BD.12 (365-547)
NUM                       5304   0         358 88:04:04 13:43
SMALL                    531A   0         919 88:04:04 13:43

```

TOTAL = 11 PREFIX = A79999A

13999 SECTOR(S) AVAILABLE

DO:LIST S.'A70152C.GB06'

PASS NUMBER FOR GB06:=R5X8K&W%M%B#

#0

#0600714 WIE FROH BIN ICH , DAS ICH WEG BIN !

#0600714 BESTER (FREUND , WAS IST DAS (HERZ DES (MENSCHEN !

#0600715 DICH ZU VERLASSEN , DEN ICH SO LIEBE , VON DEM ICH UNZERTRENNLICH WAR
, UND FROH ZU SEIN !

#0600716 ICH WEIS , DU VERZEIHST MIR'S .

#0600717 WAREN NICHT MEINE =UBRIGEN (VERBINDUNGEN RECHT AUSGESUCHT VOM (SCHICK
SAL . UM EIN (HERZ WIE DAS MEINE ZU =ANGSTIGEN ?

#0600719 DIE ARME (LEONORE !

#0600719 UND DOCH WAR ICH UNSCHULDIG .

#0600720 KONNT' ICH DAF=UR , DAS , W=AHREND DIE EIGENSINNIGEN (REIZE IHRER (SC
HWESTER MIR EINE ANGENEHME (UNTERHALTUNG VERSCHAFFTEN , DAS EINE (LEIDENSCHAFT
IN DEM ARMEN (HERZEN SICH BILDETE ?

#0600723 UND DOCH - BIN ICH GANZ UNSCHULDIG ?

.....

ファイルの一覧表示とファイルの形式

GOETHE'S WERKE

Hamburger Ausgabe in 14 Bänden

ページ

- | | | |
|---------|---------|--|
| Band 6 | 7-124 | Die Leiden des jungen Werther |
| | 125-241 | Unterhaltungen deutscher Ausgewanderten |
| | 242-490 | Die Wahlverwandtschaften |
| | 491-513 | Novelle |
| Band 7 | 9-610 | Wilhelm Meisters Lehrjahre |
| Band 8 | 7-486 | Wilhelm Meisters Wanderjahre |
| | 487-516 | Wilhelm Meisters theatralische Sendung |
| Band 9 | 7-598 | Dichtung und Wahrheit |
| Band 10 | 7-187 | Dichtung und Wahrheit |
| | 188-363 | Campagne in Frankreich |
| | 363-400 | Belagerung von Mainz |
| | 401-428 | Sankt-Rochus-Fest zu Bingen |
| | 429-528 | Tag- und Jahreshefte |
| | 529-547 | Autobiographische Einzelheiten |
| Band 11 | 9-349 | Italienische Reise |
| | 350-556 | Zweiter römischer Aufenthalt |
| Band 12 | 7-223 | Schriften zur Kunst |
| | 224-364 | Schriften zur Literatur |
| | 365-547 | Maximen und Reflexiomen |
| Band 13 | 7-52 | Zur Naturwissenschaft im allgemeinen |
| | 53-250 | Morphologie |
| | 251-303 | Zur Geologie |
| | 304-313 | Zur Witterungslehre |
| | 314-523 | Zur Farbenlehre |
| Band 14 | 7-269 | Materialien zur Geschichte der Farbenlehre |

3. 小文字化

SIGMA領域に格納されている全テキスト・データは大文字のみであるが、利用者が、検索結果をリファイルした場合、そのファイルを大小文字混りの普通のテキストとしてプリンターに出したい場合には次の方法が用意されている。

- 1) DO : REF
- 2) REPORT (Y/N) ? N
- 3) QUESTION : = 1
- 4) NEW RECORD DELIMITER : = #
- 5) NUMBERING (N/Y) ? N
- 6) OUTPUT-FILE : = W
- 7) QUESTION : = ↙
- 8) DO : S. SMALL
- 9) INPUT-FILE : = W
- 10) OUTPUT-FILE : = W
- 11) REPORT (N/Y) ? N
- 12) INPUT-FILE : = ↙
- 13) DO : LOOK

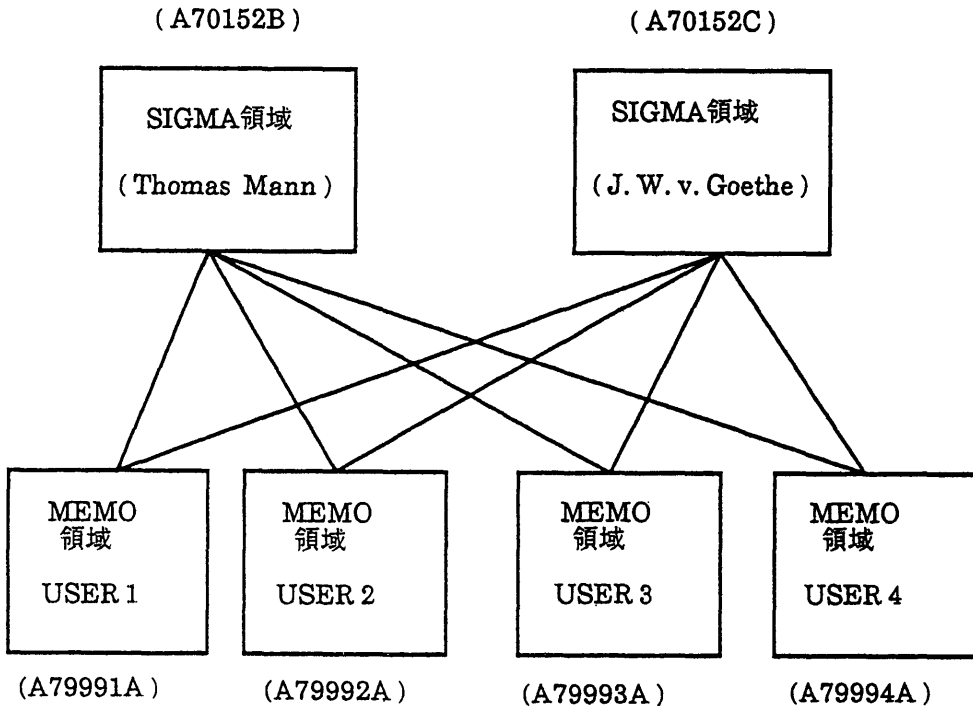
- 1) 検索結果をリファイルする。
- 2) レポートは必要か? 不要。
- 3) 何番目の質問か?
- 4) レコード区切り記号として#を用いる。
- 5) 通し番号は不要。
- 6) どの区域にファイルを作るか?
- 7) 次に何番目の質問をリファイルするか。不要。
- 8) ここでW.1にあるファイルを小文字化するためSIGMA領域に用意されたSMALLというログファイルを呼び出す。
- 9) どのファイルを小文字化するか?
- 10) どこへ小文字化したファイルを置くか?
- 11) レポートが必要か。不要。
- 12) 次にどのファイルを小文字化するか。しない。
- 13) 小文字化したファイルはW.1にあるから、これをLOOKコマンドで見る。

4. パスナンバーについて

ゲート・ファイルの場合、各ファイルにアクセスするためには、パスナンバーのプロンプトに対して答えなければならない。この番号は提供責任者が利用申込者に対して個別に通知するもので、公表はしない。アクセスの度ごとにプロンプトがあるために若干の不便さがあるが、パスナンバーの設定はファイルの保護を行うためのひとつの実験的な試みでもあるから、御協力願いたい。利用者は必ず

事前に連絡をとってパスナンバーを受領し、これを秘密扱いにさせていただきたい。問題が生じた場合にはパスナンバーの変更もありうるので、不明の場合は提供者に問い合わせられたい。

5. 各領域間の概念図とその説明



任意のユーザ(USER 1)の課題番号を仮に“A79991”とする。USER 1はREADYモードで“SIGMA”と入力すると、初めての場合に限りMEMO領域をつくるかどうかを尋ねてくるので、これに対してMEMO領域をつくるように指示しなければならない。MEMO領域の作成がすめば、トーマス・マン・ファイルにも、ゲーテ・ファイルにもアクセス可能の状態になる。

ところで、何れのSIGMA領域のファイルもそれぞれ、“A70152B”および“A70152C”という課題番号を各ファイルの先頭にもっており、USER 1がこれらのSIGMA領域内のファイルにアクセスするためには、完全名を使用しなければならない。例えば、“S.'A70152B.MANN.BB'”あるいは“S.'A70152C.GB06'”のように。しかし、このように複雑なファイル名をいちいち入力しなくてもすむ方法が次のように用意されている。

1)PROFILEコマンドの利用

SIGMAモードまたはDOモードで“PROFILE”と入力し、続いてプロンプト

PROF: に対して“PREFIX”と入力すると、システムがMEMO OR SIGMA(M OR S)?と聞いてくるので、SIGMA領域のプレフィックスを指定するときには“S”と入力する。すると、PREFIX: というプロンプトになるので、ここで各ファイルの先頭部分の共通な部分を指定すれば、以後は各ファイルの指定が簡単になる。

例えば、
PREFIX: = A70152B.MANN

とすれば、トーマス・マン・ファイルの各ファイルはこのプレフィックスを省略した形式、すなわち、例えば、“S.BB”と指定すればよいことになる。ゲータ・ファイルにアクセスするためには、プレフィックスを“A70152C”と指定しなければならない。プロンプトPROF: に対して、“SAVE”と入力すると、設定したプレフィックスが次のシステム利用時にも有効となる。

2)ファイル名ファイルの利用

比較的多数のファイルをたびたび検索する場合には、予め指定するファイル名を完全名で別につくっておき、これをファイルの指定に際して読み込ませる方法がある。トーマス・マン・ファイルの場合にはI~VII巻のすべてをまとめた“FNAME.ROMAN”と、REDEN UND AUFS=ATZEの1~4(IX~XII巻)をまとめた“FNAME.RA”をSIGMA領域に用意しているので、それぞれ、“S.FNAME.ROMAN”および“S.FNAME.RA”のように入力することによって、それらのファイルをまとめて検索することができる。

3)ワイルド・カードの利用

ゲータ・ファイルにおいては、各ファイルに順序番号を付ける方法を採用した。SIGMAシステムにおいてはファイル名の指定にワイルド・カードが使えるので、例えば、“S.G*”とファイル名を指定するとSIGMA領域のファイルで、先頭に“G”の文字を持つすべてのファイルが検索されることになる。この方法をトーマス・マン・ファイルに適用して“S.MANN*”のように指定することはもちろん可能である。しかし、トーマス・マン・ファイルにおいては各ファイルがアルファベット順に並んでおり、巻数に従って順序よく配列しているわけではないから、この方法をとるのは不都合である。

備考1.

ファイル名ファイルを利用すれば、PROFILEコマンドで課題番号(プレフィックス)を変更しなくても、両方のSIGMA領域から一連の必要なファイル呼び出すことができる。もちろん、完全名さえ用いれば、随時、何れの領域からも個々のファイル呼び出せることはいうまでもない。SIGMA領域の完全名とは先頭に“S.”を持ち、その後にシングルクォート'で前後をくくった文字列が続くファイル名で、'でくくられた部分の先頭に課題番号“A70152B.”(トーマス・マン・ファイルの場合)、または“A70152C.”(ゲータ・ファイルの場合)を持つファイル名である。

(例: “S.'A70152B.MANN.BB'”, “S.'A70152C.GB06'”)

備考2.

MEMO領域は各USERの“MEMO87.DATA”というデータセットとしてつくられているから、不必要な時点で、これを消去しない限りいつまでもこのデータセットは存在し、これに対しては期間に応じて課金される。消去はREADYモードで

DEL MEMO87.DATA

とすればよい。

6. 利用例と解説

利用例)

```

DO:PROF
PROF:PREFIX
MEMO OR SIGMA(M OR S)?S
1) PREFIX:=A70152C
PROF:✓
DO:SEA
RECORD DELIMITERS
D01:=#
D02:=✓
ITEM DELIMITERS
D02:=✓
KEYWORDS
2) A01:=F=AHIG
A02:=TR=ACHTIG
A03:=✓
LOGICAL FORMULAS
Z01:=A1
Z02:=A2
Z03:=✓
REPORT(Y/N)?✓
3) FILE:=S.GB*
PASSNUMBER:=R5X8K&W%M%B#
SEARCHING S.A70152C.GB06 ...
RETRIEVED TEXTS
QUESTION 01 (Z01) =      22      22
QUESTION 02 (Z02) =       1       1
TOTAL                =      23      23
CPU (SEC/1000)       =     462     462
SEARCHING S.A70152C.GB07 ...
RETRIEVED TEXTS
QUESTION 01 (Z01) =      80     102
QUESTION 02 (Z02) =       2       3
TOTAL                =      82     105
CPU (SEC/1000)       =     973    1435
SEARCHING S.A70152C.GB09 ...
RETRIEVED TEXTS
QUESTION 01 (Z01) =      49     151
QUESTION 02 (Z02) =       6       9
TOTAL                =      54     159

```


解 説

```

CPU (SEC/1000) = 659 2094
  SEARCHING S.A70152C.GB10 ...
  RETRIEVED TEXTS
QUESTION 01 (Z01) = 32 183
QUESTION 02 (Z02) = 2 11
TOTAL = 34 193
CPU (SEC/1000) = 625 2719
  SEARCHING S.A70152C.GB11 ...
  RETRIEVED TEXTS
QUESTION 01 (Z01) = 25 208
QUESTION 02 (Z02) = 0 11
TOTAL = 25 218
CPU (SEC/1000) = 266 2985
  SEARCHING S.A70152C.GB12 ...
  RETRIEVED TEXTS
QUESTION 01 (Z01) = 37 245
QUESTION 02 (Z02) = 0 11
TOTAL = 37 255
CPU (SEC/1000) = 297 3282
4) FILE:=S.GC*
  PASSNUMBER:=R5X8K&W%M%B#
  SEARCHING S.A70152C.GC13 ...
  RETRIEVED TEXTS
QUESTION 01 (Z01) = 37 282
QUESTION 02 (Z02) = 0 11
TOTAL = 37 292
CPU (SEC/1000) = 393 3675
  SEARCHING S.A70152C.GC14 ...
  RETRIEVED TEXTS
QUESTION 01 (Z01) = 14 296
QUESTION 02 (Z02) = 0 11
TOTAL = 14 306
CPU (SEC/1000) = 216 3891
5) FILE:=S.MAXIMEN
  RETRIEVED TEXTS
QUESTION 01 (Z01) = 15 311
QUESTION 02 (Z02) = 2 13
TOTAL = 17 323
CPU (SEC/1000) = 132 4023
FILE:=✓
6) DO: REF
  REPORT(Y/N)? N
7) QUESTION: = 1
  NEW RECORD DELIMITER: = #
  NUMBERING(N/Y)? ✓
8) OUTPUT-FILE: = W
9) PASS NUMBER FOR GB06: = R5X8K&W%M%B#
  PASS NUMBER FOR GB07: = R5X8K&W%M%B#
  PASS NUMBER FOR GB09: = R5X8K&W%M%B#
  PASS NUMBER FOR GB10: = R5X8K&W%M%B#

```

- PASS NUMBER FOR GB11:=R5X8K&W%M%B#
 PASS NUMBER FOR GB12:=R5X8K&W%M%B#
 PASS NUMBER FOR GC13:=R5X8K&W%M%B#
 PASS NUMBER FOR GC14:=R5X8K&W%M%B#
- 10) QUESTION:=2
 NEW RECORD DELIMITER:=#
 NUMBERING(N/Y)?✓
 OUTPUT-FILE:=W
 PASS NUMBER FOR GB06:=R5X8K&W%M%B#
 PASS NUMBER FOR GB07:=R5X8K&W%M%B#
 PASS NUMBER FOR GB09:=R5X8K&W%M%B#
 PASS NUMBER FOR GB10:=R5X8K&W%M%B#
 PASS NUMBER FOR GB11:=R5X8K&W%M%B#
 PASS NUMBER FOR GB12:=R5X8K&W%M%B#
 PASS NUMBER FOR GC13:=R5X8K&W%M%B#
 PASS NUMBER FOR GC14:=R5X8K&W%M%B#
 QUESTION:=✓
- 11) DO:CAT
 (A1 A2 ... AN ==) B (N=(99))
 FROM-FILE01:=W. 1
 FROM-FILE02:=W. 2
 FROM-FILE03:=✓
 TO-FILE00:=DOLDE
- 12) DO: S. SMALL
 INPUT-FILE:=DOLDE
 OUTPUT-FILE:=W
 REPORT(N/Y)?N
 INPUT-FILE:=✓
- 13) DO:LOOK
 #0635537 wo jeder Trost niedertr=chtig und Verzweiflung Pflicht ist.
 #0747806 " abscheuliche, niedertr=chtige Kupplerin !
 #0755909 man sieht doch ganz niedertr=chtig aus !
 #0901701 dieser =Ubergang hatte manches Unangenehme: denn indem man die bisher
 zu Hause abgesondert, reinlich edel, obgleich streng gehaltenen Kinder unter
 eine rohe Masse von jungen Gesch=opfen hinunterstie\$, so haben sie vom Gemeine
 n, Schlechten, ja Niedertr=achtigen ganz unerwartet alles zu leiden, weil sie
 aller Waffen und aller F=ahigkeit ermangelten, sich dagegen zu sch=utzen.
 #0943712 je weiter ich kam, meine Kleidungsst=ucke anzulegen, desto!
- 14) DO:COPY W X. DOLDE1. TEXT
- 15) DO:END END
 READY

解説)

- 1) プレフィックス“**A70152C**”を指定する。
- 2) ...fähigおよび...trächtigという形容詞の造語例をすべて取り出したい。語尾変化の部分は空白にしないようにすると、全ての変化形を含むことになる。
- 3) **SIGMA**領域(**A70152C**)のファイルの中で“**GB**”に始まるすべてのファイルを指定する。
- 4) “**GC**”に始まるすべてのファイルを指定する。
- 5) “**MAXIMEN**”も指定する。
- 6) リファイルのコマンドを投入する。
- 7) 質問項目1(...fähig)のリファイル指定。
- 8) リファイルしたものをどの区域におくのかを指定する。作業区域のトップに置く。
- 9) アクセスするファイルにパスナンバーがついていれば、そのすべてについて問い合わせてくる。
- 10) 次に、質問項目2(...trächtig)のリファイル指定。
- 11) **CATENATE**(ファイルの接続)コマンドを用いて、質問項目1と2のファイルをひとつのファイルにまとめ、今度はメモ区域に“**DOLDE**”というファイル名をつけておく。作業区域はスタックになっているので、**W.1**には質問項目2に対する結果が、**W.2**には質問項目1に対応する結果がはいっている。
- 12) 上記ファイルを小文字化してみたいので、**SIGMA**領域のログファイル“**SMALL**”を呼び出す。出力先は作業区域のトップとする。
- 13) **LOOK**コマンドでファイルを見る。
- 14) このファイルを**MEMO**領域から取り出し、外部データセットに作る。データセット名は“**DOLDE1.TEXT**”
- 15) **SIGMA**を終了する。

参考文献

1. 樋口, 篠原 テキスト・データベース「トーマス・マン・ファイル」の完成と再編成について, 九州大学大型計算機センター広報, Vol.20, No. 6, 1987, pp. 582-596.
2. 有川 他 テキスト・データベース管理システム**SIGMA**第2版について, 九州大学大型計算機センター広報, Vol.20, No. 6, 1987, pp. 517-581.

使用上の注意

「ゲータ・ファイル」のソース・データ自体のコピーを作ることは如何なる目的のためであっても、これを禁止する。