

公用データベース「トーマス・マン・ファイル」の ファイル追加について

樋口, 忠治
九州大学教養部

篠原, 武
九州大学大型計算機センター

<https://doi.org/10.15017/1468113>

出版情報：九州大学大型計算機センター広報. 18 (2), pp.73-79, 1985-03-15. 九州大学大型計算機センター

バージョン：

権利関係：

公用データベース「トーマス・マン・ファイル」のファイル追加について

樋口 忠治*, 篠原 武**

1. データの追加

このたびトーマス・マン全集(全12巻)のうち文学作品を収めた第1巻から第8巻までの全作品のテキスト・データ入力が終わり、「トーマス・マン・ファイル」[1]に追加しました。その他、第9巻の「評論」も合わせて収めています。今回のファイル追加によって、トーマス・マンの文学作品のすべてが検索の対象として扱えることになったわけです。

追加したデータのうち、第4巻と第5巻の「ヨーゼフ物語」はページが通し番号になっています。第4巻は9ページから903ページまで、第5巻は913ページから1822ページまでとなっています。ファイルのデータ中の「文番号」は7ケタで、先頭2ケタが巻、次の3ケタがページ、最後の2ケタが行を表わしています。第5巻の1000ページを超える部分については、▼0500036▼は1000ページを、▼0500320▼は1003ページをといた具合に、下3ケタのみが文番号に収められていますので注意してください。

今回の追加によってファイル総数は78になりました。また、誤りの多かった▼ERZ.TONIO▼および▼ERZ.VENEDIG▼の訂正も行いました。追加・修正後のファイルの一覧を表1に示します。表1のファイル名は、すべて共通した部分F1683.MANNを省略しています。完全名は、S.▼1683.MANN.ERZ.TONIO▼のようになります。前回のお知らせ[2]以後、追加したファイルは、「ヨーゼフ物語」の▼JS.N01▼-▼JS.N15▼、第9巻「評論」の▼RA.N01▼-▼RA.N07▼、および、短篇の▼ERZ.BAJAZZO▼、▼ERZ.GLADIUS▼、▼ERZ.STUNDE▼です。

2. SEARCHの用法について

トーマス・マン・ファイルはSIGMAシステム[3]を用いて利用します。SIGMAの詳しい利用法については文献を参照してください。ここでは、検索コマンドSEARCHの用法について、検索結果から「文番号」だけを取り出す方法、複数のファイル名を一括して入力する方法、および、検索結果を再ファイル化する方法について紹介します。これらはSIGMAの新しい機能ではなく、LOGファイルの機能を用いてコマンドを組み合わせることで実現されています。

* 九州大学教養部

** 九州大学大型計算機センター

表1. トーマス・マン・ファイルのファイル一覧

ファイル名	作成更新	文字数	ファイル名	作成更新	文字数
BB.N01	83:02:16	189521	JS.N04	84:03:24	297893
BB.N02	83:02:16	181629	JS.N05	84:03:24	323120
BB.N03	83:02:16	185585	JS.N06	84:03:24	219701
BB.N04	83:02:16	127367	JS.N07	84:03:24	331241
BB.N05	83:02:16	210072	JS.N08	84:03:24	209260
BB.N06	83:02:16	188230	JS.N09	84:03:24	332311
BB.N07	83:02:16	185391	JS.N10	84:05:29	298167
BB.N08	83:02:16	205319	JS.N11	84:05:29	285277
BB.N09	83:02:16	274073	JS.N12	84:05:29	306952
DF.N01	83:02:16	202066	JS.N13	84:05:29	217211
DF.N02	83:02:16	171537	JS.N14	84:05:29	198581
DF.N03	83:02:16	208395	JS.N15	84:05:29	238542
DF.N04	83:02:16	207433	KH.N01	83:05:10	262898
DF.N05	83:02:16	228104	KH.N02	83:05:10	141448
DF.N06	83:02:16	157461	KH.N03	83:05:10	246699
DF.N07	83:02:16	164583	KH.N04	83:05:10	181993
DF.N08	83:02:16	210407	LT.N01	83:02:16	248688
ERZ.BAJAZZO	84:03:24	209973	LT.N02	83:02:16	197060
ERZ.BETRO	83:02:16	175200	LT.N03	83:02:16	138307
ERZ.GESETZ	83:07:14	156500	LT.N04	83:02:16	190385
ERZ.GLADIUS	84:03:24	198449	LT.N05	83:02:16	161251
ERZ.HERR	83:02:16	216466	RA.N01	84:11:27	379896
ERZ.KOEPFE	83:07:14	219840	RA.N02	84:11:27	263122
ERZ.LEID	83:07:14	227419	RA.N03	84:11:27	297170
ERZ.MARIO	83:02:16	128239	RA.N04	84:11:27	232116
ERZ.STUNDE	84:03:24	184196	RA.N05	84:11:27	249376
ERZ.TONIO	84:11:27	158171	RA.N06	84:11:27	324585
ERZ.TRISTAN	83:02:16	108059	RA.N07	84:11:27	373189
ERZ.VENEDIG	84:11:27	191196	ZB.N01	83:02:16	195348
EW.N01	83:07:14	217113	ZB.N02	83:02:16	211415
EW.N02	83:07:14	226248	ZB.N03	83:02:16	171045
EW.N03	83:07:14	149806	ZB.N04	83:02:16	223121
FK.N01	83:02:16	193046	ZB.N05	83:02:16	229796
FK.N02	83:02:16	236247	ZB.N06	83:02:16	205143
FK.N03	83:02:16	230496	ZB.N07	83:02:22	264544
FK.N04	83:02:16	266583	ZB.N08	83:02:16	235798
JS.N01	84:03:24	245464	ZB.N09	83:02:16	244479
JS.N02	84:03:24	282099	ZB.N10	83:02:16	228052
JS.N03	84:03:24	335830	ZB.N11	83:02:16	110647

2.1 文番号の取り出し

検索の目的によっては全ファイルを対象にする必要があるでしょうが、これを1回で済ませようとすると、出力データ量が膨大になる可能性があります。そのような場合には検索結果から文番号だけを取り出すことができます。

文番号を取り出すためにはLOGファイルを用います。まず、次の手順で、このためのLOGファイルを利用者自身のMEMO域に取り込んでください。以下の説明では、下線部が利用者の入力を表わします。

```
READY
SIGMA
SIGMA> LOAD S.'F1683.LOG.NUM'
DO: PUT NUM 0
DO:
```

これで、準備は終了です。確認のためLISTコマンドでこのファイルの内容を表示して見てください。

次のようになるはずですが、

```
DO: LIST_NUM
DO: REFILE

  QUESTIONS:=.K.1
  RECORD DELIMITER:=#
  NUMBERING (N/Y)?
  SORT ON:
DO: SEA D
D1:=!
D2:=
D3:=
A1:=#
A2:=
Z1:=A1
Z2:=

FILE:=W

FILE:=
LIST OF RESULTS (N/Y)?
DO: REF

  QUESTIONS:=1
  RECORD DELIMITER:=
  NUMBERING (N/Y)?
  SORT ON:
DO: MOVE T.2 T
DO: TERM LI 73
TERM:
DO: END

DO: _
```

LOGファイルNUMの内容

この使用法を図1の例に沿って説明します。

- (1) SEARCHコマンドにより検索を行います。
- (2) レコード区切を指定します。
- (3) キーワードを登録します。空白記号も意味を持つので注意してください。この例では、単語としてのDAMITを検索するために、両側に空白を置いています。
- (4) 論理式の登録をします。ここでは、単にキーワードA1を持つものの検索を指示しています。
- (5) 検索対象のファイル名を入力します。
- (6) 検索結果が34件あったことが表示されています。
- (7) 引き続き検索するファイル名を入力します。
- (8) 検索結果が15件あったことが表示されています。
- (9) 検索するファイルがない場合には、単に復改キーを押します。
- (10) 検索結果を表示しないので、単に復改キーを押します。ここでYを入力すると検索結果の表示をします。

以上で、SEARCHコマンドによる検索が終了しています。次に、この検索結果から文番号だけを取り出すために、LOGファイルNUMを実行します。LOGファイルを実行するためには、先頭にピリオドを付加したファイル名と空白とYまたはN

.LOGファイル名 {Y|N}

```

(1) DO: SEA D

      RECORD DELIMITERS
(2)  D1:=#
      D2:=

      KEYWORDS
(3)  A1:= DAMIT
      A2:=

      LOGICAL FORMULAE
(4)  Z1:=A1
      Z2:=

(5)  FILE:=S.'F1683.MANN.JS.NO7'

      RETRIEVED TEXTS
(6)  TOTAL                =    34
      QUESTION 1 (Z1 ) =    34
      CPU TIME =          350

(7)  FILE:=S.'F1683.MANN.JS.NO8'

      RETRIEVED TEXTS
(8)  TOTAL                =    15
      QUESTION 1 (Z1 ) =    15
      CPU TIME =          214

(9)  FILE:=
(10) LIST OF RESULTS (N/Y)?

(11) DO: NUM N

(12) QUESTIONS:=1

      RETRIEVED TEXTS
(13) TOTAL                =    49
      QUESTION 1 (Z1 ) =    49
      CPU TIME =          51
(14) DO: LOOK
      #0476113 #0476232 #0476931 #0477005 #0477501 #0477536 #0478218 #0478317-
      #0478628 #0479009 #0479106 #0479527 #0479805 #0480311 #0480314 #0481209-
      #0481306 #0481423 #0481435 #0483036 #0483316 #0485333 #0486009 #0486035-
      #0487010 #0488132 #0488135 #0488721 #0489212 #0489218 #0489628 #0489705-
      #0489721 #0490227 #0592102 #0593726 #0594134 #0594406 #0594801 #0594929-
      #0595313 #0595927 #0597928 #0598124 #0598409 #0599032 #0599509 #0599722-
      #0500002
      DO:_

```

図1. 文番号の取り出し

を入力します。YはLOGファイルの内容を表示しながら実行するときに、Nは表示しないときに用います。

(1) LOGファイルNUMを実行します。

(2) どの検索結果の文番号を取り出すかを指示します。ここでは、質問1の結果を処理することを指示しています。複数の結果がある場合には必要な質問番号(論理式番号)を空白で区切って指示します。

(3) 検索結果が合計49件であったことを表示しています。

これで、検索結果から文番号が取り出されました。この結果は、WORK ファイルに置かれます。

(4) LOOK コマンドによって取り出された文番号を表示しています。

この文番号を保存するためには、PUT コマンドを用います。保存したファイルの内容表示には、LIST コマンドを用います。次の例を参考にしてください。

```
DO: PUT KEKKA 0
DO: LIST KEKKA
#0476113 #0476232 #0476931 #0477005 #0477501 #0477536 #0478218 #0478317-
#0478628 #0479009 #0479106 #0479527 #0479805 #0480311 #0480314 #0481209-
#0481306 #0481423 #0481435 #0483036 #0483316 #0485333 #0486009 #0486035-
#0487010 #0488132 #0488135 #0488721 #0489212 #0489218 #0489628 #0489705-
#0489721 #0490227 #0592102 #0593726 #0594134 #0594406 #0594801 #0594929-
#0595313 #0595927 #0597928 #0598124 #0598409 #0599032 #0599509 #0599722-
#0500002
DO: _
```

2.2 ファイル名の一括入力

LOG ファイルを利用すれば、SEARCH コマンドで複数のファイル名をまとめて入力することができます。この方法を用いれば、作品を単位とした検索が容易になります。このためには、まず、次のように、TSS の EDIT コマンドを用いて外部データセットにファイル名を入力しておきます。

```
READY
E JS.DATA
KEQ52320I DATA SET NOT FOUND, ASSUMED TO BE NEW
INPUT
00010 _
00020 FILE:=S.'F1683.MANN.JS.N01'
00030 _
00040 FILE:=S.'F1683.MANN.JS.N02'
...
00290 _
00300 FILE:=S.'F1683.MANN.JS.N15'
00310 _
00320 FILE:=.E
00330
E
UNNUM
E
END S
KEQ52460I SAVED IN DATA SET 'F9999.JS.DATA'
READY
```

次に、この外部データセットをMEMO域に取り込みます。

```
READY
SIGMA
SIGMA> LOAD X.JS.DATA
DO: PUT JS 0
DO: _
```

これで、準備は完了です。次の例は、このLOGファイルJSを用いて検索したものです。LOGファイル実行は、SEARCHコマンドのファイル名に、ピリオド+LOGファイル名+空白+Yを指示します。ここでは、ファイル名がJSですから、`▼.JS Y▼`を入力します。Yは実行しているLOGファイルの内容を表示するために指示します。ここでは、検索しているファイル名を表示することになります。

解 説

```
DO: SEA D

RECORD DELIMITERS
D1:=#
D2:=

KEYWORDS
A1:= WENN ... AUCH
A2:=

LOGICAL FORMULAE
Z1:=A1
Z2:=

FILE:=.JS Y

FILE:=S.'F1683.MANN.JS.NO1'

RETRIEVED TEXTS

TOTAL           =      8
QUESTION 1 (Z1) =      8
CPU TIME =      234

...

FILE:=S.'F1683.MANN.JS.N15'

RETRIEVED TEXTS

TOTAL           =      6
QUESTION 1 (Z1) =      6
CPU TIME =      241

FILE:=.E

FILE:=
LIST OF RESULTS (N/Y)?
DO: _
```

この例のように、結果が多い場合には、文番号だけを取り出すこともできます。そのためには、すでに説明したように、LOG ファイル NUM を用います。

```
DO: .NUM N

QUESTIONS:=1

RETRIEVED TEXTS

TOTAL           =      210
QUESTION 1 (Z1) =      210
CPU TIME =      251
DO: LOOK
#0400932 #0401336 #0402915 #0405932 #0406611 #0407504 #0407619 #0409132-
#0413501 #0415509 #0416812 #0417401 #0417931 #0419222 #0419627 #0419814-
#0419902 #0420104 #0422601 #0424525 #0425418 #0425524 #0425707 #0426027-
#0426735 #0426904 #0427104 #0427304 #0427703 #0427706 #0427832 #0428122-

...

#0569609 #0570005 #0570109 #0570412 #0572532 #0572703 #0574907 #0576322-
#0576502 #0578501
DO: _
```

2.3 検索結果の再ファイル化

検索結果が多量になった場合の解決法の一つは、文番号だけを取り出すことですが、これは、原典が手元になれば有効とはいえません。もし原典が手元にない場合には、検索結果をREFILEコマンドによって再ファイル化すると便利です。再ファイル化された結果は、SEARCHコマンドによって検索することもでき、SAVEコマンドによって外部データセットに書き出して、TSSのEDITコマンド等で自由に扱うこともできます。次の例では、再ファイル化した結果にKEKKA1と名前を付けて保存しています。

```
DO: REF

QUESTIONS:=1
RECORD DELIMITER:=#
NUMBERING (N/Y)?
SORT ON:
DO: PUT KEKKA1 0
DO:
```

LOGファイルNUMでは、文番号だけを取り出すために、まずREFILEによって検索結果の再ファイル化を行い、そのファイルに対してSEARCHコマンドで文番号だけを検索し、その結果を再ファイル化しています。従って、LOGファイルNUMを実行すると、作業ファイルのトップから2番目に検索結果を再ファイル化したものが残るので、これを利用することができます。この内容を表示するためには、

```
DO: LIST W.2
```

とします。また、これに名前を付けて保存するためには、

```
DO: MOVE W.2 KEKKA1 0
```

のようにMOVEコマンドを用います。さらに、これを外部データセットに書き出すためには、COPYコマンドを用います。

```
DO: COPY KEKKA1 X.EXTERNAL.DATA
```

これで、TSSのEDITコマンドを用いて編集したり、LISTコマンドでラインプリンタに出力したりすることができます。

参考文献

1. 樋口, 篠原 公用データベース トーマス・マン・ファイル/SIGMAの公開, 九大大型計算機センター広報, 16, 4, 1983, 379-393.
2. トーマス・マン・ファイルについてのお知らせ, 九大大型計算機センターニュースNo. 279.
3. 有川, 篠原, 白石, 玉越 研究者向き情報システム SIGMAについて, 九大大型計算機センター広報, 14, 4, 1981, 550-573.