公用データベース　トーマス・マン・ファイル/SIGMAの公開

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/10.15017/1468080

出版情報：九州大学大型計算機センター広報．16（4），pp.379-393，1983-07-25．九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／ＳＩＧＭＡの公開

樋口　忠治*，篠原　　武**

1. はじめに

トーマス・マン・ファイルは，この作家の文学作品に現われる全語形の索引を作る目的で，10年以上前に機械可読化を始めたものである．しかしながら，その目的の遂行には，相応の研究資源を費やさねばならない．一方，研究者向き情報システムＳＩＧＭＡ［２，３］を用いて，トーマス・マン・ファイルをオンライン・テキスト・データベースとして利用すれば，索引の作成等の労力を必要とせずに，様々な用例の検索を自由に行うことができる．そこで，これを公用データベース［１］として多くの研究者に公開することにした．トーマス・マン・ファイルが一人でも多くのドイツ語学・文学の研究者に利用されることを希望する．また，このトーマス・マン・ファイルが試金石となって，ドイツ語のみならず他の言語研究の分野においても，この種のテキスト・データベースの作成が盛んになることを願うものである．

2. 言語研究とテキスト・データベース

言語の研究にとっては，テキストの中において語や句の用法がどのような意味を持つのか調べる作業が欠かせない．辞書は一般的な意味を参照するためにあるのであって，具体的な用例を「公式化」したものにすぎない．従って，言語の研究は具体的なテキストに基づいて行わなければならない．個々の単語や句はそれだけでは具体的な意味は持たない．それらは，テキストの中にあって初めて躍動する具体的な意味を持つのである．このような「生きた」語の意味を失うことなく研究の対象とするためには，ばらばらに解体された「単語」ではなく，常にテキストの中の「語」，文の中の「語」として取り扱う必要がある．

外国語の研究にとって，研究の対象となるテキストは実に様々である．それはまず書き言葉と話し言葉に大別できる．言語の研究は主として書き言葉を対象としてきた．話し言葉の研究は記録方式に本質的な困難さがあるためにそう簡単には進展しないであろう．一方，書き言葉の研究の歴史は永い．多くの研究者は著名な文章家（詩人や作家）の用法を書き抜いて，これによってある語や句などの用法を説明したのであり，今日でもその習慣は続いている．しかし，現代のように出版物があふれ言葉が氾濫している時代では，取り扱うべき資料も多くなり，これまでのように特定の文章家の模範的な文章だけを対象に研究していたのでは，言語の実体や言語の変化は捉えられない．

従って，言語の研究にとって，テキスト・データベースは非常に有用であるといえる．すなわち，研究対象のテキストを機械可読な形式にしさえすれば，言語の研究にとって基本的な作業である用例を集めるという機械的作業の大部分を計算機に代行させることができ，人間でなければできないような仕事に研究者は専念することができるのである．トーマス・マン・ファイルは，ドイツ語はもちろんその他の言語研究におけるテキスト・データベースの有用性・有効性を実証するために今回公開す

＊　九州大学教養部
＊＊　九州大学大型計算機センター

るものである．

トーマス・マン・ファイルは，研究者向き情報システムSIGMAを用いており，大学間ネットワークを通じて全国の研究者が利用可能なオンライン・テキスト・データベースである．これまでは，テキスト・データの作成に莫大な時間と労力を要するという問題があったが，これもテキスト・データベースの有用性・有効性が実証されれば，ネットワークを利用した作業分担などによって容易に解決できよう．今回の公開が，今後多くの研究者が様々なテキスト・データベースを作成し有効に活用するために役立つと信じる．

3. 研究者向き情報システムSIGMA

SIGMAシステムは，次のような点においてこの種のテキスト・データベースを扱うのに適している．

1）対象となるデータはすべて文字列である．

2）文字列の中から任意の「語」や「句」を含む部分列を検索することができる．

3）検索に際して，複数のキーワードを任意に組合わせた複数の質問を同時に処理できる．

4）研究者が協同し，そのグループ用のデータベースを構築できる．

1）は別の目的で作成されたデータをほとんどそのまま扱えることを意味し，また形式を問わないので新たにデータを作成する場合の作業が容易である．2）及び3）は特に言語の研究においては欠かせない重要な機能である．たとえば動詞のすべての変化形を求める場合など3）の機能により極めて効率よく処理できる．4）は公用データベースとしてトーマス・マン・ファイルなどのテキスト・データベースを構築・維持するのに必要な機能である．

SIGMAシステムの領域は，各利用者の作業のためのMEMO領域と利用者グループが協同してデータベースを構築したり，でき上がったデータベースを共用するためのSIGMA領域とから構成されている．このようすを表わしたのが図1であり，利用者UID2のSIGMA領域は，UID1，…，UIDnによって共有されている．トーマス・マン・ファイルはひとつのSIGMA領域として提供される．

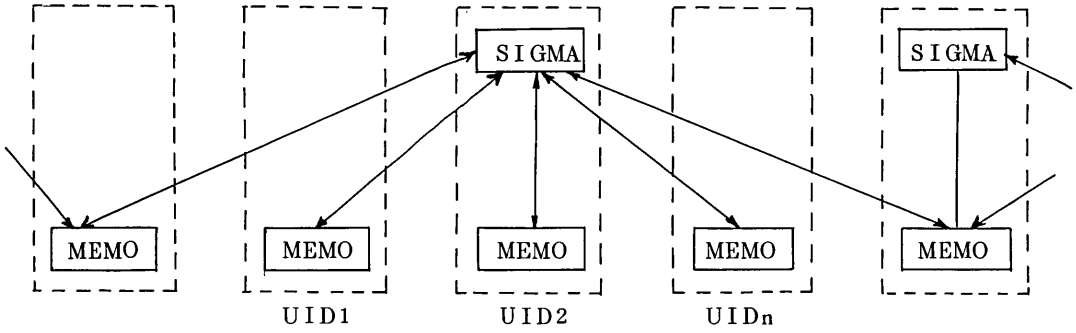

図1. SIGMAシステムの領域

SIGMAシステムのコマンドとその機能は次のとおりである．下線部は，コマンドの最短省略形であり，例えばSEARCHはSE，SEA，SEAR，SEARC，SEARCHのいずれを用いてもよい．

1. END … DO：を終える，ＳＩＧＭＡを終える．
2. SETMEMO … MEMO領域を初期化する．
3. SETSIGMA … ＳＩＧＭＡ領域を初期化する．
4. DIRECTORY … ディレクトリ（ファイル名一覧）を表示する．
5. DDIRECTORY … ディレクトリをファイルの作成日付，長さと合わせて表示する．
6. DELETE … ファイルを除去する．（消去ではない）
7. TERMINAL … システム定数を表示したり，設定したりする．
8. PUT … WORKファイルを指定のところに置く．
9. GET … 指定されたファイルをWORKへ取ってくる．
10. MOVE … ファイルを移動する．
11. NICKNAME … ファイルに別名をつける．
12. SAVE … WORKファイルの内容を指定のファイルに書き出す．
13. LOAD … 指定されたファイルの内容をWORKに読み込む．
14. COPY … ファイルの複製を作る．
15. LIST … ファイルの内容を表示する．
16. LOOK … WORKファイルの内容を表示する．
17. KEYIN … キーボードから直接WORKファイルに入力する．
18. SEARCH … 逐字検索システムSEARCHを起動する．
19. REFILE … SEARCHの検索結果を再ファイル化してWORKに置く．
20. REPLACE … 複数文字列の同時置き換えをする．
21. CATENATE … ファイルの連接をする．
22. SORT … レコードのソーティングをする．
23. KSORT … 重複をカウントして単語のソーティングをする．
24. TSS … TSSコマンドを呼び出す．
25. EDIT … ファイルを編集する．（現在使用できない）

ＳＩＧＭＡシステムの詳細については文献［２］を参照していただくことにして，本稿では以下，トーマス・マン・ファイルの内容及びSEARCHコマンドによる用例検索の方法を中心に説明していくことにする．

4. トーマス・マン・ファイルの構成

現在トーマス・マン・ファイルに収められている作品は西独S. Fischer 版Thomas Mann全集のうち

第１巻　Buddenbrooks，
第２巻　Königliche Hoheit，
　　Lotte in Weimar，
第３巻　Die Zauberberg，

第6巻　Doktor Faustus，
第7巻　Bekenntnisse des Hochstaplers　Felix Krull，
第8巻　Tristan，
　　Tonio Kröger，
　　Der Tod in Venedig，
　　Herr und Hund，
　　Mario und der Zauberer，
　　Die Betrogene

の12作品である．今後追加してファイル化する予定の作品は，

第7巻　Der Erwählte，
第8巻　Die vertauschten Köpfe，
　　Das Gesetz

であり，データが完成しだい追加していく．

4.1　ファイルの分割とファイル名

トーマス・マン・ファイルのファイル名と作品との対応を示したものが表2である．それぞれのファイルはもとのテキストにして約100頁分程度の大きさにまとめられている．従って，長編の作品はいくつかに分割されている．検索は作品を単位として行うのが望ましいが，システム全体のバランスからこのようにファイルの大きさを制限することにした．しかし，ファイルの区分は章や節を単位として行っているので文の途中でファイルが切れているというような不都合は一切ない．ファイルの大きさに不揃いがあるのはこのためである．

ファイルの完全名は

S. 'F1683. MANN. x. y'

の形式である．xは作品名を表わす省略記号で，ただし全集第8巻に限りERZ(Erzählungen，短編の意)である．yは分割した作品の順序番号で，形式は先頭がN，次に2ケタの数01，02，…が続く，ただし第8巻の場合はここが作品名(長いものは短縮形)である．

4.2　ファイル名の入力に関する注意

次章で説明されるSEARCHコマンドの使用に際して検索対象を指示するためにファイル名を用いるのであるが，すべての場合にいちいち完全名を使用するのは煩わしいので，次のようにTERMINALコマンドを用いてシステム定数であるPREFIXを変更して省略形を用いると便利である．SIGMAシステムが呼び出された時点においては，PREFIXは利用者の課題番号(ユーザID)となっているので，トーマス・マン・ファイルを利用する場合には，まず始めにPREFIXを変更すべきである．

```
SIGMA〉 TERM (または DO: TERM)
TERM: P
PREFIX:=F1683. MANN
```

表2. トーマス・マン・ファイルのファイル名

ファイル名	作品名	大きさ（文字数）
S.'F1683.MANN.BB.N01'	Buddenbrooks	189521
S.'F1683.MANN.BB.N02'		181629
S.'F1683.MANN.BB.N03'		185585
S.'F1683.MANN.BB.N04'		127367
S.'F1683.MANN.BB.N05'		210072
S.'F1683.MANN.BB.N06'		188230
S.'F1683.MANN.BB.N07'		185391
S.'F1683.MANN.BB.N08'		205319
S.'F1683.MANN.BB.N09'		274073
S.'F1683.MANN.DF.N01'	Doktor Faustus	202066
S.'F1683.MANN.DF.N02'		171537
S.'F1683.MANN.DF.N03'		208395
S.'F1683.MANN.DF.N04'		207433
S.'F1683.MANN.DF.N05'		228104
S.'F1683.MANN.DF.N06'		157461
S.'F1683.MANN.DF.N07'		164583
S.'F1683.MANN.DF.N08'		210407
S.'F1683.MANN.ERZ.BETRO'	Die Betrogene	175200
S.'F1683.MANN.ERZ.HERR'	Herr und Hund	216466
S.'F1683.MANN.ERZ.MARIO'	Mario und der Zauberer	128239
S.'F1683.MANN.ERZ.TONIO'	Tonio Kröger	158405
S.'F1683.MANN.ERZ.TRISTAN'	Tristan	108059
S.'F1683.MANN.ERZ.VENEDIG'	Der Tod in Venedig	190399
S.'F1683.MANN.FK.N01'	Bekenntisse des Hochstaplers Felix Krull	193046
S.'F1683.MANN.FK.N02'		236247
S.'F1683.MANN.FK.N03'		230496
S.'F1683.MANN.FK.N04'		266583
S.'F1683.MANN.KH.N01'	Königliche Hoheit	262898
S.'F1683.MANN.KH.N02'		141448
S.'F1683.MANN.KH.N03'		246699
S.'F1683.MANN.KH.N04'		181993
S.'F1683.MANN.LT.N01'	Lotte in Weimar	248688
S.'F1683.MANN.LT.N02'		197060
S.'F1683.MANN.LT.N03'		138307
S.'F1683.MANN.LT.N04'		190385
S.'F1683.MANN.LT.N05'		161251
S.'F1683.MANN.ZB.N01'	Der Zauberberg	195348
S.'F1683.MANN.ZB.N02'		211415
S.'F1683.MANN.ZB.N03'		171045
S.'F1683.MANN.ZB.N04'		223121
S.'F1683.MANN.ZB.N05'		229796
S.'F1683.MANN.ZB.N06'		205143
S.'F1683.MANN.ZB.N07'		264544
S.'F1683.MANN.ZB.N08'		235798
S.'F1683.MANN.ZB.N09'		244479
S.'F1683.MANN.ZB.N10'		228052
S.'F1683.MANN.ZB.N11'		110647

TERM:

DO:

この状態では

完全名： S.'F1683. MANN. ERZ. TONIO' を

省略名： S.ERZ. TONIO で代用することができる.

ここで注意しなければならないことは，先頭の"S. "はいかなる場合でも省略できないことである．もしもファイル名の先頭が"S. "で始まっていなければ，それは各利用者自身のMEMO ファイルを表わすことになるからである.

上の例では，PREFIXが"F1683. MANN"であるが，さらに長いPREFIXを指定することもできる．PREFIXを"F1683. MANN. ZB"とすれば，完全名： S. 'F1683. MANN. ZB. N01'を単にS. N01で指定することができる．この場合は検索の対象がDer Zauberbergに限られている時に便利である．ただし，このように長いPREFIXを用いている場合には，常に現在のPREFIXの状態を知っていなければ誤ったファイルを検索することになるので注意しなければならない．現在のPREFIXの状態を知るためのコマンドが用意されているので次のように使用すればよい.

DO: TERM DISP

4.3 ファイルの形式

言語研究のための用例検索の単位は主として文章であり，検索された文章の原作品中での出現位置を知ることも重要な目的の一つである．SIGMAシステムでは検索の単位（＝レコード）は指定された記号列によってはさまれた部分である．この記号列をレコード区切り語（record delimiters）と呼ぶ.

トーマス・マン・ファイルでは#を文章の区切りとして挿入してあるので，検索時にはレコード区切り語に#を指定すればよい．さらに文章の出現位置を表わす7ケタの数字列を各文の先頭に付加した．つまりファイルの形式は，

…#VVPPPLL 文章 #VVPPPLL…

であり，VVは巻，PPPは頁，LLは行を表わした数字列である．また，パラグラフの区切りには@を挿入しているので，パラグラフ単位の検索も可能である.

ドイツ語文章を計算機に格納する場合には，いくつかの問題が生じるが，トーマス・マン・ファイルでは，次のような形式で文章を格納することにした.

1）文章は・（ピリオド）？（疑問符）！（感嘆符）で区切られるものとして，機械的に処理する.

2）単語の前後に必ず空白を入れる.

3）アルファベットは，すべて大文字を用いる.

4）名詞の先頭は本来大文字表記すべきものであるから，その直前に＜記号を置く.

5）変母音ä，ö，üはそれぞれ＝A，＝O，＝Uとし，それらが大文字である場合は＜＝A，＜＝O，＜＝Uとする.

6）文頭の大文字は，それが名詞である場合を除いて，＜記号を付けない.

7）ß（エスツェット）はSSではなく通貨記号$で表わす.

８）引用記号〈及び〉はともに〉で表わす．

９）分離動詞の分離した前つづりはその前に＋記号を付ける（前置詞と同形のもののみ）．

例えば，Der Zauberberg の冒頭部分は，次のような形式で格納されている．

```
#@
#0300907 DIE <GESCHICHTE <HANS <CASTORPS , DIE WIR ERZ=AHLEN WOLLEN , - NICHT
UM SEINETWILLEN  ( DENN DER <LESER WIRD EINEN EINFACHEN , WENN AUCH ANSPRECHEN
DEN JUNGEN <MENSCHEN IN IHM KENNENLERNEN  ) , SONDERN UM DER <GESCHICHTE WILLE
N , DIE UNS IN HOHEM <GRADE ERZ=AHLENSWERT SCHEINT ( WOBEI ZU <HANS <CASTORPS
<GUNSTEN DENN DOCH ERINNERT WERDEN SOLLTE , DA$ ES SEINE <GESCHICHTE IST , UND
 DA$ NICHT JEDEM JEDE <GESCHICHTE PASSIERT  > : DIESE <GESCHICHTE IST SEHR LAN
GE HER , SIE IST SOZUSAGEN SCHON GANZ MIT HISTORISCHEM <EDELROST =UBERZOGEN UN
D UNBEDINGT IN DER <ZEITFORM DER TIEFSTEN <VERGANGENHEIT VORZUTRAGEN .
#@
#0300917 DAS W=ARE KEIN <NACHTEIL F=UR EINE <GESCHICHTE , SONDERN EHER EIN <VO
RTEIL ; DENN <GESCHICHTEN M=USSEN VERGANGEN SEIN , UND JE VERGANGENER , K=ONNT
E MAN SAGEN , DESTO BESSER F=UR SIE IN IHRER <EIGENSCHAFT ALS <GESCHICHTEN UND
 F=UR DEN <ERZ=AHLER , DEN RAUNENDEN <BESCHW=ORER DES <IMPERFEKTS .
#0300921 ES STEHT JEDOCH SO MIT IHR , WIE ES HEUTE AUCH MIT DEN <MENSCHEN UND
UNTER DIESEN NICHT ZUM WENIGSTEN MIT DEN <GESCHICHTENERZ=AHLERN STEHT : SIE IS
T VIEL =ALTER ALS IHRE <JAHRE , IHRE <BETAGTHEIT IST NICHT NACH <TAGEN , DAS <
ALTER , DAS AUF IHR LIEGT , NICHT NACH <SONNENUML=AUFEN ZU BERECHNEN ; MIT EIN
EM <WORTE : SIE VERDANKT DEN <GRAD IHRES <VERGANGENSEINS NICHT EIGENTLICH DER
<ZEIT , -  EINE <AUSSAGE , WOMIT AUF DIE <FRAGW=URDIGKEIT UND EIGENT=UMLICHE <
ZWIENATUR DIESES GEHEIMNISVOLLEN <ELEMENTES IM <VORBEIGEHEN ANGESPIELT UND HIN
GEWIESEN SEI .
#@
#0300930 UM ABER EINEN KLAREN <SACHVERHALT NICHT K=UNSTLICH ZU VERDUNKELN : DI
```

5. テキスト・データの検索

ＳＩＧＭＡシステムにおけるテキスト・データの検索はＳＥＡＲＣＨコマンドを用いて行なう．ＳＥＡＲＣＨコマンドは，テキスト・データを何ら加工することなくそのままの形で前から後ろに一読する間にすべての検索処理を行う逐字サーチの方法を用いている．ＳＥＡＲＣＨは逐字サーチの方法を有効に活用し，複数の検索キーワードを用いた複数の質問を同時に処理できる特徴を持っている．これにより，通常では何度もファイルを読まなければならないような，動詞のすべての変化形の用例を求めるといった処理を極めて高速に行うことができるのである．

具体的なＳＥＡＲＣＨコマンドの使用例は６章にまとめて示すことにして，まずトーマス・マン・ファイルを検索するために必要な事項を説明していくことにする．実際のＳＥＡＲＣＨコマンドの機能の詳細は文献［２］に示されているので，ここでは読者の理解を妨げぬよう最低限必要な事項の説明に留めておく．

まずSEARCHコマンドの簡単な使用例を与えておき，以下順を追って説明していくことにする．

```
READY
SIGMA
SIGMA> TERM
TERM: P
 PREFIX:=F1683.MANN
TERM:
DO: SEARCH
 VERSION (D/E)? D

   RECORD DELIMITERS
 D1:=#
 D2:=

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

   LOGICAL FORMULAE
 Z1:=A1
 Z2:=

 FILE:=S.ERZ.TONIO

  RETRIEVED TEXTS

TOTAL              =      3
QUESTION  1 (Z1 ) =      3
CPU TIME =             288

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

 QUESTIONS:=1


                                                    QUESTION  1 (Z1 ) =      3
                                                                 NO.   1(  269)
0828624 NEIN , NEIN , SEIN <PLATZ WAR DENNOCH HIER , WO ER SICH IN <INGE'S <N-
=AHE WU$TE , WENN ER AUCH NUR EINSAM VON FERNE STAND UND VERSUCHTE , IN DEM <-
SUMMEN , <KLIRREN UND <LACHEN DORT DRINNEN IHRE <STIMME ZU UNTERSCHEIDEN , IN-
 WELCHER ES KLANG VON WARMEM <LEBEN .

                                                                 NO.   2(  177)
0828708 HATTE ETWA NICHT K=URZLICH EINE <ZEITSCHRIFT EIN <GEDICHT VON IHM ANG-
ENOMMEN , WENN SIE DANN AUCH WIEDER EINGEGANGEN WAR , BEVOR DAS <GEDICHT HATT-
E ERSCHEINEN K=ONNEN ?

                                                                 NO.   3(  343)
0830809 ZUWEILEN IN DIESEN DREIZEHN <JAHREN , WENN SEIN <MAGEN VERDORBEN GEWE-
SEN WAR , HATTE IHM GETR=AUMT , DA$ ER WIEDER DAHEIM SEI IN DEM ALTEN , HALLE-
NDEN <HAUS AN DER SCHR=AGEN <GASSE , DA$ AUCH SEIN <VATER WIEDER DA SEI UND I-
HN HART ANLASSE WEGEN SEINER ENTARTETEN <LEBENSF=UHRUNG , WAS ER JEDESMAL SEH-
R IN DER <ORDNUNG GEFUNDEN HATTE .

DO: END
SIGMA> END
READY
```

5.1 VERSIONの選択

SEARCHコマンドを投入すると，VERSIONをDにするかEにするかをたずねてくるのでDを入力するか単に復改すればよい．VERSION Eについては，ここでは説明しないので文献を参照されたい．またSEARCH Dと入力すればただちに次のレコード区切り語の登録へとうつる．

5.2　レコード区切り語（Record Delimiters）

ＳＥＡＲＣＨコマンドの検索の単位は，レコード区切り語によってはさまれた部分で，レコードと呼ばれる．レコード区切り語として任意の文字列を用いることができるが，トーマス・マン・ファイルでは，文章と文章の間に＃記号を挿入しているので，通常は＃をレコード区切り語に用いればよい．

ここで注意しなければならないことは，もしレコード区切り語を何も指定しない場合には，ファイル全体が１つのレコードとみなされてしまい，検索結果は０か１のいずれかになることである．

5.3　キーワード（Keywords）

レコード区切り語の指定を終えると，次にキーワードの登録をしなければならない．キーワードもまた任意の文字列である．従って，１つの単語'△ＮＡＣＨ△'や２つの単語からなる句'△ＮＡＣＨ△＜ＨＡＵＳＥ△'も正しいキーワードである．（△は空白記号を表わすものとする．）また'＋'をキーワードとして用いれば，分離動詞の分離した前つづりを含む文を検索することができ，'＋ＡＵＦ△'を用いれば分離した前つづりのうち'＋ＡＵＦ△'を含む文のみを求めることもできる．

ＳＥＡＲＣＨではキーワードを同時に複数登録することができ，入力された順にＡ１，Ａ２，…と名前をつける．ＳＥＡＲＣＨでは質問に直接キーワードを用いないで，この名前を任意に組み合わせて作られる論理式を用いて質問をする．

5.4　空白（ブランク）

空白も１つの文字であるから，キーワードの登録に際しては十分注意しなければならない．例えば'△ＵＭ△'は単語としてのumだけを検索できるが，'ＵＭ'はこれ以外にもzumなどの単語中に現われるumも検索してしまうのである．トーマス・マン・ファイルではすべての単語の前後に空白が置かれているので，キーワードにおける空白の用い方しだいで様々な検索が可能である．通常の単語の検索の場合には，原則として前後に空白を置いたキーワードを用いればよい．また，接頭辞un－の検索には'△ＵＮ'，接尾辞－ungの検索には'ＵＮＧ△'を用いればよい．

5.5　大文字シフト記号（＜）

トーマス・マン・ファイルでは，テキスト・データはすべて大文字であるため，本来大文字表記すべき名詞の先頭などは，大文字シフト記号として＜が置かれている．従って，名詞のWürdeと動詞のwürdeはそれぞれ'△＜Ｗ＝ＵＲＤＥ△'，'△Ｗ＝ＵＲＤＥ△'として区別できる．

5.6　'Ｘ．．．Ｙ'形式のキーワード

キーワード'Ｘ・・・Ｙ'はＸがまず先にあり続いてＹがある文字列に対応する．この形式のキーワードを用いれば，単に２つの単語が同時に現われているというだけでなく，その順序を検索の条件にすることができる．具体的な例をあげると，um…zuの構文をピックアップしたい場合などがこれに該当する．この場合キーワードとして'△ＵＭ△・・・△ＺＵ△'を用いるが，'・・・'にあたる部分は様々な場合をすべて含むので，望んでいない形式も検索されることがある．これを完全に除去する方法は直接目で見て判断するほかない．しかし，より細かい質問の指示を与えることによってあ

る程度は排除することができる．例えば，項目区切り語（ item delimiters ）と呼ばれる補助的な区切り語に＇，＇（コンマ）を用いることなどがある．項目区切り語については文献［２］を参照されたい．

5.7 論理式（Logical Formulae）

キーワードの登録が終わると質問のための論理式を入力しなければならない．論理式は，登録済みのキーワードの名前を組み合わせて作る．代表的な組み合わせの例が次に示してある．その意味は右にある通りである．

Ｚ１：＝Ａ１．Ａ２	Ａ１かつＡ２
Ｚ２：＝Ａ１，Ａ２	Ａ１またはＡ２
Ｚ３：＝へＡ１	Ａ１でない（へはASCⅡ端末の場合で，EBCDIC端末では￢を用いる．）

これらの論理式が検索の質問となるのである．SEARCHコマンドでは，このようにして複数のキーワードを組み合わせた複数の質問を指定するのである．これ以外の論理式の指定の仕方は文献［２］に説明してあるので参照されたい．

ここで注目してほしいことは，SEARCHの検索処理時間は質問の複雑さにはほとんど無関係であるということである．正確にいえば，処理時間は，検索するファイルの長さ（文字数）に比例する時間と，検索されるレコードの個数に比例する時間の和によって決まるのである．従って，SEARCHを有効に利用するためには，あらかじめ質問事項をよく検討し，必要であれば適当な短いファイルに対して検索を試行したりして，この同時処理機能を活用することが望まれる．なお，システムの制限値は，キーワードは９９個まで，質問のための論理式は３２個までである．

5.8 ファイル名の指定

論理式の設定で質問の準備がすべて終わると，最後に検索すべきファイル名を入力しなければならない．入力の形式は4.2で述べた通りである．指定されたファイルの検索が終了すると，見つかったレコード数，検索に要したCPU時間（単位はミリ秒）が表示され，さらに検索すべきファイルがあるかどうかをたずねてくる．ここでファイル名を入力せずに単に復改キーを押せば，検索結果の表示の処理にうつる．

5.9 検索結果の表示

SEARCHの検索結果は，MEMO領域にあるSEARCHのための作業用ファイル上に，レコードの位置情報と質問番号の関係のみを格納した形で保存される．検索結果の表示は，この作業用ファイルから実際のレコードを復元して行われる．検索の処理を終えると結果を表示するかどうかをたずねてくるので，表示する場合にはＹを入力する．Ｙを入力した場合には，どの質問に対する結果を表示するかをたずねてくるので，表示すべき質問番号を入力する．複数の質問番号の入力は，それらを空白で区切って行う．すると，それぞれの質問ごとに検索結果が端末に表示される．

5.10　検索結果の再ファイル化（ＲＥＦＩＬＥコマンド）

ＳＥＡＲＣＨコマンドの処理は，5.9で説明した検索結果の表示で終わるが，ＳＩＧＭＡシステムでは，ＳＥＡＲＣＨの作業用ファイルを用いて検索結果をMEMO領域のファイル上に書き出すことができる．この検索結果の再ファイル化のためのコマンドがＲＥＦＩＬＥである．ＲＥＦＩＬＥコマンドの簡単な使用例が次章に示されている．詳細については文献［２］を参照されたい．

6.　トーマス・マン・ファイルを用いた用例検索の具体例

ここにあげた例は，実際の用例検索に際して，より目的に適した質問を作成するための参考のためのものである．言語研究にとって，ＳＥＡＲＣＨシステムは，大量のテキスト・データから目的の文章を取り出すための補助的手段にすぎず，あくまでも検索された用例が質問の意図にあっているかどうかは，研究者自身の手によって吟味されなければならないことは言うまでもない．

１）話法の助動詞 können の用例を調べる．その変化しうる形をリストアップし，それらをキーワードに登録して検索する．この例ではすべてのキーワードの前後に空白を置かねばならないことに注意してほしい．

```
DO: TERM P F1683.MANN
TERM:
DO: SEA D

   RECORD DELIMITERS
 D1:=#
 D2:=

   KEYWORDS
 A1:= K=ONNEN
 A2:= KANN
 A3:= KANNST
 A4:= KONNTE
 A5:= KONNTEST
 A6:= KONNTET
 A7:= KONNTEN
 A8:= K=ONNE
 A9:= K=ONNEST
A10:= K=ONNET
A11:= K=ONNTE
A12:= K=ONNTEST
A13:= K=ONNTEN
A14:= GEKONNT
A15:=

   LOGICAL FORMULAE
 Z1:=A1
 Z2:=A2
 Z3:=A3
 Z4:=A4
 Z5:=A5
 Z6:=A6
 Z7:=A7
 Z8:=A8
 Z9:=A9
Z10:=A10
Z11:=A11
Z12:=A12
Z13:=A13
Z14:=A14
Z15:=
```

```
 FILE:=S.BB.N01

  RETRIEVED TEXTS

TOTAL                =     50
QUESTION  1 (Z1 ) =      9
QUESTION  2 (Z2 ) =     19
QUESTION  3 (Z3 ) =      1
QUESTION  4 (Z4 ) =     15
QUESTION  5 (Z5 ) =      0
QUESTION  6 (Z6 ) =      0
QUESTION  7 (Z7 ) =      0
QUESTION  8 (Z8 ) =      1
QUESTION  9 (Z9 ) =      0
QUESTION 10 (Z10) =      0
QUESTION 11 (Z11) =      5
QUESTION 12 (Z12) =      0
QUESTION 13 (Z13) =      0
QUESTION 14 (Z14) =      0
CPU TIME =             533

 FILE:=S.BB.N02

  RETRIEVED TEXTS

TOTAL                =     75
QUESTION  1 (Z1 ) =     19
QUESTION  2 (Z2 ) =     32
QUESTION  3 (Z3 ) =      1
QUESTION  4 (Z4 ) =     19
QUESTION  5 (Z5 ) =      0
QUESTION  6 (Z6 ) =      0
QUESTION  7 (Z7 ) =      0
QUESTION  8 (Z8 ) =      3
QUESTION  9 (Z9 ) =      0
QUESTION 10 (Z10) =      0
QUESTION 11 (Z11) =      5
QUESTION 12 (Z12) =      1
QUESTION 13 (Z13) =      0
QUESTION 14 (Z14) =      0
CPU TIME =             599
```

2) 話法の不変化詞，例えばmal やnur など任意の語について用例を見る．不変化詞は語形の変化がないので，いくつもの語をまとめて検索する方がよい．実際の検索例は1) と同様なので省く．

3) 接頭辞，接尾辞の調査，造語法上のun - ，ur - などの接頭辞や，- ung ，- heit ，- keit などの接尾辞の用例検索をする．この場合5.3 で注意したようにキーワードの前後における空白の用い方が問題となる．

```
DO: SEA D

  RECORD DELIMITERS
 D1:=#
 D2:=

  KEYWORDS
 A1:= UN
 A2:= UR
 A3:=

  LOGICAL FORMULAE
 Z1:=A1
 Z2:=A2
 Z3:=

 FILE:=S.ERZ.VENEDIG

  RETRIEVED TEXTS

TOTAL             =   615
QUESTION  1 (Z1 ) =   615
QUESTION  2 (Z2 ) =     1
CPU TIME =            389

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

 QUESTIONS:=1


                                                    QUESTION  1 (Z1 ) =   615
                                                              NO.   1(  364)
0844406 <GUSTAV <ASCHENBACH ODER VON <ASCHENBACH , WIE SEIT SEINEM F=UNFZIGST-
EN <GEBURTSTAG AMTLICH SEIN <NAME LAUTETE , HATTE AN EINEM <FR=UHLINGSNACHMIT-
TAG DES <JAHRES 19.. , DAS UNSEREM <KONTINENT MONATELANG EINE SO GEFAHRDROHEN-
DE <MIENE ZEIGTE , VON SEINER <WOHNUNG IN DER <PRINZREGENTENSTRA$E ZU <M=UNCH-
EN AUS ALLEIN EINEN WEITEREN <SPAZIERGANG UNTERNOMMEN .

                                                              NO.   2(  597)
0844411 =UBERREIZT VON DER SCHWIERIGEN UND GEF=AHRLICHEN , EBEN JETZT EINE H=-
OCHSTE <BEHUTSAMKEIT , <UMSICHT , <EINDRINGLICHKEIT UND <GENAUIGKEIT DES <WIL-
LENS ERFORDERNDEN <ARBEIT DER <VORMITTAGSSTUNDEN , HATTE DER <SCHRIFTSTELLER -
DEM <FORTSCHWINGEN DES PRODUZIERENDEN <TRIEBWERKES IN SEINEM <INNERN , JENEM -
" MOTUS ANIMI CONTINUUS " , WORIN NACH <CICERO DAS <WESEN DER <BEREDSAMKEIT B-
ESTEHT , AUCH NOAH DER <MITTAGSMAHLZEIT NICHT <EINHALT ZU TUN VERMOCHT UND DE-
N ENTLASTENDEN <SCHLUMMER NICHT GEFUNDEN , DER IHM , BEI ZUNEHMENDER <ABNUTZB-
ARKEIT SEINER <KR=AFTE , EINMAL UNTERTAGS SO N=OTIG WAR .

                                                              NO.   3(  183)
0844421 SO HATTE ER BALD NACH DEM <TEE DAS <FREIE GESUCHT , IN DER <HOFFNUNG -
, DA$ <LUFT UND <BEWEGUNG IHN WIEDERHERSTELLEN UND IHM ZU EINEM ERSPRIE$LICHE-
N <ABEND VERHELFEN W=URDEN .
```

この例では，un およびur で始まる単語を含む文を求めている．キーワードには'△UN '，'△UR 'を用いた．検索された文章は，un で始まる単語を含むものが615 ，ur で始まる単語を含むもの

が１であった．結果を表示してみると，un で始まる単語として und が非常に多く現われていることがわかる．また，見つかった３番目の文章には und 以外に un で始まる単語は現われていない．それでは un で始まる und 以外の単語を含む文を検索するためにはどうすればよいであろうか，すぐに思いつくことは，キーワードを

Ａ１：＝△UN

Ａ２：＝△UND△

として，論理式を

Ｚ１：＝Ａ１．∧Ａ２

とすることである．しかしこれでは，und を含む文がすべて除外されることになるのでうまく行かない．正しくは，キーワードはこのままにして論理式を

Ｚ１：＝Ａ１〉Ａ２

とすればよいのである．これは５章では説明しなかったキーワードの組み合わせ方であるが，'△UN'が現われている回数が'△UND△'が現われている回数より多いものを求めるものである．次の例でＡ１．∧Ａ２とＡ１〉Ａ２を同時に質問して結果を比較して見た．

```
DO: SEA D

   RECORD DELIMITERS
 D1:=#
 D2:=

   KEYWORDS
 A1:= UN
 A2:= UND
 A3:=

   LOGICAL FORMULAE
 Z1:=A1.^A2
 Z2:=A1>A2
 Z3:=

 FILE:=S.ERZ.VENEDIG

  RETRIEVED TEXTS

TOTAL              =   174
QUESTION  1 (Z1 ) =    34
QUESTION  2 (Z2 ) =   174
CPU TIME =            372

 FILE:=S.ERZ.TONIO

  RETRIEVED TEXTS

TOTAL              =   137
QUESTION  1 (Z1 ) =    17
QUESTION  2 (Z2 ) =   137
CPU TIME =            327

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

４）ＲＥＦＩＬＥの使用例．ＳＥＡＲＣＨコマンドを用いて検索されたレコードを再ファイル化する．ＲＥＦＩＬＥによって検索結果を再ファイル化することによって，検索結果をさらにＳＥＡＲＣＨコマンドで検索したり，様々に利用することができる．次はその簡単な使用例である．

```
DO: SEA D

  RECORD DELIMITERS
 D1:=#
 D2:=

  KEYWORDS
 A1:= WENN ... AUCH
 A2:= WIE ... AUCH
 A3:=

  LOGICAL FORMULAE
 Z1:=A1
 Z2:=A2
 Z3:=

 FILE:=S.ERZ.TONIO

  RETRIEVED TEXTS

TOTAL              =      6
QUESTION  1 (Z1 ) =      3
QUESTION  2 (Z2 ) =      3
CPU TIME =             309

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

 QUESTIONS:=1
 TOTAL RECORDS =    3
 RECORD DELIMITER:=#
 NUMBERING (N/Y)?
 SORT ON:
DO: LOOK
#0828624 NEIN , NEIN , SEIN <PLATZ WAR DENNOCH HIER , WO ER SICH IN <INGE'S <-
N=AHE WU$TE , WENN ER AUCH NUR EINSAM VON FERNE STAND UND VERSUCHTE , IN DEM -
<SUMMEN , <KLIRREN UND <LACHEN DORT DRINNEN IHRE <STIMME ZU UNTERSCHEIDEN , I-
N WELCHER ES KLANG VON WARMEM <LEBEN .
#0828708 HATTE ETWA NICHT K=URZLICH EINE <ZEITSCHRIFT EIN <GEDICHT VON IHM AN-
GENOMMEN , WENN SIE DANN AUCH WIEDER EINGEGANGEN WAR , BEVOR DAS <GEDICHT HAT-
TE ERSCHEINEN K=ONNEN ?
#0830809 ZUWEILEN IN DIESEN DREIZEHN <JAHREN , WENN SEIN <MAGEN VERDORBEN GEW-
ESEN WAR , HATTE IHM GETR=AUMT , DA$ ER WIEDER DAHEIM SEI IN DEM ALTEN , HALL-
ENDEN <HAUS AN DER SCHR=AGEN <GASSE , DA$ AUCH SEIN <VATER WIEDER DA SEI UND -
IHN HART ANLASSE WEGEN SEINER ENTARTETEN <LEBENSF=UHRUNG , WAS ER JEDESMAL SE-
HR IN DER <ORDNUNG GEFUNDEN HATTE .

DO: PUT
 TO-FILE:=WENN.AUCH
 PASS NUMBER:=[illegible]
DO:
```

この例では，△WENN△．．．△AUCH△を含むものを求めて，検索結果を再ファイル化している．再ファイル化の結果はWORKに置かれるので表示のためにLOOKコマンドを用いている．最後にPUTコマンドを用いてこのファイルにWENN. AUCHという名前をつけて保存している．

＊使用上の注意

トーマス・マン・ファイルは，西独Ｓ．Ｆischer社の許可を得て，専ら言語研究のために提供するものである．このテキスト・データを何らかの方法により加工して，読書用のテキストとして使用することは禁止する．この点には注意を喚起しておきたい．

また機械可読なデータとしてのテキスト・データ自体の“著作権”は著者の側にあるので，みだりにコピーを作るなどの方法はとらないようにしていただきたい．

謝　辞

今回の公開にあたって多くのご支援・ご助言をいただいた松尾研究開発部長ならびに二村データベース室長に深く感謝の意を表します．

参考文献

1. 松尾，二村，高木　公用データベースについて，九州大学大型計算機センター広報，**15**，2，1982，222－227．
2. 有川，篠原，白石，玉越　研究者向き情報システムＳＩＧＭＡについて，同上，**14**，4，1981，550－573．
3. S. Arikawa, T. Shinohara, S. Shiraishi, Y. Tamakoshi SIGMA－An Information System for Researchers Use, Bulletin of Informatics & Cybernetics, **20**, No. 1～2, 1982, 97－114.