ゲーテ・ファイルの完成について

樋口, 忠治
九州大学言語文化部

https://doi.org/10.15017/1468192

出版情報：九州大学大型計算機センター広報. 22 (1), pp.1-5, 1989-01-25. 九州大学大型計算機センター
バージョン：
権利関係：

ゲーテ·ファイルの完成について

樋口　忠治*

1. まえがき

Goethes Werke (Hamburger Ausgabe in 14 Bänden) のうち6巻から14巻までのテキストについては1988年度から公開提供を開始していたが†, このたび残りの1巻から5巻までのテキストのデータ化も終了したので, これらのファイルの追加を行った. 本稿では, これらの追加ファイルの内容に関して説明する.

2. レコード形式について

今回追加されるテキストの内容は別掲のリストに示す通りであるが, 若干の説明をつけ加える. 全体を詩の形式(韻文)と散文とに大別して, 詩形式の場合は行を以てレコードの単位とした. 散文の場合は従来と同様に「文」を以てレコードの単位としている. 詩行をレコードとしたことにより, 検索結果はキーワードを含む行のみが出力されるから, 「文」を単位とした場合に較べてかなり出力量が減少する. 換言すれば, ヒットしたレコードの件数と等しい行数だけが出力されることになる.

詩行には全体に通し番号をつけたもの(Faust など)もあるし, 各章ごとの通し番号をつけたものもあり, さまざまであるが, Faust (第3巻)のみはページ番号を採用せず, すべて通し番号のみを採用している. その他の詩行はすべてページ番号との併用とした. この場合, 行数を表す数字は2桁の範囲内で利用するため, 3桁以上の数字は末尾の2桁のみを示していることに注意されたい. 同一のページの中に複数の詩が含まれている場合には, 行数としても全く同一の数字が割り当てられることが起こる. こうした問題は書物の行数字を利用する場合には回避できない. しかし, 検索に関しては何ら問題はなくて, 書物のテキストと照合するのに何の障害もないのである.

3巻の中のUrfaustの場合は4桁の通し番号をもつ詩行の部分と散文とが混在しているが, 通し番号の部分はその行数を採用している. 但し, 4桁の数字の前に'U'の文字を1桁加えて, Urfaustの行数であることを示した. 散文の部分については他と同様にページ数と行数の組み合せとなっている.

4巻と5巻のテキストは詩行の形式をとったものと散文形式の違いはあるが, 総じてドラマである. そこで, これらを分割して, 詩行のものだけを1ファイルとし, 散文のものだけを集めて1ファイルとした. 従って検索に際しては, 作品の内容が詩行形式の場合はGA04, 散文形式の場合にはGA05のファイルに含まれていることに注意されたい. もっとも, 検索は作品単位で行うことを前提としているわけではないし, 出力データの不揃いを避けるという長所もあるわけであるから, この分離方式はやむを得ないものと考える.

以上, 1巻から5巻までのテキストのレコード形式に2種類があることを述べたが, 一般的に, レコードやファイルの編成を書物としてのテキストの配列にどのように対応させるべきかについては, 参考になる例は皆無であるから現時点で最良と考えられる方法によったが, 将来, より良い編成方法があれば, そうした方法に改めることも考慮したい. 原則論としては, 書物としての配列方法をそのまま完全に踏襲するか, 検索の方法全体にとって効果的な配列をとるかの選択の問題であるともいえるが, 書物としてのテキストの形式や行数その他の表現形式の利用の仕方については, 今後なお検討していく余地があろう.

昭和63年11月22日受理

*九州大学言語文化部

†文献1を参照.

```
READY
SIGMA
SIGMA)DDIR
FILE:=S.'A70152C.*'
PASSNUMBER:=↙
OPTION:=A
 FILENAME                                    ID ALIAS      SIZE     DATE  TIME
GA01                                       833B    0     709509 88:10:13 18:56
     BD. 1 (  7-391) BD. 2 (  7-125)
GA02                                       992P    0     813742 88:12:21 14:02
     BD. 2 (126-270;285-536)
GA03                                       798Y    0     719694 88:10:13 18:57
     BD. 3 (  9-364;367-420)
GA04                                       942L    0     757659 88:10:13 18:57
     BD. 4 (  7- 72;176-202;455-462) BD. 5 (  7-167;215-308;332-454)
GA05                                       698N    0     813468 88:08:25 12:00
     BD. 4 ( 74-175;203-454) BD. 5 (168-214;309-331)
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                                       221O    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                                       857T    0    1266739 88:09:06 13:42
     BD.11 (  9-556)
GB12                                        28A    0    1122320 88:11:21 14:32
     BD.12 (  9-364;365-547)
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)
NUM                                        5304    0        358 88:04:04 13:43
SMALL                                      531A    0        919 88:04:04 13:43

        TOTAL =  15          PREFIX = A70152C
 1137 SECTOR(S) AVAILABLE
DO:LIST S.'A70152C.GA01'
#0100700 @@@            @@
#0100701  ERHABNER ⟨GRO$PAPA !   EIN ⟨NEUES ⟨JAHR ERSCHEINT ,
#0100702 DRUM MU$ ICH MEINE ⟨PFLICHT UND ⟨SCHULDIGKEIT ENTRICHTEN ,
#0100703 DIE ⟨EHRFURCHT HEI$T MICH HIER AUS REINEM ⟨HERZEN DICHTEN ,
#0100704 SO SCHLECHT ES ABER IST , SO GUT IST ES GEMEINT .
#0100705  ⟨GOTT , DER DIE ⟨ZEIT ERNEUT , ERNEURE AUCH ⟨IHR ⟨GL=UCK ,
#0100706 UND KR=ONE ⟨SIE DIES ⟨JAHR MIT STETEM ⟨WOHLERGEHEN ;
.....
```

ファイルの一覧表示とファイルの形式

GOETHES WERKE

Hamburger Ausgabe 14 Bänden

Band 1	7−391	Gedichte
Band 2	7−270	West−östlicher Divan
	271−284	Die Geheimnisse
	285−436	Reineke Fuchs
	437−514	Hermann und Dorothea
	515−536	Achilleis
Band 3	9−364	Faust
	376−420	Urfaust
Band 4	7− 27	Die Laune des Verliebten
	28− 72	Die Mitschuldigen
	73−175	Gotz von Berlichingen
	176−187	Prometheus
	188−202	Satyros
	203−215	Gotter, Ilelden und Wieland
	216−259	Claudine von Villa Bella
	260−306	Clavigo
	307−351	Stella
	352−369	Die Geschwister
	370−454	Egmont
	455−462	Proserpina
Band 5	7− 67	Iphigenie auf Tauris
	68− 72	Nausikaa
	73−167	Torquato Tasso
	168−214	Die Aufgeregten
	215−299	Die Naturliche Tochter
	300−308	Palaophron und Neoterpe
	309−331	Elpenor
	332−365	Pandora
	366−399	Des Epimenides Erwachen

3. ファイル名について

ファイル名の形式は，全体としては全集の巻数に対応する番号形式にしているわけであるが，ファイル名の先頭文字はすべてGAとし01から05までの数字を付している．6巻から12巻まではGB，13巻と14巻はGCを先頭にもつようにファイル名をつけているから，

1) 1巻から5巻までを一括して検索したい場合には，S.GA*を指定する．
2) 6巻から12巻までを一括して検索したい場合には，S.GB*を指定する．
3) 13巻と14巻を一括して検索したい場合には，S.GC*を指定する．
4) 1巻から14巻までを一括して検索したい場合には，S.G*を指定する．

もちろん，各ファイル毎にファイル名を指定することができるが，その場合には，例えばS.GA01のように指定することになる．上記1)から4)までのファイル名一括投入の方法は，出力データが少ないことが予想される場合には簡単で便利であるが，出力データが過大な場合はその取り扱いが困難になることがあるので注意を要する．なお，SIGMA状態においてPROFILE及びPREFIXコマンドの使用によってプレフィックスの指定をしていなくても，ファイルの完全名を指定すれば，ファイルへのアクセスは可能である．完全名とは，ファイル名の前にプレフィックスA70152Cをつけ，全体を引用符号でくくった形式であり，例えばS.GA01の場合はS.'A70152C.GA01'が完全名を指定した形式となる．しかしながら，なるべくは事前にプレフィックスの指定をしておくことが望ましい．

4. ゲーテ・ファイル(ハンブルク版)の完成について

今回のファイル追加によって，ハンブルク版ゲーテ全集14巻のすべてのテキストが，テキスト・データベースとして利用できることになった．これによって，全14巻の全集の中から指定した語や句を含む文や詩行を約6秒(CPU)ですべて調べることができる．ゲーテのある言葉を憶えていても，それがどこに出て来たのか思い出せない，という経験をした人は多いであろう．このような場合，ゲーテ・ファイルの検索によって直ちにその出現個所を知ることが可能である．テキスト・データベースはすべての語(単語)を単独で，または組み合わせて検索することができるから，あいまいな形で記憶しているにすぎない場合も，方法によっては何とか捜し出すことが可能である．言語研究のためばかりでなく，文学研究(特にゲーテ研究)のためにも，このテキスト・データベース・システムは大いに役立つであろう．

ハンブルク版全集14巻はワイマル版と較べると確かに規模が小さい．しかし，ワイマル版のほぼ半分は日記と書簡からなっており，その部分を除いて考えると，主要な著作物のほとんどはハンブルク版に収められているといえる．また，ワイマル版においては語の綴りが現在のように統一されていなかったが，ハンブルク版はほぼ現代ふうに統一されている．このような事情があるから，ワイマル版に基づいたテキスト・データを作成した場合には，これを検索するに当たっては，語の綴りの特性を十分知っているのでなければ，正しいキーワードの指定をしたことにならない．従って本来の正しい結果も得ることはできない．つまり，ハンブルク版は一般性を持っているのに対して，ワイマル版は専門的な予備知識を必要とするのであるから，利用者が限定される．しかしながら，本来の「ゲーテ」のテキスト・データベースとしては，著作物の全てを含むワイマル版全集が出来れば望ましいに違いない．その場合にも，綴りの問題をどのように扱うかが重要なポイントになるであろう．このような場合には，当時の綴りをそのまま採用した版と，現代の綴りに改めて統一し

た版との2種類をつくることも考えられるが，テキスト･データベースの本質が検索であるとすると，書物のままの綴りに固執するよりも，検索に適した綴りを選ぶほうがよいのかも知れない．いずれにせよ，これは今後の問題として検討を要するであろう．

5. あとがき

ハンブルク版ゲーテ全集が完全な形でテキスト･データベースとして提供できることになったが，この際，従来のパスナンバーを除去して，すべてパスナンバー無しで検索できるように改める．パスナンバーの問い合わせに対しては，そのまま，リターンキーを押すか，0を入力すればよい．この方式を採用することにより，ファイルの保護が難しくなる欠点はあるが，他方，ファイル･アクセスのたび毎にパスナンバーを入力するという繁雑さからは解放される長所があり，結局，後者を選択することにした．検索の目的でのみ利用されることを特に強く要望しておきたい．

参考文献

1. 樋口　テキスト･データベース「ゲーテファイル」の公開，九州大学大型計算機センター広報，Vol. 21, No. 3,4, 1988, pp. 167-176.
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

「ゲーテ･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