

歴史地震

第 5 号

平成 1 年

歴史地震研究会

歴史地震史料集のデータベース化について—パソコンで読める史料集—

歴史地震史料データベース化委員会

岩崎伸一（国立防災科学技術センター・平塚海洋防災研究支所）

吉井敏尅・都司嘉宜（東京大学地震研究所）

石橋克彦（建設省建築研究所）

笠原敬司・小見波正隆（国立防災科学技術センター）

歴史地震史料の一部がパソコン（PC98シリーズ）のディスプレイ上で表示、検索できるようになった、印刷も可能である。システムは現在のところ2枚のフロッピーディスクに収まっている。その内容は1. コード化されたテキストファイル。2. キーワードインデックス。3. 日付インデックス。4. これら进行处理するプログラム。の4種類に分けられる。この4つについて順次説明する。

1. コード化されたテキストファイル。TEXT128. M, S, R.

M, S, Rは、それぞれ増訂大日本地震史料、新収日本地震史料、理科年表を表す識別子である。使用したコードはシフトJISであり、JIS第2水準までの漢字を用いることとし、外字は設けなかった。コード化した史料の範囲は以下のものである。

・TEXT128. M

増訂大日本地震史料第1巻

頁：605—723

年月日：西暦1596/9/5—1624/6/19

慶長元年閏7月13日—寛永元年5月4日

記載地震数：270個

・TEXT128. S

新収日本地震史料第2巻本巻

頁：58—111

年月日：西暦1596/9/5—1623/3/27

慶長元年閏7月13日—元和9年2月27日

記載地震数：191個

・TEXT128.R

理科年表(1989)日本付近の主な被害地震年代表

頁：地198(816)—地228(846)

年月日：西暦416/8/23—1987/12/17

尤恭5年7月14日—昭和62年12月17日

記載地震数：432個(4つの外国の地震を含む)

基本的には、本文の内容をそのまま1行1レコードとして入力した。

変更を加えたものは、

(1) JIS第2水準に含まれない文字、あるいは、読めない文字。

基本的には井を代わりに入力。新収史料中に表れる白抜き下向き矢印は、下黒三角で代用した。

(2) 2行割書き。

1行の途中にあるものは、1行に直し2重括弧でくくった。1行全部が割書きになっている場合は、2重括弧でくくるのみとした。

(3) ふりがな、返り点、傍注など

括弧でくくって直後に挿入

等であるが、その他細かいところで変更を加えたものがある。

読み出しの手がかりとして、以下のものを最終バイトから最大12バイトまでに埋め込んだ。

(1) 頁インデックス。8バイト。

頁の最初に、史料の種類と頁を判断できるようにつけた。たとえば、今回はS200086P、あるいはM100662P等である。S、Mについては前述した。20は第2巻の本巻、0086は86頁、Pはこの行が頁インデックスであることを示す識別子である。

(2) 地震インデックス。12バイト。

例を挙げて説明する。1596090501AZというインデックスは、

この1行が地震の記事であり(Z)、西暦の1596年9月5日に起こったもので(15960905)、この頁の最初の地震の記事で(01)、ひと続きの地震の記事の1行目に当たる(A)。

(3) 史料名インデックス。4バイト。

これも例を挙げて説明する。04BDというインデックスは、この1行が史料名を表すものであり(D)、この頁に出てくる4番目の史料名の行であり(04)、ひと続きの史料名の記事の第2行目に当たる(B)。

(4) 年月日インデックス。8バイト。

史料名、史料本文、後述する史料区切りの全てに西暦年月日を埋め込んだ。

(5) 史料区切り。3バイト。

新収史料中、一つの史料名の中に複数の史料を含むものがあり、それらを縦波線で区別している。これらのものは、レコードの先頭に史料区切りと入力し、最終3バイトに01V等のインデックスをつけた。この意味は、この行は史料の区切りであり(V)、同一史料名中の第一番目の区切りである(01)。

2. キーワードインデックス。

検索処理の迅速化のため、キーワードインデックスを作った。一般のキーワードインデックスは、ある任意の最小単位の文章にいくつかのキーワードを予めつけておき、文書を検索する形式である。しかしながら今回のデータベース化に当たっては、史料の性質上、キーワードを後で幾らでも増やせるようにしたいこと。また、文書の最小単位をどの様に決めるのが合理的であるかがはっきりしなかった。さらにそれに、はじめからキーワードをつけるのは手間から考えて不可能であること。等の理由から、以下の方式にした。まず最小単位を1行とした。これには地震、史料名、史料本文など全てのもが含まれる。次にTEXTのレコード番号を文番号とした。さらに、キーワードの方に、このキーワードを含む行の文番号をぶら下げることとした。この様にすることにより、キーワードは幾らでも増やせる。また、新しい史料がつけ加わったときも、キーワードに文番号をぶら下げただけで対応が出来る。

3. 日付インデックス

地震の起こった西暦の日付と、文番号の対応表である。このインデックスを活用することにより、文番号を与えての史料の表示、西暦の日付を与えての史料の表示が迅速に行える。

4. 処理プログラム。

以上のインデックスを活用し、表示、検索、印刷を統合して行うものである。簡単にその概要を述べる。

表示

- ・文の番号を指定する。頭に#をつける。例 #1200
- ・地震の年月日を指定する。例 1601/2/3

以上のいずれか一つによって指定されたレコードを先頭にして史料の内容がディスプレイに表示される。

この表示は ROLLUP ROLLDOWN によりスクロールすることができる。f1で印刷が可能。地震の年月日が間違っていた場合、その年に地震があれば、その年の地震のあった日の日付が表示される。また、年だけを入力しても同じ結果となる。

検索

このモードでは、自動的に漢字変換機能が使えるようになる。適当な文字列を与えることにより検索ができる。

検索の表示の中で詳しく見たいところの内容を表示することができる。

表示モードにはいることによって印刷機能が使える。

初めての文字列での検索の時は、ファイルをまじめに捜しに行く。このプログラムには学習機能があるため、2度目以降からの検索表示では、上述したキーワードインデックスを参照して非常に高速に表示する。

現在の進展状況など

昭和63年度にコード化した史料は主な史料集の100分の1の量である。今年度もほぼ同じ量のコード化を現在進めている。このままでは単純計算で100年かかることになる。一方、我々だけでなくいろいろな所で、いろいろな立場の方々がデータベース化に着手していると聞く。お互いの仕事を効率よく進ませるためにも、連絡を密にとってゆきたい。

プログラムは、現在のままでも十分使えるが、若干の改良を進めている。

このプログラムは最終的な決定版ではありませんが、使ってみてほしい方は、ぜひ下記の所まで連絡して下さい。送らせていただきます。

連絡先・国立防災科学技術センター・平塚海洋防災研究支所・岩崎伸一

〒254 平塚市虹が浜9-2

t e l 0463-32-7159

f a x 0463-33-3635