

平和研究文献用簡易検索システム PRINSE の 改良について

松 尾 雅 嗣

平和科学研究センター

Improvement of PRINSE, A Simple Retrieval System for Peace Research Documents

Masatsugu MATSUO

Institute for Peace Science, Hiroshima University

Summary

The present paper is a sequel to the report in the preceding issue of this journal, and discusses improvements and extensions of PRINSE, a microcomputer program for the retrieval of peace research documents. Main improvements discussed here are as follows.

Improvements of the main program PRINSE with the change of the operating system from N88 BASIC to MS-DOS. Development of a program to transform list of documents into PRINSE files, that is, files which can be read by PRINSE. (Input utility) Development of programs to output the list of documents arranged according to the alphabetical order of authors and keywords either to the printer or to the disk. (Author index and keyword index utilities)

はじめに

本報告は、本誌前号に報告した平和研究文献用簡易検索システム PRINSE (文献1)の機能の拡張,改良についての報告である。改良,拡張された主な点は以下のとおりである。

N88 BASIC から MS - DOS への OS の変更

検索プログラムの機能拡張と高速化

PRINSE への入力プログラムの開発

著者名のアルファベット順による全文献リスト(著者索引)出力プログラムの開発

キーワードのアルファベット順による全文献リスト(キーワード索引)出力プログラムの開発

このうち、PRINSEの主プログラムである検索用プログラムに関しては、OSの変更も含め、既に竹花誠児氏と共同で報告済みであるので(文献2)、ここでは割愛し、新たに開発したPRINSEへの入力プログラム、著者索引、キーワード索引作成のプログラムについて報告する。なお、本報告では、OSをMS-DOSに変更したことにより、MS-DOS関連の用語を断りなしに使用することがあるので、MS-DOSのマニュアル、入門書(例えば、文献3,4)の類を、必要に応じ、参照されたい。

1 ユーティリティプログラムの実行開始と終了

本稿で報告する3つのプログラム(以下ユーティリティプログラムと総称する)は、すべて“UTILITY”という名前の付いたディスクで提供される。

ユーティリティプログラムの実行環境,終了方法等は、主プログラムに準ずるので、文献2に譲り、ここではユーティリティプログラムの実行開始方法についてのみ述べる。UTILITYディスクをドライブAにセットするとディスクに収められたプログラム(コマンド)の名称と簡単な説明が表示される。コマンド(プログラ

ム) とその機能は次のとおりである。

IN: PRINSE へのデータ入力

AI: 著者索引の作成

KI: キーワード索引の作成

この直後に MS-DOS のプロンプトが表示されるので、実行したいコマンド名を入力すればよい。個々のコマンドの使用方法については、次節以下に述べる。

2 PRINSE への入力 (IN コマンド)

2.1 入力データ

PRINSE による文献の検索は、文献データを PRINSE で利用できる形に変換して始めて可能となる。平和科学研究センターあるいは利用者個人が作成した文献データを PRINSE 用の形式に変換することを、ここでは、PRINSE への入力と呼び、変換される前の文献データを入力データと呼ぶことにする。入力用プログラムの説明に先立って、まず入力データについてごく簡単に触れておく。

PRINSE への入力データは、次の条件を満たすものでなくてはならない。

MS-DOS のテキストファイルであること。

1 件の文献は、3 行 (3 レコード) から成り、先頭から順に著者名、表題、書誌事項を示す行であること。

1 行の長さは最大 254 文字以下であること。

2 バイトコード (漢字等) を含まないこと。

PRINSE は、入力データの作成そのものは、特にサポートしていないので、入力データ自体は、英文ワープロ、スクリーンエディタ等を利用して作成する。その際、ワープロソフトでは文字以外の制御コードが混入することがあるので、MS-DOS のテキストファイル作成機能を有するソフトを利用する方が安全である。(後述のように、PRINSE への入力時に、入力ユーティリティ、IN コマンドの編集機能を利用してこのような制御コードを削除できる場合もあるが、保証の限りで

はない。)

1 件の文献に関するデータを構成する 3 行 (3 レコード) は、上述のように先頭から順に、著者名、表題、書誌事項を示すが、以下、これらをそれぞれ、著者名レコード、表題レコード、書誌事項レコードと呼ぶことにする。各レコードについては若干の説明が必要である。

まず、著者名レコードには、必ず編著者名が与えられていなければならない。編著者名等が不明の時には必ず斜線 (/) を与えなくてはならない。著者名は、次の規則に従って与える。

編著者名は、姓、名の順で与える。(例 1 参照)

この時、姓、名の区切りとして、カンマ (,) を用いてはならない。(例 1 参照)

複数の編著者があるときには、カンマと空白 (,) で区切る。(例 1 参照)

(例 1)

Woodward M, Henderson L, Jackson E

(ファーストネーム等は、必ずしも、この例のようにイニシャルである必要はない)

著者名レコードに、編著の別、版、翻訳、出版年等の、編著者名以外の情報も与えておきたいときには、次の例 2 のように編著者名の後に () で括って与える。(後述の書誌事項レコードに与えることもできる。) 著者名レコードに左括弧 (があるとき、それ以降のデータは著者名とは看做されず、検索のためのキーとして登録されない。

(例 2)

Woodward M, Henderson L, Jackson E (eds.) (tr)

表題レコードには文献の表題を与える。このレコードには、表題、副題は言うまでもなく、著者名以外の、検索用キーワードとしておきたいすべての語、句、記号等を与えてよい。このとき、表題と区別するため適当な区切り記号を使用してもよ

い。このレコードに与えられたデータは、入力時に、空白、カンマ(,)、ピリオド(.), コロン(:), セミコロン(;), 疑問符(?), 感嘆符(!), パーセント記号(%), 斜線(/), 二重引用符("), 左右括弧((と)), 左右角括弧([と]), 左右中括弧(|と|), 下線(), 等号(=), 縦線(|), 波線(~)の20種の記号と、行頭、行末を区切りの標識として単語に分割される。そして、表題レコード中の単語は、利用者の指示がない限り、すべてキーワードとして登録される。但し、数詞(正確には、数字で始まる単語)は、たとえ表題レコード中であっても、キーワードとしては扱われない。この点に関しては、ストップワードの指定、例外の処理等を含め、詳しくは、後の2.3, 2.5に述べる。また、上述の区切り記号によって分離される文字列はすべて単語として扱われるから、“United States”, “United Nations”等は、それぞれ2語として処理される。(これは、主プログラムで検索するときに、“United States”というキーでは直接には検索できないことを意味する。)このような複合語の扱いに関しては、2.3で述べる。

書誌事項レコードは、本来、検索キーとしない書誌事項を与えることを意図したものであるが、実際にはここに何を与えても構わない。注意すべき点はふたつ、与えるべきデータが無いときでも、省略はできないので、空白以外のデータ、たとえば斜線、を、少なくともひとつは与える必要があることと、このレコード中の如何なる文字、記号、単語も検索キーとすることはできないということである。

以上、入力データの構造を簡単に説明した。例3に、実際の入力データの一部を例として挙げておく。

(例3)

Engelbrecht Helmuth Carol, Hanighen Frank Cleary (1937)
Merchants of Death
75-291
Read Herbert (1949)
Education for Peace
75-292
Brock Peter (1968)
Pacifism in the United*States, from the Colonial Era to the First World War
75-293
Kellogg Walter Guest (1970)
The Conscientious Objector
75-294/76-28
Bernadotte Folke (1949)
Instead of Arms

75-296
Hartill Percy (ed.) (1941)
Into the Way of Peace
75-297
Brittain Vera (1942)
Humiliation with Honour
75-298
Abelson Robert P., Zimbaro Philip G. (c1970)
Canvassing for Peace; a Manual for Volunteers
75-299
Brockway Fenner, Mullally Frederic (1944)
Death Pays a Dividend
75-300/83-341
Paciter Thomas (c1971)
People to People for World Peace
75-301
Franks Lucinda (c1974)
Waiting out a War; the Exile of Private John Picciano
75-302
Society of Friends, American Friends Service Committee (c1966)
Peace in Vietnam; a New Approach in Southeast Asia
75-303
Callin George, Brittain Vera, Hodges Sheila (comp.) (1945)
Above All Nations; an Anthology
75-304
Lieberman Mark (c1972)
The Pacifists; Soldiers Without Guns
75-305
Broucker Jose de, Briffault Herma (tr.) (c1969)
Dom Helder Camara; the Silence of a Peacemaker
75-306

2.2 IN コマンドによる入力処理の概要

IN コマンドを入力すると、PRINSE への入力用プログラムの実行が開始されるが、このプログラムの具体的な利用法を述べるに先立って、ここでまず、その概要を説明しておく。

PRINSE への入力処理は、文献を 1 件ずつ処理するという形を取るが、文献の処理は利用者の側からすればふたつの、プログラムの側からすれば 3 つの、ステップに分れる。第一のステップは、1 件の文献を構成する 3 レコードを画面表示し、修正、変更等が必要であれば、これを行うというデータ編集のステップである。このステップでは、データの削除、挿入、圧縮文字列(後述)の表示、登録等も行われる。次のステップは、このデータについて、どの単語をキーワードとして登録するかを利用者が指定するキーワード処理のステップである。キーワード処理のス

テップでは、キーワードの指定だけでなく、ストップワードの指定、キーワードからの除外指定、キーワード処理モードの変更、登録済みキーワードの表示、などの処理を行うことができる。また、このステップから、データ編集ステップに戻ってデータそのものを再び編集することもできる。このステップが終了するとプログラムは、キーワードとして指定された単語の登録、アルファベット順での順位付け、索引の作成などを行うが、これは利用者の目には見えない。このようにして1件の文献の処理が終わると、次の文献の処理に移る。このように書くと、1件の文献の処理に相当に時間がかかるように思われるかも知れないが、我々の経験では、1件の文献の処理に要する時間は、ある程度処理に慣れれば、平均して30秒程度である。

入力データをすべて処理してしまえば、プログラムが終了するのは、言うまでもないが、入力データがまだ残っていても、入力処理はいつでも中断できる。このときには、次に IN コマンドを起動すると、それ以前に処理された文献の次の文献から処理が始まる。また、中断、終了いずれの場合でも、新しく作成あるいは追加されたデータを PRINSE のキーワードディスクに書込む作業が行われ、これが終わって始めて IN コマンドの実行終了となる。

2.3 IN コマンドの実行開始

IN コマンドを実行して文献データを PRINSE へ入力するためには、予め次のようなディスクを用意しておく必要がある。

PRINSE キーワードディスク（既存のものに追加するときには、それをそのまま使用する。独立のデータベースを新規に作成するときには、MS-DOS でフォーマットしたものを用意する）

PRINSE データディスク（既存のものに追加するときには、それをそのまま使用する。独立のデータベースを新規に作成するときには、MS-DOS でフォーマットしたものを用意する）

入力データの入ったディスク

作業用ディスク（MS-DOS でフォーマットしたもの）

また、IN コマンドにおいては、PRINSE の他のプログラムと異なり、1 文字の応答が要求される場合には、文字どおり 1 文字のみ応答すればよい。例えば、

.....(ret / yes , else / no)

という、IN コマンドでしばしば用いられるプロンプトに対して、“no” の応答をするときには、(リターンキー以外の) 任意のキーのみ入力すればよく、その後でリターンキーを押す必要はない。

IN コマンドを投入すると、まず、「ユーティリティディスクを取り出せ」という指示があるから、これに従う。続いて、登録可能なキーワードの種類数、今回処理可能なキーワードの延べ数が、画面に表示される。この数字は、使用しているパソコンのメモリに依存するので、必ずしもプログラムの制限値ではない。また、よほど大量の文献を中断なしに処理するのでない限りは、特に気にする必要はない。

これに続いて、

Always confirm your response (ret/yes, else/no)?

というプロンプトが現れる。応答の意味は次のとおりである。

リターンキー：問い合せに対する利用者の応答がある度に、

sure (ret/yes, else/no)?

という確認の問い合せをする。(確認モード、オン)

それ以外の任意のキー：確認の問い合せを一切行わない。(確認モード、オフ)

確認モードをオンにした場合、IN コマンドの使用に慣れると煩わしいこともあるが、最初のうちはオンにしておいた方が安全であろう。確認モードは、後のキーワード処理のステップで自由に変更できる。

次に、

Create a new system (ret/yes, else/no, addition)?

という問い合せがある。応答の意味は次のとおり。

リターンキー：PRINSE のデータベースを新たに作成する。

それ以外の任意のキー：既存の PRINSE データベースにデータを追加する。

PRINSE, 具体的には IN コマンドでは, どのような既存のデータベースにどのようなデータを追加してもよいし, また新たに独立のデータベースを作成してもよい。但し, 同一データベース内での文献の重複, 二重入力は一切チェックしない。

ここで, 追加と応答したときには, 「キーワードディスクをドライブ A にセットせよ」という指示があるので, これに従う。そして, 既に登録済みのキーワード, ストップワード, 入力途中の入力データの有無など, 以後の処理に必要な情報を読み込むと, 「キーワードディスクを取り出せ」という指示がある。

これに対して, 新規作成を指定したときには, 以下の問い合せに答える形で, 新しいデータベースを定義しなければならない。この定義中に誤りがあったり, 定義をやり直したいときには, HELP キーをおせば, 定義を最初からやり直すことができる。また, ここでの定義の大部分は, IN コマンド実行時, あるいは主プログラム PRINSE の実行中に変更, 追加, 再定義できる。

Printer (ret/PC PR 201, else/others)?

Line size (= characters per line) (ret/80, 40-120)?

Different fields in different printing lines (ret/yes, else/no)?

Words may be divided at line ends (ret/no, else/yes)?

Printing font (ret/elite, P/pica, C/condensed)?

以上 5 つの問い合せは, 検索された文献の印刷に関するものであり, PRINSE の U コマンドとまったく同じであるので文献 2 を参照されたい。

Keyword registration mode (ret/inclusive, else/exclusive)?

IN コマンドでは, キーワードの登録を原則として自動的に行うが, 登録の方法はふたつある。ひとつは, 包括モードで, このモードの下では, 表題レコード中の単

語は、ストップワードとして既に指定されていない限り、すべてキーワードとして扱われる。これに対し、選別モードでは、表題レコード中の単語は、既にキーワードとして登録されていない限り、キーワードとしては扱われない。登録モードは、キーワード処理ステップで何度でも切換えることができる。なお、著者名レコード中の著者名は、登録モードの如何に拘らず、常に包括モードで処理される。

Connecting symbol for a compound word(ret/ + , else/others)?

前述のように“United Nations”のような複合語は、特に指定しない限り、2語(あるいはそれ以上)に分割される。複合語として処理したいときには、構成要素を、ここで定義する連結記号で結んでおかなければならない。標準の連結記号はプラス記号(+)であるが、それ以外のものを使用してもよい。但し、連結記号は、2.1で挙げた単語の区切り記号であってはならない。標準の記号を使うときには、リターンキーのみ応答する。このとき、入力データ中で、“United + Nations”、“Third + World”等としておけば、“United Nations”、“Third World”(キーワードとしては連結記号が空白で置換されていることに注意)が、それぞれ1語として処理される。これ以外の記号を使用するときには、ここでリターンキー以外を応答する。すると、「記号を入力せよ」という指示があるから、任意の(単語の区切り記号でない)記号を入力する。ここで定義した連結記号は、変更できない。(なお、複合語の扱いをしないときにも、上の問い合せにリターンキーのみ応答しておく。)

Do you define stopwords (ret/yes, else/no)?

ストップワードとは、入力データ中には存在するが、検索のためのキーワードとはしない単語をいう。欧語文献の場合、冠詞、前置詞などは一般にストップワードにするのが普通である。PRINSEの場合、ストップワードの指定は、必ずしも必要ではないが、ディスクやメモリの効率的な使用、処理速度といった観点からすれば、“and”、“the”など出現度数の大きい語は、ストップワードに指定しておくことが好ましい。ただ、ストップワードは必ずしもここで指定する必要はない。キー

ワード処理ステップで対象となる単語が文献データの中に出現したときに、個別に指定してもよい。ここで、ストップワードを指定すると答えたときには、指示に従ってストップワードを入力する。入力は大文字でも小文字でもよい。ストップワードは、1語最大15文字までで、キーワード処理ステップで追加されるものも含め、最大120語まで指定できる。

Do you define strings to be compressed (ret/yes, else/no)?

圧縮文字列を定義するか否かを答える。圧縮文字列とは、文献データをディスクに記憶するとき、1ないし2バイトの特殊なコードに圧縮変換して記憶する文字列である。出現度数が大きく、長い文字列を指定すれば、データディスクの利用効率を相当に高めることができる。圧縮文字列は、ここで定義すると応答したときにも、あるいは後のデータ編集ステップでも定義あるいは追加できる。いずれの場合でも、1件最大62文字まで、最大60件まで指定できる。また、圧縮文字列では、大文字と小文字は区別される。入力は指示に従ってキーボードから行うが、特例として、圧縮文字列を最初に定義するときには、ディスクから（即ち、既に作成され、ディスクに記憶されているリストを）入力できる。このときには、ファイル名の定義、入力ディスクのセット等を、指示に従って行えばよい。

Do you use a keyword list already created (ret/yes, else/no)?

この問い合せに対しては、リターンキー以外の一文字を応答しておけばよい。

データの追加時で、入力データにまだ処理されていない文献がある場合を除き、次には、入力データの入ったファイル名の問い合せが行われる。そして、入力データのあるディスクをドライブAにセットせよという指示があるのでこれに従う。続いて、PRINSE データディスクを指示に従い、ドライブBにセットする。これでPRINSE への入力処理の準備がすべて整ったことになる。

2.4 データ編集ステップ

データ編集ステップでは、1件の文献について、必要に応じ、修正、削除等を行う。また、圧縮文字列の表示、登録も行う。このステップでは、まず入力データから1件の文献を読み込み、それを、例4のように、著者名レコード、表題レコード、書誌事項レコードの順に画面の上半分に表示する。そして画面中央に、次の処理を問うプロンプトが表示される。画面の下半分は、システムからの様々な出力、利用者の入力のために使用される。この画面構成は、次のキーワード処理ステップでもまったく同じである。

利用者は表示されたコマンドのうちからひとつを選択して、その一文字を入力することにより次の処理を指示する。コマンドの機能を忘れたときには、単にHを押せば、各コマンドの簡単な説明が表示される。なお、コマンドは、大文字、小文字のいずれで入力してもよい。

(例4)

Data Number= 12

Roberts A

Research on a possible alternative to military defense

Bulletin of Peace Proposals (1970) Vol.1, No.3, p.286-

next step (H=help) (ret, C, S, I, D, R, L, H, E)

利用者の応答や入力に何等かの誤りがあったときには、エラーメッセージが表示され、更に疑問符が表示される。このときには、任意のキーを押せば、エラーの直前の処理が再開される。

修正、変更がまったく必要ないとき、あるいは修正等の作業が終わったときには、プロンプトに対しリターンキーのみ応答して、次のキーワード処理ステップに進む。

以下、コマンドとその機能、利用法を個別に説明する。リターンキーとEコマンドを除けば、同一文献処理中に同一コマンドを何度用いてもよい。また、コマンドによっては、コマンドの実行終了後、直ちにコマンド入力待ち状態、即ち次のデータ編集用コマンドを入力できる状態、に復帰するものもあるが、そうでないコマンドの場合、

Any key to return to editing

というプロンプトが表示されるので、任意のキーを押せば、コマンド入力待ちの状態に戻る。

リターンキー

データ編集ステップを終了し、次のキーワード処理ステップに進む。キーワード処理ステップのBコマンドで、再びこのデータ編集ステップに戻ることもできる。

Cコマンド

データ中の文字列を置換する。このコマンドを投入すると、

source string =

result string =

というふたつのプロンプトが現れるので、それぞれ、置換前の文字列、置換後の文字列を与える。文字列を削除するには、置換後の文字列としてリターンキーのみ入

力する。このコマンドでは、文献データの先頭から、つまり著者名レコードの先頭から、データを走査して、最初に見つかった文字列のみを置換する。置換の結果は、画面では黄色で表示されるので、これにより置換が意図したとおりに行われたかどうかを確認する。

S コマンド

現在処理中の、つまり現在表示されている文献をスキップし、PRINSE には入力しない。入力データ中で削除される訳ではない。このコマンドを投入すると、スキップしてよいかどうか確認するプロンプトが出るので、スキップしてよければリターンキーのみ、そうでなければそれ以外のキーを入力する。スキップしてよいと応答したときには、直ちに次の文献が読み込まれ、次の文献の編集ステップに入る。

I コマンド

このコマンドと次のDコマンドは、入力データとして必要な、著者名、表題、書誌事項いずれかのレコードの重複、脱落に対処するためのものである。このコマンドはレコードを挿入し、Dコマンドはレコードを削除する。レコードの挿入、削除は次のような仕組で行われる。今、図1のように、画面にa、b、cという3つのレコードが、この順で、表示されているとする。このとき、xというレコード(x自体は、後述のように、端末から入力される)をbの位置に挿入すると、b、cの位置がずれて、図2に示すように、画面には、a、x、bが表示される。そして、a、x、bがそれぞれ、著者名、表題、書誌事項と解釈されることになる。このとき、元あったcは、システムの用意した退避領域に退避される。更に、図2の状態ではyというレコードをaの位置に挿入すると、結果は図3のようになる。逆に、図3の状態ではaを削除すると、xの位置が繰り上がり、空になった3番目の位置には、bが退避領域から呼び戻され、退避領域にはcだけが残る。結果は図4のようになる。退避領域が空のとき、例えば、図1の状態ではbを削除したときには、入力データの次のレコードが読み込まれる。

退避領域には、最大3レコード(即ち文献1件分)まで退避することができる。退

避されたレコードは先入後出 (Last in, first out) の順で呼び戻される。I, D コマンド使用時の退避, 呼出については上で述べたとおりである。また, 1 件の文献, 例えば, 図 4 の y, x, b から成る文献の処理が終わって次の文献が読み込まれるときには, まず退避領域にあるレコード (図 4 の状態では c) が呼出される。これにより, 退避領域は空になる。そして, その後入力データの次のレコードが必要なだけ (この例ではふたつのレコード d, e が) 読み込まれる。その結果, c が著者名レコード, 入力データから読み込まれたふたつのレコード d, e が順に表題レコード, 書誌事項レコードと解釈されることになる。(図 5 参照)

I コマンドを入力すると,

field number (1-3) = ?

というプロンプトが出るので, 新しいレコードを挿入する位置を数字で答える。数字の意味は次のとおり。

- 1 著者名レコード
- 2 表題レコード
- 3 書誌事項レコード

続いて, 指示に従い, 挿入すべきレコードをキーボードから入力する。

D コマンド

このコマンドは, 画面に表示されているレコードのいずれかを削除する。削除後の処理については I コマンドの項で述べたとおりである。削除するレコードは, I コマンドと同一のプロンプトに対し, 数字で指示する。数字の意味も同じである。このコマンドで削除されたレコードは, (I コマンドで利用者が再入力しない限り) 復活できない。

R コマンド

圧縮文字列の登録を行うことを指示する。このコマンドを与えると, 短い説明とともに, 圧縮文字列の入力要求があるから, 指示に従い圧縮したい文字列を入力す

る。

Lコマンド

既に登録済みの圧縮文字列を表示する。

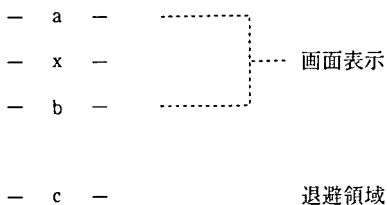
Hコマンド

データ編集用コマンドの機能を画面表示する。

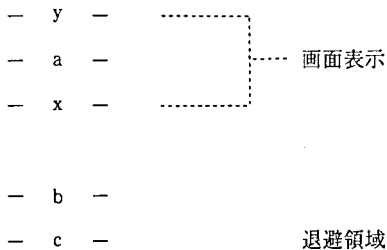
(図1)



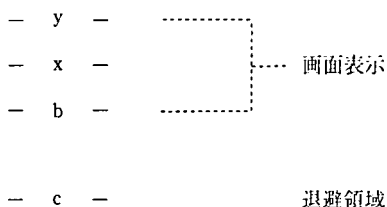
(図2) 図1 bの位置にxを挿入した結果



(図3) 図2 aの位置にyを挿入した結果



(図4) 図3の a を削除した結果



(図5) 図4の文献の処理終了後、次の文献が表示された状態



(c は退避領域から, d, e は入力データから)

E コマンド

PRINSE へのデータ入力処理を一旦終了する。前述のように、まだ入力されていない文献が残っていても差し支えない。実際、処理すべき文献が無くなったときには、プログラムが処理を自動的に終了するので、このコマンドを与えるのは、中断の場合だけである。次に、IN コマンドを起動したときには、このコマンドを投入したときに画面に表示されていた文献から入力が再開される。このコマンド投入後の手順については、2.6で述べる。

2.5 キーワード処理ステップ

データ編集ステップが終わると、キーワード編集ステップに移るが、その間に、小さな補助的ステップがある。これは、次のステップで、キーワードとして扱われる単語について、英語の複数形の処理を簡略化するためのステップである。このス

テップでは、キーワードの候補のうちで、sで終わるものについて、sで終わる形が既にキーワードとして登録されていなければ、どちらの形で登録するかを問う。条件に合うすべての単語について個別に問い合わせがあるので、その都度応答する。この問い合わせは、条件に合う単語すべてについて機械的に行われるので、ノイズも多く、時には繁雑でさえある。しかし、単数、複数の処理を首尾一貫して行うには、次のキーワード処理ステップで利用者自身が行うより、このステップを利用した方が安全であろう。この問い合わせを行うかどうかは、次のステップのMコマンドで随時切換えることができる。

キーワード処理ステップでは、キーワードの登録、除外、ストップワードの登録、削除等を行う。このステップにおける、画面構成、コマンドの入力方法、エラー処理は、上述2.4のデータ編集ステップと同じである。しかし、このステップでは、前のステップと異なり、文献データ中の単語の状態を以下のように色分けして表示する。キーワード処理ステップにおける主たる作業は、単語の状態を変えることである。

黄色：そのままの形でキーワードとして既に登録されているか、登録予定の

単語

緑：画面に表示されているのは別の形（例えば、単数形で）キーワードとして既に登録されているか、登録予定の単語

紫：ストップワード

白：上記以外の単語

以下、このステップにおけるコマンドとその機能、使用法を述べる。

リターンキーとBコマンドを除き、同一コマンドを何度使用してもよい。実行後自動的にコマンド入力待ち状態に復帰しないコマンドからキーワード処理コマンド入力待ち状態への復帰方法は、データ編集ステップの場合と同じである。

リターンキー

画面に表示されている文献のキーワード処理を終わり、次の文献の処理に移る。以後、次の文献のデータ編集ステップに移る。

K コマンド, S コマンド, C コマンド, A コマンド

この4個のコマンドは機能は異なるが、画面の操作方法は同じであるので、まずKコマンドを例として操作方法について述べる。

コマンドを投入すると、画面の下半分に、どのコマンドを実行中かの表示と、簡単な画面操作の説明が現れる。実際の作業はこの説明を見ながら行えばよい。それと同時に、画面上部のどれかの単語表示が赤にかわる。(どの単語が最初に赤で表示されるかはコマンドに依存する)。カーソル移動キーを押すと赤で表示される単語が次々と変わっていく。キーワードとして扱われていない単語をキーワードとして登録したいときには、カーソル移動キーを押して目的の単語を赤に変え、その状態でリターンキーを押す。キーワードとしたい単語がまだあれば、必要なだけこの操作を繰り返せばよい。処理が終わったら、エスケープキーを押す。勿論、何もせずにエスケープキーを押してもよい。すると、コマンド入力待ちの状態に戻る。ここに挙げた他の3つのコマンドの場合もやり方はまったく同様である。以下、4つのコマンドの機能を簡単に述べる。

Kコマンドは、キーワードを登録する。そのままの形でキーワードとして扱われていない単語(画面で黄色以外の色で表示されている単語)であれば、どの単語でも直接キーワードとすることができる。但し、2.1で述べたように、著者名レコードの左括弧以降の単語、文字列と、書誌事項レコードの単語はキーワードとして登録できない。なお、一度キーワードとして登録された単語は削除できない。誤って意図しない単語を登録した場合、処理中の文献に関しては、Cコマンドを使用してキーワードから除外すればよいが、Kコマンドの効力は、後述のように以後の文献にも持続し、その単語が以後の文献に現れる度に、キーワードとして扱われるから、その都度Cコマンドで除外しなければならない。ただ、無意味なキーワードが10や20あっても苦にならなければ、Cコマンドで削除する手間は省いて構わない。余程出現頻度の大きい単語でない限り、実害は無く、審美感の問題である。出現頻度が大きく、メモリやディスクの記憶効率を圧迫するようなら、Sコマンドでストップワードにしてしまうという抜け道もある。

S コマンドは、ストップワードを登録する。既に登録されたストップワードの削除は、後述のLコマンドにより行う。

K コマンド、S コマンドによる指定は、現在処理中の文献だけでなく、以後処理される文献に関するも有効である。これに対し、次のA、C 両コマンドの効力は、現在処理中の文献の外には及ばない。

A コマンドは、画面に表示されたのとは異なった形のキーワードとして登録する。このコマンドの実行中にリターンキーを押すと、問題の単語をどのような形でキーワードとして登録するかを問うプロンプトが表示される。これに対しては、任意の単語または文字列を答えてよい。例えば、“US”を“America”、“United States”、“USA”あるいは“XYZ”等のいずれの形で登録してもよい。入力は大文字、小文字いずれでもよい。また、ここで、新しい形を与えるときには、アスタリスク(*)を用いた略記法が使用できる。*に任意の文字または文字列を続けると、問題の単語、つまり画面で赤で表示されている単語の語尾から*に続く文字(列)を削除した形を与えたものと解される。例えば、“taxes”を“tax”という形で登録したいときには“*es”と入力すればよい。指定された形がまだ登録されていないときには、その形をキーワードとして新たに登録するか否かの確認が行われる。ここまでの作業が終わると、再び、画面操作可能な状態に戻る。

このコマンドの効力は処理中の文献限りであるので、“taxes”、“tax”のような変換は、必要に応じその都度行わなければならない。

C コマンドはキーワード(さして意味は無いが、場合によっては、ストップワード)として扱われている単語を、キーワード(ストップワード)から除外する。但し、これは、このコマンドの効力が現在処理中の文献だけにしか及ばないことからして、現在の文献ではキーワードとして扱わないことを意味するだけであり、一旦登録されたキーワードをシステムから削除することを意味しない。

D コマンド

現在までにキーワードとして登録されていない単語で、画面にも表示されてい

ないが、何らかの理由でキーワードとして登録しておきたい単語があるとき、直接キーボードからキーワードを登録する。登録すべき語を入力せよという指示があるから、これに従って入力する。入力は、大文字、小文字いずれでもよい。尚、ある語句がキーワードとして既に登録されているかどうかは、Lコマンドにより知ることができる。

Mコマンド

キーワード登録モード、応答確認モード、複数形処理モードを表示、変更する。

Which mode (ret/registration, C/confirmation, A/assignment)?

というプロンプトが表示されるので、表示、変更したいものを選択する。応答の意味は次のとおり。

リターンキー：キーワード処理モード(包括モード、選別モード)の表示、変更

C : 応答の確認モード(オン、オフ)の表示、変更

A : sで終わる単語の自動処理モード(オン、オフ)の表示、変更

いずれかを選択すると、そのモードの現在の状態が表示され、変更するかどうかの問い合せがあるので、これに答える。

Lコマンド

キーワードの点検、ストップワードの表示、削除を行う。

Keywords or stopwords(ret/keywords, else/stopwords)?

というプロンプトに対する応答により、いずれかを選択する。

キーワードを選択したときには、任意の語句が既にキーワードとして登録されているか否か、登録されているとすれば今回の処理で何件の文献に出現しているかを知らせる。このときには、チェックしたい語句を入力せよという指示があるから、これに従い最大6語まで入力する。6語未満で入力を終わるときには単語入力

要求のプロンプトに対しリターンキーのみ入力する。登録されていない単語は、その旨のメッセージが表示されるが、現在処理中の文献で登録予定のものは、未登録の扱いを受ける。登録されたキーワードの場合、表示される文献数は、今回の処理開始後その語句が出現した文献数であって、新規作成時を除き、作成中のデータベース全体での出現文献数ではなく、しかも現在処理中の文献を含まない。

ストップワードを指定したときには、ストップワードがアルファベット順に表示される。この後、ストップワードから削除する単語があるかという問い合わせがあるので、あればその旨応答し、指示に従い削除すべき語を入力する。

Wコマンド

このコマンドは、表題レコードが長いときのキーワードの遺漏を点検することを目的としたもので、処理中の文献のキーワード、ストップワード、それ以外の単語のいずれかのみを画面に表示する。

どれを表示するかは、

Word types to show (ret/keywords, S/stopwords, N/non-registered)?

に対する応答として指示する。応答の意味は次のとおり。

リターンキー：キーワードだけを表示する

S ：ストップワードだけを表示する

N ：キーワード、ストップワード以外の単語だけを表示する。

I コマンド

現在までの登録済みのキーワード数、ストップワード数、処理済の文献数、入力済の行数、メモリの残りバイト数などの情報を表示する。

Bコマンド

データ編集ステップに戻る。このとき、キーワード編集ステップで与えられたCコマンド、Aコマンドの効果は失われるが、それ以外のコマンドの効果は失われない。

H コマンド

キーワード編集ステップの各コマンドの機能を説明する。

2.6 IN コマンドの実行終了

PRINSE へのデータ入力には次のふたつの場合に正常に終了する。

入力データをすべて読み終ったとき

利用者がデータ編集ステップの E コマンドで終了を指示したとき。

言うまでもないことだが、データ入力に異常終了することもある。これはプログラムのバグによることもありうるが、ほとんどの場合メモリの不足が原因である。前者の場合、“ERROR” に続けてふたつの数字が表示されるので、筆者に連絡されたい。後者の場合は、異常終了する前に、メモリが残り少ないという警告メッセージが出るので、その時点でデータ編集ステップの E コマンドを用いて処理を中断する。なお、異常終了しても、それ以前に作成されたデータは損われることはない。

正常終了の場合は、指示に従い以下の手順でディスクを出し入れする。

入力データの入ったディスクを抜取る。

PRINSE データディスクを抜取る。

PRINSE キーワードディスクを A ドライブにセットする。

作業用ディスクを B ドライブにセットする。

作業用ディスクをセットした後、暫く待つと、すべてのディスクを抜取れという指示がある。この指示が出たら、ディスクを抜取り、電源を切ってよい。指示に従って、リターンキーを押すと MS-DOS のエラーメッセージが出るが、これは無視してよい。

3 著者索引、キーワード索引の作成

3.1 実行開始とオプションの指定

作成された PRINSE キーワードディスクとデータディスクを用いて、著者名のアルファベット順による全文献のリスト(著者索引)と表題中のキーワードのアルファベット順による全文献のリストを出力することができる。出力は、ディスクとプリンタのいずれかを選択できる。また、特定の著者、あるいはキーワードのみを対象とするオプションと、著者名あるいはキーワードを見出しとして出現する文献数を付けて出力するオプションのふたつのオプションがある。著者索引は AI コマンド、キーワード索引は KI コマンドにより作成するが、オプションの指定方法など使用方法も、出力の形式もまったく同じであるので、ここでは、紙幅の関係も在り、キーワード索引についてのみ説明する。

コマンドを投入すると、PRINSE の初期画面が現われ(文献 2 参照)、キーワードディスクをドライブ A に、データディスクをドライブ B にセットせよという指示があるから、これに従う。次に、

Outout device (ret/disk, else/printer)?

というプロンプトに対し、希望の出力先を答える。続いて、

Header (ret/yes, else/no)?

という問い合せに対しては、個々のキーワードとその出現する文献数を出力するか否かを応答する。出力を選択したときには、個々のキーワードについて、そのキーワードを有する文献のリストの出力に先立ち、

キーワード (出現度数)

という見出しを出力する。これに続いて、

Any selection (ret/no, else/yes)?

というプロンプトに対し、出力するキーワードを選択するか否かを応答する。選別を行わないと応答したときには、すべての(表題中の)キーワードについて、それが出現する全文献が出力される。選別を行うと応答したときには、以下に述べる方法で利用者の指定したキーワードを有する文献のみが出力の対象となる。選別を行うときには、出力の対象とすべきキーワードを入力せよという指示があるから、これに従って出力したいキーワードをひとつずつ入力する。キーワードは最大100個まで与えることができる。途中で入力を終えるときには、入力要求のプロンプトに対し、リターンキーだけを入力する。ここで注意すべきは、実際に出力されるのは、

ここで指定されたキーワードと完全一致あるいは前方一致するキーワードを含むすべての文献だということである。従って、場合によっては、相当のノイズが含まれることもあるが、逆に、索引をアルファベットの1文字あるいは数文字ごとに分けて出力することができるという利点もある。

以後、処理が終了するまで利用者は何もする必要がない。

3.2 出力の形式

出力はキーワードのアルファベット順に行われる。即ち、アルファベット順で最初のキーワードを有する文献がまずすべて出力され、次に、2番目のキーワードを有する文献がすべて出力されるという形で対象となったすべてのキーワードについてそれが出現するすべての文献が出力される。この説明からも明らかのように、1件の文献が場所を変えて複数回出力される可能性はきわめて高い。見出しが出力されるときには、そのキーワードの出現する文献の出力に先立って行われる。ひとつのキーワードについて出力される文献の順序は、入力データからの入力順である。また、1件の文献は著者名レコード、表題レコード、書誌事項レコードの順で出力される。

ディスク出力の場合も出力形式はここで述べたとおりであるが、若干の補足説明が必要である。まず、ディスク出力は、MS-DOSのテキストファイルとして、常にBドライブのディスク、即ち、PRINSEデータディスクに出力される。出力ファイルの名称は、著者索引が“ALIST”，キーワード索引が“KLIST”と定められている。出力では、見出し、著者名レコード、表題レコード、書誌事項レコードがそれぞれ1行(1レコード)である。このとき、見出しでは、行の先頭にふたつのアスタリスク(*)が付される。ディスク出力で注意すべきは、出力は、BドライブのPRINSEデータディスクに空きがある限りにおいてのみ可能であるということである。将来的には、PRINSEデータディスク、キーワードディスク以外のディスクに出力できるよう改良する予定であるが、差し当たりは、ディスク出力を行うときには、以下の点に留意して行うのが安全である。

実行前にMS-DOSのコマンドを使って、PRINSEデータディスクの残り

容量を調べておく。残り容量がディスク容量の2/3以下であれば、全キーワードあるいは全著者を対象とした索引の出力は失敗する可能性が高い。

実行前に、PRINSE データディスクのバックアップを取っておく

実行時には、キーワード選択機能を利用して、索引の出力を何度かに分けて行う。

データディスクに出力された索引は、速やかに他のディスクにコピーし、データディスク上の“ALIST”あるいは“KLIST”は削除しておく。

3.3 実行終了

終了方法は、主プログラムの場合と同じである。文献2を参照されたい。

結び

本報告では、前回の報告(文献1)の末尾に挙げた問題点のうち、ごく一部の改良しか報告できなかった。将来的なシステム全体の大幅な改良はともかくとして、差し当たり、次の点は早急に改善、改良すべきであろう。

作成されたデータベースに含まれる文献の削除(この機能があれば、追加は用意のできるから文献の修正そのものも可能になる)

作成されたデータベース中のキーワードの削除、併合、置換等の編集機能

文献

1. 松尾雅嗣(1986)「パーソナルコンピュータによる平和研究文献の簡易検索システム」広島平和科学第8号 pp.297-309
2. 松尾雅嗣・竹花誠児(1986)「パーソナルコンピュータによる平和科学研究センター所蔵文献の検索システム」平和科学研究通信 vol.9, no.1, pp.1-10
3. 村瀬康治 入門MS-DOS, アスキー出版局
4. アスキー出版局編(1986)標準MS-DOSハンドブック, アスキー出版局