

NRDFシステムの強化

北海道大学理学部 富樫雅文

1 概要

NRDFの管理システムについて次のような機能強化を行なった。

- (1) コード辞書の付加。
- (2) 情報コマンドの追加
- (3) DISPLAYコマンドの改良
- (4) キー項目の追加

NRDFのシステムでは各種の情報の表現に統制された簡略語彙(コード)を使用していたが、これまではそれらは人間によって管理されておりシステムの直接の管理下にはないものであった。そこで、これらのコードの体系を計算機向きに整備してNRDFの管理するシステムファイルの一部とする改善を行なった。

これによって人間によるコード体系の管理において生じがちな一貫性の不備を防ぎ、データ作成時や検索時あるいは入力データの検査時に同一の辞書ファイルを使用することによってコードの使用法における混乱を避けることができる。さらに、これまでは入力データの中に不当なコードが使用されていてもシステム自身がコード表を持っていないためにエラーとして検出できなかった点を回避することができるようになる。また、このコード辞書のファイルに種々の利用者向け案内情報を含ませ検索時に随時これを参照できるようにして利用者の便宜を図ることにした。

次に検索コマンドの一つであるDISPLAYコマンドでは、これまでは指定した検索結果のすべての情報を表示するようになっており、検索の結果、大量のデータが得られた場合、それを表示することは実際上不可能なことがあった。これに対して、DISPLAYコマンドに幾つかのパラメータを加え、出力すべき情報を選択できるようにした。このとき、コマンドの文法として多くのパラメータを容易に指定できるように考慮を払った。

最後に検索の時に使用できるキー項目として数値の表に付けられたデータ項目の名前と単位の名前を新たに追加した。これまでの使用によって、これらの情報を手掛かりとして検索したいという要求が多かったためである。以下にこれらの強化点について個別に述べる。

2 コード辞書の付加

これまではコード表としてカードイメージのファイルを持っていた。これらの中にはNRDF固有のコードとEXFOR(国際核データ交換フォーマット)におけるコードの2種類の情報源からのコードが混在していた。EXFORからは雑誌及び研究所コードが移入されている。システム組み込みの辞書ファイルを作成するに当たってはこの既存のコード表を交換し必要な情報を付加する形で行なった。

辞書に含まれるコードを次の5種類に区分する。

(1) 項目名 (Field)

Fタイプ辞書

データ記述文の項目名として左辺に現れるコード

(2) 項目値 (Value)

Vタイプ辞書

データ記述文の項目値として右辺に現れるコード

(3) 基本語彙 (Word)

Wタイプ辞書

ハイフンでつないだ複合コードを構成するための基本コード

(4) システム用語 (System)

Sタイプ辞書

NRDFシステムに関連する非統制一般語彙

(5) 検索コマンド名(文法の為に) (Command)

Cタイプ辞書

コマンドの文法が説明として入っている。

(6) 検索コマンド名(使用例の為に) (Example)

Eタイプ辞書

コマンドの使用例が説明として入っている。

各辞書の各語には次のような情報が含まれる。

- (1) コード名 (見出し語)
- (2) コードの展開形
- (3) コメント
- (4) タイプ (F, V, W, S, C, E)
- (5) (Fタイプ辞書のとき)
適合する項目値コードのクラス(複数個指定できる)
- (Vタイプ辞書のとき)
所属するクラス(1から14までに分類されている)
- (6) 情報源(現在はEXFORコードのときのみ記録)
- (7) 作成/更新日付 yyyymmdd
- (8) (単位名のコードのとき)
基本単位名
- (9) (単位名のコードのとき)
基本単位に対する換算比率

Vタイプ辞書に対するクラスの分類は次のようになっている。

クラス番号 分類

- | | |
|----|------------------|
| 1 | 研究所名 |
| 2 | 雑誌名 |
| 3 | 核反応の型 |
| 4 | 加速器に関する情報 |
| 5 | 検出器に関する情報 |
| 6 | 分析法に関する情報 |
| 7 | 物理量 |
| 8 | 標的核に関する情報 |
| 9 | YESとNO |
| 10 | 未知及び不確定値 |
| 11 | 光学模型のポテンシャルパラメータ |
| 12 | その他 |
| 13 | 粒子名 |
| 14 | 単位名 |

辞書ファイルはカードイメージの外部形式を持ちVSAMファイル上の内部形式との間の変換をおこなうことができる。外部形式は次のようになっている。

第1カラム<----->第72カラム

コード名

- 1個以上の空白 展開形
1個以上の空白 /+制御情報
1個以上の空白 /*コメント

制御情報はタイプ、クラス、情報源、日付、基本単位、換算比率についての記述の集まりで、各項目について次の形式を

持ったものである。

TYPE=タイプコード

CLASS=クラス番号

SOURCE=情報源

BASE=基本単位名

RATE=換算比率

DATE=日付

コード名の最大長は31文字であるが、展開形、制御情報及びコメントは2行以上になってもよい。

内部形式では先頭にフィールドディレクトリーをもつ可変長レコードとし、各項目の値を連結している。

辞書ファイルの保守

辞書ファイルの保守を行なうために、従来のユーティリティプログラムを拡張した。

保守の作業として次のようなことが行なえる。

- (1) IJNRDFコード表からの外部形式辞書データの作成
- (2) EXFORコード表からの外部形式辞書データの作成
- (3) 外部形式辞書データから内部形式への変換(ローディング)
- (4) 辞書ファイルから外部形式データへの部分/全体変換(アンローディング)
- (5) 個々のコードの表示/追加/消去/更新

この辞書データの編集に当たってはTSSの画面エディターを起動してカードイメージで行なえるようにした。

辞書ファイルの利用

辞書ファイルは次のような使われかたをする。

- (1) 入力データの作成の時の手引き
- (2) 入力データの自動検査

作成された入力データはこれまで構文的な検査のみを行なっていたが、この辞書ファイルの組み込みによってコード使用の正当性の検査も合わせて行なえるようになった。入力データ検査プログラムでは入力された記述文に対して次の様な検査を行なう。項目名がFタイプ辞書からまた項目値をVタイプ辞書からおのおの検索する。項目名及び項目値のクラス指定が一致するかどうかを調べる。すなわち、両者のクラスが交わり(空でない積集合)をもてばよい。なお著者名などのような項目には適合クラスの指定はなく、項目値のコード検査は行なわない。

- (3) 検索時の辞書引き

オンライン検索を行なうとき、検索条件の指定にコード名を必要とすることがある。また検索結果として表示されたデータに現れたコードの意味が分からない時など、この組み込み辞書を参照できるように辞書表示コマンド(LISTDCコマンド)を用意した。コマンドの構文は、

LISTDC (辞書タイプ 比較演算子 指定文字列);

で、辞書タイプはF, V, W, S, C, Eのいずれか。

比較演算子は文字列比較のための次の6種類のうち一つ。

=, ^=, >, >=, <, <=, ^>, ^<

意味は文字型の索引に対する検索時と同じである。

指定文字列には知りたいコード名またはその一部を与える。また@は「すべて」の意味で用いられる。

3 情報コマンドの追加

辞書のうちSタイプ、Cタイプ及びEタイプ辞書は、正確には、統制されたデータ用コード表ではなく利用者のための案内情報であるが、これを組み込み辞書に加える事によって辞書ファイルへの共通アクセラレーターを利用した情報コマンドを設けた。これは特定の見出し語について辞書を調べるにはLISTDCコマンドより簡便である。

WHATコマンド

これはコードの意味や検索コマンドの文法を表示させる：

WHAT [見出し語]；

HOWコマンド

これは検索コマンドの使用例を表示させる：

HOW [コマンド名]；

どちらのコマンドもパラメータを省略した場合はすべての検索コマンドについての情報を表示する。

一般にオンラインシステムでのHELP機能と呼ばれるものでは指定されたコマンドや事項の説明を表示することになっているが、未習熟の利用者が一つのシステムについて使用法を覚えていくのは、まずそのシステムのさまざまな使用例を見ることによる場合が多い。このため単なるHELPではなく、これをWhat Howという疑問詞を使って視点を交えた問い合わせを行なえるようにした。

4 DISPLAYコマンドの改良

出力すべき情報を次の観点から制限選択できるようにした。

- (1) 出力する検索結果集合の指定 (OFパラメータ)
- (2) 検索結果集合の中のレコード指定 (INパラメータ)
- (3) 各レコードから取り出すセクションの種類 (FORパラメータ)
- (4) 表示する項目の指定 (WITHパラメータ)

各パラメータは<前置詞><空白><指定値>の組で表現され、指定順は任意である。

DISPLAYコマンドは次のような形式となる。

```
DISPLAY [(OF) 集合名]
          [(FOR n1 [, n2 [, n3]])]
          [(IN セクション名リスト)]
          [(WITH 項目リスト)]
          [(PLOT)]
```

N1, N2, N3は 第N1レコードから第N2レコードまでN3個刻みで出力することを示している。セクション名としてはBIB, EXP, DATAから選択する。

OFパラメータは先頭にある場合のみ前置詞「OF」を省略できる。また各パラメータは省略可能で、その時は省略時解釈が適用される。

このコマンドのように多くのパラメータを指定しなければならない時には混乱や誤りが生じやすい。このような場合、一般に位置パラメータではなくキーワードパラメータを採用することが多い。しかしそれによってもコマンド入力の際のぎこちなさは残る。ここにたいし自然語(英語)に倣って前置詞を導入すると、言語による思考の流れにある程度沿うことができるため、混乱や心理的抵抗を回避することが期待できる。一つの試みとして提案したい。

5 キー項目の追加

新たに数値の表に付けられたデータ項目名とその単位名を検索用キーとして使えるようにした。原則としてこれまでは検索の為の索引情報は入力データのうち、<項目名=値>の形式をもつ記述文からとり、表の内容は索引されなかった。しかし、表の内容を大まかに示すこの「データ項目名」という情報は正確な検索を期するために重要な貢献をしてくれるので、これを新たにキー項目とすることにした。

6 数値情報データベースの展望

自然言語やフォーマット化されたコード情報に対するデータベース設計と構築運用の理論と技術は安定期に入ったといえる。しかしながら学術研究や個人的情報管理などスキーマ抽出の困難な世界に対する情報の的確な管理方式については漸進的な問題解決とともに新しい視点による理論と技術が要求されているといわなければならない。

一つの課題として数値情報のデータベース化がある。数値情報は確立された演算体系と表現形式のもとに電子計算機の伝統的な処理対象であった。情報としての細部を切り落とし、実用上の制限を課した数値情報の集まりをデータベース化することはこ

れまでも行われてきた。しかし真に要求されていることはある種の不確定さと多様な環境の下における数値情報の柔軟な把握と処理ではなからうか。また、数値計算プログラムや実験装置とこのデータベースの相互作用が円滑に行われるならば、それは研究の生産性を飛躍的に向上させることになる。

このような意識を以って数値情報データベースのありうべき姿を展望するとき、次のような乗り越えられるべき困難のあることがわかる。

- (1) 不確定性を含めた数値情報の表現形式の不備
有効数字、各種誤差、各種空値など
- (2) 単位系の計算機による認識が無い
- (3) プログラミング言語における非データベースの入出力機能
- (4) 文字列志向のファイル管理

これらに対してまず第1に数値情報の拡張された表現形式と演算体系を定める必要がある。そこでは、単位を有し不確定さのある数値が自在に取り扱えるようではなくてはならない。次にFORTRANを始めとする数値処理用語の入出力仕様をデータベース向きに強化する必要がある。

一方、既存のOSの下でのデータ管理サブシステムでは事務系の処理の影響もあってか数値情報、特に浮動小数点数を意識した基本アクセスメソッドがない。レコードのアクセスキーとして数値を用いる事は極く自然であり、必要である。

これらの問題を解決したとして、そのうえで数値情報のデータベースの持つべき特性を考えると、

- (1) 動的なスキーマの設定/解放
必要に応じて枠組みを設ける
- (2) スキーマとしての理論値及び統計値の取り扱い
理論値との比較による検証、仮設生成と検証、及びスキーマ化
- (3) 外部表現の多層化という観点からのビューのサポート
図形表示、単位変更など
- (4) 数値計算ライブラリーや統計パッケージとの結合

このためにはさらにソフトウェアデータベースを要する

等が挙げられる。これらは既存のデータベース理論の外縁に位置付けられるものでもあり、その実現はひとり数値情報のみが恩恵を被るものではない。