

検索例と検索システムの問題点

北海道大学情報処理教育センター 岡部成玄

ユーザが一般にデータベースシステムに対し望んでいることは、必要なデータを必要な時に必要な形で手にいれることができるということであろう。最初の条件はデータの質と量に、2番目は主として運用に、最後のは検索システムに関係するものである。

NRDFの検索システムの詳細は「荷電粒子核反応データファイルNRDF使用説明書(第2版)」(北海道大学大型計算機センター 千葉正喜氏)にある。ここでは、検索の概要を示すとともに検索して感じたことを記することにする。

NRDFの検索システムは、著者(ATH)、標的核(TGT)、物理量(PHQ)等の索引項目により検索条件を与えると高速に検索を行い、データを表及びグラフの形でディスプレイ或いはプリンタ上に出力してくれる。

NRDFのデータは論文から採取している(現在のところ日本国内で生産された測定データを対象としている)。これがNRDFのデータの質と量を基本的に規定している。検索の面でいうと検索データは少なくとも論文毎には区切られているということである。これは検索結果の表示においても見ることができる。

NRDFでは検索データの単位をDATA SETと呼んでいる。それは数値データ(DATA SECTIONと呼ぶ)に実験条件(EXP SECTIONと呼ぶ)と論文の書誌情報(BIB SECTIONと呼ぶ)を加えたものである。書誌情報は当然一論文に一つであるので、表題(TITLE)、著者(ATH)、研究機関(INST-ATH)、雑誌(REF)といった索引項目での検索は論文を探すことを意味する。

さて、例として ^{16}O の p 散乱の微分断面積のデータの検索を考える。まづ、各計算機システムの手続きに従って処理を行い(北海道大学大型計算機センターではLOGONを行い、READYモードでHDBとキーイン送信し、問い合わせに答えればよい)、NRDFの空間にはいると次の表示になる。

KEY-IN COMMAND,PLEASE.

:

終了は

END;

とキーイン送信する。

どんなコマンドが使えるかをみるには、辞書表示コマンドLISTDCで、

LISTDC (C=@);

とする。ここでCはコマンドを、@は全部を指す。又、= + * > <等の演算子は数値、比較、論理、集合の各演算に使われる。個々のコマンドの使い方を知りたいときは、情報コマンドHOWを使って、

HOW LISTDC;

とする。検索を行うためには検索条件を(TGT=160)の様に(索引項目 演算子 値)を組み合わせた形で与えることが必要である。TGT(標的)といった索引項目のリストは「使用説明書」にあるが、

LISTDC (F=@);

で、又値のリストは

LISTDC (V=@);

で見ることができる。いづれも大量であるので注意する。核種名は¹⁶Oを160の様に与える。

検索するまえにどんなデータが登録されているか知りたいときは索引表示コマンドLISTX3で、

LISTX3 (YEAR=@);

発行年

LISTX3 (ATH=@);

著者

LISTX3 (INST-ATH=@);

研究機関

LISTX3 (REF=@);

雑誌

等とする。研究機関や雑誌名は略称を用いている。例えば略称2J A P O S Aの正式名称を知りたい時は情報コマンドWHATを使って

WHAT 2J A P O S A;

大阪大学

とする。

いよいよ検索を始める。

KEY-IN COMMAND,PLEASE.

:

(TGT=160)*(PRJ=P);

標的核=¹⁶O 入射粒子=陽子(p)

COMMAND=

(TGT=160)*(PRJ=P);

検索結果はレジスタ(後で参照するときの名称は#)に確保

* HIT-COUNT=

100

KEY-IN COMMAND,PLEASE.

:

*(PHQ=DSIGMA/DOMEGA)=SAVE1;

上の結果に測定物理量として微分断面積を

指定し、結果を SAVE1 の名前で確保

COMMAND= *(PHQ=DSIGMA/DOMEGA);

* NEW SET SAVE1 CREATED

* HIT-COUNT= 90

KEY-IN COMMAND,PLEASE.

:

DISPLAY SAVE1;

結果の表示

```
000000000 000000000 000000000 DISPLAY SAVE1 000000000 000000000
0 000000000
=====
| DATA SET 2049 |
=====
| SECTION 2820 |
=====
| BIB SECTION | 2049~2059 |
=====
| D# | D276 |
| TITLE | MICROSCOPIC DESCRIPTION OF 800-MEV POLARIZ |
.....
| SECTION 2821 |
=====
| EXP SECTION | 2049~2059 |
=====
| IRCT | 160(P,P)160 |
.....
| IPHQ | DSIGMA/DOMEGA | 1 |
| IPHQ | ANALPW | |
| IPHQ | OPT-POTL-PARA | |
| IPHQ | DNSTY-DSTRN | 12 |
.....
| SECTION 2822 |
=====
| DATA SECTION | 2049 |
=====
| INC-ENGY-LAB | 800MEV |
.....
| TABLE |
|-----|
| THTC | DSIGMA/DOMEGA | DELTA-DSIGMA/DOEMGA |
| (DEG) | (MB/SR) | (MB/SR) |
| 6.14 | 9.10E+02 | XXXXXXXX |
| 6.40 | 7.90E+02 | XXXXXXXX |
.....
| DATA SET 2050 |
=====
| BIB SECTION | ----> SEE SECTION 2820 |
=====
| EXP SECTION | ----> SEE SECTION 2821 |
```

```

=====
SECTION          2823
=====
DATA SECTION    | 2050
=====
INC-ENGY-LAB   | 800MEV
=====

```

TABLE

| THTC (DEG) | ANALPW (NODIM) | DELTA-ANALPW (NODIM) |
|---------------|-------------------|-------------------------|
| 5.99 | 2.13E-01 | 2.55E-02 |
| 6.29 | 2.23E-01 | 2.04E-02 |

```

=====
DATA SET        2051
=====
BIB SECTION    | ----> SEE SECTION 2820
=====
EXP SECTION    | ----> SEE SECTION 2821
=====

```

```

=====
SECTION          2824
=====
DATA SECTION    | 2051
=====
RSD             | 160
EXC-ENGY        | 0.0MEV
J-PI            | 0+
INC-ENGY-RANGE-* | 800MEV
*LAB            |
=====

```

TABLE

| A (NODIM) | Z (NODIM) | DATA1*1' (FM) | DATA1-ERR (FM) |
|--------------|--------------|------------------|-------------------|
| 16 | 8 | 2.66 | 0.2 |

```

COMMENT          | *1' DATA1 = NEUTRON ROOT MEAN SQUIRE RADIUS |
COMMENT          | COMMENT = CALCULATIONS USING EQUAL PROTON |
                  | AND NEUTRON DENSITY DISTRIBUTION |
                  | IONS ARE NOT SIGNIFICANTLY DIFFERENT FROM |
                  | THOSE OBTAINED USING THE OPTIMIZE FIT |
                  | TO THE NEUTRON DENSITY. HENCE OUR RESULTS |
                  | ARE CONSISTENT WITH EQUAL PROTON AN |
                  | D NEUTRON DENSITY DISTRIBUTION |
=====

```

```

=====
SECTION          3026
=====
BIB SECTION    | 2192~2202
=====
D#             | D296
TITLE          | SPIN OF THE 14.174 MEV T=3/2 LEVEL IN 17F |
=====

```

論文の表題のみ見たいときは

DISPLAY SAVE1 WITH TITLE;

とする。他の索引項目を見たいときはそれを指定するとよい。

レジスタに確保されているものの表示は

DISPLAY;

とする。

表示を中止したいときは割込キーを押すと次の問い合わせがある。

<<<*** ATTENTION ACCEPTED ***>>>

<<<*** ENTER S TO STOP, G TO GO (S/G) ?:

終わりにいくつか注意事項を記しておく。

(1) 文字列の一致

| 例 | 検出されるTITLE |
|--------------|-----------------|
| TITLE=ALPHA | ALPHA |
| TITLE>ALPHA |ALPHA..... |
| TITLE>ALPHA* | ALPHA..... |
| TITLE>*ALPHA |ALPHA |
| TITLE<ALPHA* | A AL ALP ALPH |
| TITLE<*ALPHA | A HA PHA LPHA |
| TITLE<ALPHA | L LP LPH P PH H |

(2) 数値の指定

INC-ENGY-CM=20 (単位を省略するとMEV) といった指定は完全一致を求めており計算機に於ける数値表現の精度を考えると適当ではない。むしろ次のように範囲指定をするのがよい。

| 例 | 意味 |
|----------------------|----------|
| INC-ENGY-CM>5 | E< 5 |
| INC-ENGY-CM<30.5 | E> 30.5 |
| INC-ENGY-CM=EE(5,30) | 5<=E<=30 |
| INC-ENGY-CM=ET(5,30) | 5<=E< 30 |
| INC-ENGY-CM=TE(5,30) | 5< E<=30 |
| INC-ENGY-CM=TT(5,30) | 5< E< 30 |

但し、指定範囲に登録データが含まれていないといけない。つまり、指定範囲が登録データの一部である時は該当データ無しとなる。つまり、大きめの範囲指定が無難である。

(3) エネルギーの指定

| | |
|--------------|----------------|
| INC-ENGY | (実験室系、重心系の区別無) |
| INC-ENGY-LAB | (実験室系) |
| INC-ENGY-CM | (重心系) |

を区別していることに注意する。

以上、簡単に検索例を示したが、NRDFの検索システムは検索が高速で良いがユーザインタフェースという点では改善すべき所がまま見られる。

例えば、ある論文に複数のエネルギー点において測定された散乱断面積の角分布がのっているときは各々の角分布の数値データの集まりが、それぞれNRDFにおけるDATA SETを構成する。従ってこの場合一論文から複数のDATA SETが作られる。このような時、複数のDATA SETが同じEXP SECTIONを共有することが多い。この場合特に測定物理量の記述がEXP SECTIONにあるため、検索がすっきりしない事がある。つまり、EXP SECTIONに複数の測定物理量の記述がある場合、測定物理量とDATA SECTIONとの対応がとれていないので、測定物理量(PHQ)で検索すると、ユーザの念頭にない同一論文に載っている他の数値データ(DATA SECTION)も一緒に引き出されることになる。更に、EXP SECTIONの測定物理量の記述では明瞭ではないデータからなるDATA SECTIONもあり一層複雑である(この場合はデータの説明がCOMMENTとして与えられている。但し、このCOMMENTは検索対象外である)。この点に関しては次の様になっていると使いやすいのではないか。

(1) 測定物理量の記述がDATA SECTIONにある。

(2) そのとき検索システムに認められていない測定物理量(実際上はDATA 1等で表示)の数値データを見ることができるためにも、同一論文内にどんな測定物理量があり、どういうコードで引用できるか表示させることができるようにする。

これに関連してDATA SETやSECTIONの番号も検索・表示対象になっている事が望ましい。その他検索結果の表示の仕方を含めコンピュータの会話型利用を念頭においた改善が望まれる。

最も改善が望まれる点は、ユーザが検索の結果得られたデータを加工して2次データを作成・利用することを支援することである。検索結果がユーザファイルに確保でき適当なスクリーンエディタを用いて更に編集・加工できれば明らかに便利である。グラフ表示もユーザが2次データを使って、例えばSAS/GRAPH等により行ったほうが有効である場合が多いように思われる。又グラフ出力を文字表示を使って行っているが、最近では端末における図形出力は標準的になってきており、更にグラフィックスのソフトも国際的標準規格とされるGKSが普及してきており、こういった事情を考慮すべき時期にきているようにも思われる。