

# 計算機で文献を探す話

西村 史朗\*

## 1. 夏の日の思い出

最初に筆者の個人的な経験から始めたい。1967年ごろアンドロメダ座28番星(28 And)という星が、たて座 $\delta$ 星型の変光(振幅0.08等, 周期1.7時間)を示すことを見つけた。論文を書く段になって、その星の変光がすでに知られていないかどうか、確かめておかねばならない。こういうときには、Astronomischer Jahresberichtを調べるのが、一番普通のやり方である。これは1899年に始って、1年1冊の割で刊行され、1年間に発表された天文学の論文の表題・要旨を分類して集めてある(1969年以後はAstronomy and Astrophysics Abstractに引きつがれ、1年2冊発行されている)。ある夏の1日、閲覧室の机へ数冊ずつ運んでは、変光星の項目に28 Andが出ていないか調べて行く。結局1899年以降の変光観測の報告がないことを確認したのは、長い夏の日もようやく暮れかける所で、黄ばんだ日ざしが窓外の樹々に落ちていたことをまざまざと思い出す。そのうち他の手づるで1887年に、眼視観測で80日間に0.3等の変光の報告を見つけたが、これはたて座 $\delta$ 星型の変光とは別のものと考えられるので、このことも含めて変光発見の報告を書いた。同じころ大沢氏の発見したHD 221568のときも、ほぼ半日を論文探索に費した。

このような観測の例ばかりでなく、理論の分野を含めて、研究分野についてすでにどのような論文が発表され、どんな知識が蓄積されて来たか知っておくことは、少くとも研究のある段階では不可欠のことである。研究者個人のレベルでは、ある人は記憶の中に、ある人はカード型式でという風に、多かれ少なかれ長い年月蓄積されているものであるが、1つの分野・対象について能率よくまた落ちこぼしなく探そうとすると、組織的に整備された手段が必要となり、上述のような文献抄録誌は最も有力な手段である。

しかし最近の天文学の急激な発展と、それに伴う論文などの情報量の増大、また一方では天文学の特質として、古いデータが重要な意味を持ち続けることが多い。こうして益々重要になるとともに、おびただしく労力を必要とするようになった情報の収集を、何かの手段でもっと能率よく出来ないものだろうか。“万能”と称する電子計算機は利用できないものだろうか。

## 2. 計算機と情報処理

一般に計算機とは、文字通り計算をするための道具と考えられているが、引算を応用すれば、データの中から特定の値に等しいものを探したり、大小関係で分類するのに利用できる。比較の対象になるものは数値データだけでなく、名前のような文字列データでも、計算機の内では一定の規則で数値に変えられている(コード化)ので、文字列データも比較判別できる。計算機は速さ、精確さ、根気強さの点で人間よりも遙かに勝れているが、これらの利点こそ情報検索に要求される能力である。ところが計算機にデータを処理させるには、2つの準備が必要である。1つはデータを計算機が読める形にして与えること、第2はデータを処理する手順を計算機に指令すること(ソフトウェア)である。当面データの種別は文献情報に限定するとして、文献を探索する項目をどのような種類に設定するかが問題になる。上述の文献抄録誌の場合、研究分野で分類されていて、例えばパルサーとか恒星大気とかいう項目では分類されているが、個々の恒星名で探すには便利ではない。この型の計算機用データとしてはINSPEC-A(物理)の文献データベースが、京都大学大型計算機センターで供用されはじめ、天文の文献も含まれている。

今一つの型の文献検索は、天体名を鍵とするもので、個々の天体の個性が重要な要素となる天文学の特性に根ざしたものである。特定の星の集合(輝線星やA型特異星など)に対するものが出版されているが、恒星全般に関する文献検索データは、始めから計算機による検索を目指して、計算機を利用して作製された。恒星文献索引(Bibliographical Star Index, 略称BSI)と呼ばれるが、1977年末にストラスブール(フランス)の恒星データセンターから金沢工業大学計算機センター(日本の天体データセンター)を経て、磁気テープの形でデータを手に入れたので、これを使いこなすためのソフトウェアを作った。

## 3. データを積上げる

当面の問題で最大の難関は、大量のデータを計算機に読める形にして集積することである。最近印刷あるいは手書きの文字を読取る機械もあるが、専門家の判断も加えながら、人手によって進めて行かねばならない。恒星名による文献検索のためのデータは、フランスでムードン天文台のケレルらによって、1960年代から集め始

\* 東京天文台 S. Nishimura: Computer Retrieval of Bibliographical Information

められた。データを作り上げる手順は次のようなものである。

先ず 1950 年以降の一定範囲の天文学の論文雑誌を読み通して、それぞれの論文名とその中に出て来る恒星名との対照表を作る。論文のテーマになっている星は勿論のこと、変光星の比較星や、論文のなかで一瞬触れている星の名も全部採られている。この表を星の名前によって並べ換えるのだが、論文名は著者名・所在（雑誌名・巻数・頁）・論文標題を含んでいて、長たらしい（平均 100 字弱）ので、論文名を論文番号で置き換え、論文番号と論文名の対照表を別に作っておく。このように作っておくと、文献を探す作業は、星の名前から文献番号を探すこと（1 箇のこともあれば多数箇のこともあるだろう）と、文献番号から著者名・所在・標題を引き出すという 2 段階に分けられる。

現在のデータは 1975 年までの恒星文献と 1976 年までの惑星状星雲の文献を含んでいる。収められている天体の数は 69348、文献の数は 9494。一番多くの文献に現れるチャンピオンはこと座  $\alpha$  星 (312 文献)、これに NGC 7027、こと座 RR 星が続いている。逆に 1 回しか現れない星は天体数の約 60%、10 回以下が 93% と大部分を占めている。

#### 4. データを探す

さて今や計算機の中に文献検索用のデータが入った。しかし計算機の中といっても、いわゆるメモリー（内部記憶装置）の中にはない。東京天文台の計算機は中型程度といってよいだろうが、メモリーの大きさは 0.768 MB (MB はメガバイトで 100 万バイト、1 バイトは 1 文字に相当する) しかないので入りきらないし、東京大学大型計算機センターでも 8 MB の程度で、そのうち 6 MB も占領すると他の人が困ってしまう。普通こういときには、外部記憶と呼ばれるもののうち、磁気ディスクに書き込んでおく。磁気テープだと、検索を始める前に一々テープをセットしなければならない。磁気ディスクは、いつでも計算機から呼び出せるので、不特定多数の利用者が、不定の時に利用したいと思っても便利に使える。

磁気ディスクはレコードに似ているが、磁氣的に記録するので何度でも書き直しができる。レコードの溝に相当するものは、渦巻きではなく同心円で 400 本くらいある（以下特に断らない限り、計算機についての数値例は 1979 年現在の東京天文台の FACOM 230-58）。さらに円盤が 19 面積み上げられて、レコード針の代りに磁気ヘッドがその間を出入りして、1 台で約 100 MB (1 億字) のデータを記憶する。これくらいあれば、6 MB 入って貰っても、利用度が高ければ、そう肩身が狭くはな

い。磁気ディスクから少しずつ計算機のメモリーに入れて、望みのデータを探しに行く。

ある星の名前を探しに行くとして、7 万個あるデータを片はしから調べて行くというのは、いくら計算機が速いからといってうまい方法ではない。先ず星の名前は 10 桁の数字コードに変えておいて、小さいものから大きいものへと順に並べておく。さらに磁気ディスク上のある単位ごとにその単位の中の最後の名前を目次を作っておくと、探したい星の数字コードと大小を比較するだけで、データの大体の所在が分り、あとはデータ全体のごく一部だけから探し出せばよい。さらに何単位かをまとめた大きな単位ごとの目次を作ったり、それを階層的に増して行くと、巨大なデータファイルからでも、能率よく望みのデータに到達することが出来る。これはちょうど何巻にもわたるぼう大な書物で、先頭の巻に全体の簡単な目次があり、各巻にそのもう少し詳しい目次があり、さらに時折見られるように、各章の先頭に詳細な目次がおかれているような構成によく似ている。

#### 5. 星の名前はいろいろ

最初のところで、28 And の文献探しに 1 日、HD 221568 には約半日だったと書いたが、この違いは何故だろうか。それは 28 And の方が明るい星で、BS 114、HD 2628、BD +28°75 という名前があり（その他にもいっぱいあるが、通常使われるのはこれくらい）、これに対して HD 221568 には他に BD +57°2758 しかない。28 And の文献を探すには、BS 114 などという名前にも注意を払いながら探さねばならなかった。こうした事情は BSI ではある程度解決されていて、文献の中で 28 And とでも BS 114 とでも現れると、全部 1 つの名前で（今の例では HD 2628）まとめられている。

逆に利用する方の立場では、28 And で探して文献がないと分っても、本当にこの星に文献がないということにならない。検索するときの名前を計算機用語ではキーというが、うまいキーを見つけないと知識の扉は開かない。BSI では名前についての漠然とした優先則があって、HD 番号が最優先になっているが、検索の前に 28 And の HD 番号を調べておくのも面倒である。何とか計算機に探して欲しいところであるが、ストラスプールで推奨されている方法は恒星同定カタログ (Catalogue for Stellar Identifications, 略称 CSI) を使うことである。

CSI はストラスプールで強力に編集を進めているカタログで、今も増訂中である。4 つのデータファイルから成り立っていて、それらを総合して恒星の名前の間の関係が分るようになっている。BSI のデータ編纂の際に、上に書いたように星の異名をまとめるのに CSI を利用

している。CSIを文献検索に使おうとすると、一番困ることはCSIのデータがぼう大なことで(全部で19MB)、BSI本体(3MB)よりずっと大きい。磁気ディスクは高価なので、出来るだけ節約したい。またCSIによって、例えば28 AndからBS114, HD2628……という名前の関係が得られても、どれがBSIを検索するのに有効なキーなのか分らない。

問題を逆の方向から考えてみよう。前準備としてBSIに現れる星の名前をおのに対して、CSIその他を使って、同じ星を指すすべての他の名前を並べ上げる。そうしておいて主従関係を逆転して、他の名前からBSIでの名前が分るように並べ換える。これは天文台の計算機で7時間くらいかかる大仕事であるが、1度作っておけばBSIやCSIが改訂されるまで有効に使えるデータファイルが得られる。ファイルの大きさは3MBであるが、BSIの名前を探すのに必要十分なデータを含んでいる。

実際に使うときは、探したい星の名前で先ずBSIを探しに行き、なければこの恒星名変換ファイルを探す。ここで見つければ、変換された名前は必ずBSIにある。もしどちらにも見つからなければ、その星の文献はBSIの中には1つもないということになる。

## 6. TVディスプレイの前で

もし貴方が恒星文献を探したいと思ったら、検索システムを置いてある計算機につながるテレビ端末のところへ行き、所定の簡単な手続きで検索プログラムを起動する。テレビ画面には使用歓迎のメッセージが現れ、星の名前を尋ねてくる。もし、こと座 $\alpha$ 星なら“\* ALF LYR”と鍵盤から打ち込むと、たちまち文献の表が画面に現れる。結果はプリンターに印刷することも出来る。もし途中で使い方が分からなくなったら“HELP”と打つと、計算機の方からヒントを与えてくれる。大抵の仕事は数分で済んでしまうだろう。

計算機は“万能”というけれども、人間が指令した以上のことは出来ないということは、最早一般の常識になったようである。上のような計算機とのやり取りやデータを探す手順を予じめ指定しておくプログラムが必要である。計算機と対話しながら仕事を進めて行くやり方

は、情報検索のような場合に特に有効である。例えばこと座 $\alpha$ 星を呼び出して見て、文献数が多すぎるなら1970年以後の最新の文献だけを出すようやり直すことも可能である。

データを探すことは本質的に重要なことであるが、お定まりの手順で実行できる。会話型のプログラムが便利に使えるかどうかは、むしろ枝葉のつけ方で決まるようである。例えば上の例で、実はこと座 $\alpha$ 星は計算機の中では(-12, 6010052)というコードで現わされている。しかしこれを利用者に覚えることを要求する訳にはいかない。\* ALF LYRと書けば、計算機の内部コードに変換し、逆にBSIのコードを人間に読める形にして表示することはプログラムが受持っている。星の名前の形式は種々雑多なので、プログラムも相当複雑になる。

恒星文献検索プログラムは、1978年春にまず東京天文台の計算機のために作られた。しかし東京天文台の計算機は孤立したシステムであるので、他の大学の研究者が直接使うことが出来ない。そこで同年秋までに東京大学大型計算機センターにも置くことにした。東大センターは全国から電話を利用して、テレビ型やタイプライターの端末装置をつないで利用することが出来る(また会話型でなくカード入力でも実行することも出来て、遠隔地からは郵送で依頼できる)。まもなく大学センター間のネットワーク回線が使えるようになり、益々便利になるだろう。最初は京都大学などにも置くことを考えたが、ネットワークなどを使って1ヶ所にある大量のデータを共同で利用する方が、ディスクなどの資源が有効になり、今後このような利用の傾向が強まって行くだろう。

日本国内の他分野の様子を見ると、化学をはじめとして計算機による文献検索が盛んに利用されている。天文学では天文データ全体としてみても計算機利用はやっとスタートし始めたという感を否めない。今回の検索システムの基礎となったBSIはフランスで作られたものである。プログラムと恒星名変換データは自前であるが、後者の原材料はほとんど“輸入品”である。国際間の学問的分業は勿論歓迎すべきことであるが、少なくとも“計算機大国”であるはずの日本が、輸入超過を続けてよいのだろうか。

## 掲 示 板

### 日本学術会議第12期会員選挙について

日本学術会議中央選挙管理会

昭和55年11月には、3年に一度の日本学術会議会員選挙が行われます。

この選挙は、会員を選挙する方も、会員に選挙される方も有権者でなければなりませんので、次のことに御留意ください。

- (1) 新たに有権者としての登録を希望する方は、登録用カードを早めに提出してください。
- (2) 引き続き有権者の方は住所、勤務機関、勤務地等登録カード記載事項に変更があった場合は、すみやかに異動届を提出してください。

以上について不明の点がありましたら、下記にお問い合わせください。

〒106 東京都港六本木 7-22-34  
日本学術会議会員選挙管理事務室  
03-403-6291