

STELLAR BIBLIOGRAPHY RETRIEVING SYSTEM IN JAPAN

Shiro Nishimura
Tokyo Astronomical Observatory, University of Tokyo

ABSTRACT

On-line retrieving systems of BSI in Japan is briefly described. The BSI reference file was adapted to be retrieved by some keywords other than star names. Then the retrieval system of BSI was constructed on a general retrieval language CIR. Some efforts to make the system efficient and to save the disk space are discussed.

RETRIEVING SYSTEMS BASED ON A DEDICATED PROGRAM

Since 1978 two on-line retrieving systems of BSI have been operated in Japan. Two systems are almost identical, and the one is on UNIVAC 1100 of the Tokyo Astronomical Observatory, the other on HITACHI M200-H of the Computer Center, University of Tokyo. The latter system can be accessed through the public telephone network by astronomers and related scientists in all over Japan, while the former system is used locally and for the system maintenance. The retrieving program including code and decode subprograms of star names was written in PL/I language. In addition to the BSI reference file and the inverted file of keywords (star names), a third file of the thesaurus of star names was compiled at the Tokyo Astronomical Observatory to make sure hitting of references from variant star names. Data of the file were taken from CSI and other supplementary data. Details of the systems were given elsewhere (Nishimura 1981).

So far BSI can be retrieved only by star names. The usage of delimiters and the abbreviation of journal names were inhomogeneous through the original BSI reference file. Furthermore the program was so simple that almost no logical operations could be applied to the sets of the retrieved references. In order to make the retrieving system more versatile, it will be more profitable to resort to a generalized retrieving language than to develop a complicate unique program. A retrieving language CIR (Conversational Information Retrieval) of UNIVAC Japan

was recently made available at the Tokyo Astronomical Observatory. Although CIR has rather limited functions, it was adopted to make the first trial.

ADAPTATION OF BSI TO DATABASE APPROACH

It was necessary to convert the BSI reference file into a unified form. Processing of the file on the computer was divided into many small steps, and was executed semi-automatically by manual interventions. The final form of each reference was " Author(s) < Journal name > < Volume, (Number,) Page > Title ". (In case of books it was modified as " Author(s) <%T Book title, (%E Editor(s),) (%P Publisher,) (%C Comments) > < Page > (Chapter title) ".) The author entry had a form " Family name, I. I'. ... (, Jr. or III etc.) / ..." where I and I' denote the initials of the first and middle names. The journal names were unified in accordance with the abbreviations in the Astronomy and Astrophysics Abstracts, reducing the number of journal names from 465 to 133. Similarly the number of authors was reduced from about 7000 to 6000.

RETRIEVING SYSTEM BASED ON CIR

The items of the extended keywords are as follow :--- 1. Reference numbers, 2. Star names, 3. Author names, 4. Journal names, 5. Book titles, 6. Titles. The last two items were not registered in the inverted file, and will be retrieved only through the reference sets extracted by other items. All items except star names were derived from the BSI reference file, while star names were acquired from the BSI star file (the inverted file) by re-inversion. As there was no thesaurus function in the CIR system, one could have listed up all the variant star names, which corresponded to each object appearing in a reference, and included them together with the BSI star names as the keywords of the reference. On the average an object has four names. Brighter stars, however, which have more variant names than the average, tend to appear more frequently in the bibliographies. The total number of keywords of star names reached three million, and the maximum number of star names corresponding to a single reference was eighteen thousand, which far exceeded the limitation of CIR. After all only one name of each object, that is the BSI star name, was registered as the keyword.

The CIR system affords the "Library File", which enable the macro-command of retrieval. If a library element named by an variant star name have been created and its contents dictate the retrieval with the keyword of the BSI star name, then one can invoke the retrieval of references relevant to the object simply by entering the variant name. Some star names, such as those in double star catalogues, denote many different objects in BSI. Even in such a case, the summation of the sets of references for each object will be obtained, as the library element may contain several retrieval commands including the sum operation.

Data of the thesaurus file were converted to the elements of the "Library File" in the CIR system. A shortcoming of the Library File was a bulky space in the disk necessary to the on-line retrieval. Star names in some catalogues, which will not often be used to denote an object, were deleted from the file. In spite of these efforts, the present size of the library file was 160 MB in addition to 15 MB of the fundamental data files of BSI-CIR system, compared to 8 MB of the total data space in the dedicated program system. The space of the library file will be soon reduced to a half of the present size by the revision of the CIR system program. It is expected to be furthermore compressed to about 20 MB by introducing the variable length structure, since the present file is occupied by meaningful data very sparsely. Then there would be little difficulty in the disk space to make the steady on-line retrieval feasible.

The test of retrieval proved successful. Retrieval by various kinds of keywords could be executed, and the sum, difference or product of sets of references could be obtained in an arbitrary mode. It became easier to reach the reference one wanted to have. The CIR system provides convenient tools of database maintenance. The revision of BSI original file could be performed more easily by reorganization of updated CIR database.

FUTURE PLAN

Another project is now being planned to construct a system based on ORION of HITACHI at the Computer Center, University of Tokyo. It will serve for the nation-wide demands, and the allowance of the disk space will be favorably given to such a commonly used database. If the files are too large to be located in the disk, they will be stored permanently in the "Mass Storage System" and transferred into the disk when used.

SUMMARY

A retrieval system based on a general purpose program was tested. It was revealed that this approach was successful in the retrieval of references in flexible mode of operation, and as the tool of the maintenance of BSI files. The difficulty to hold a large space in the disk will be overcome in near future.

REFERENCE

Nishimura, S. : Bull. Inf. CDS, N. 20, Pp82-85.