

A PERSONAL COMPUTER SYSTEM FOR RETRIEVAL OF NQRS DATA

Koichi MANO* · Hideaki CHIHARA**

ABSTRACT

A compact software was developed for loading/searching NQR spectral data from the source database which contains 10,155 records as of May, 1991. The software is designed for a personal computer with a hard disk storage of 20Mb and runs, hopefully, on any personal computer that can compile ANSI C language. This software, named NQR PC SEARCH, was actually tested successfully on NEC's PC9801/RA21, IBM's PS/5530Z, PS/55note, and PS/2, and several other IBM compatible machines.

INTRODUCTION

Nuclear Quadrupole Resonance (NQR) frequency reflects sensitively the electronic environment of the resonant nucleus. Therefore the measurements of NQR frequencies provide useful information about chemical bonding, molecular- and crystal-structure, phase transition, molecular motion in solid state etc. The technique can be also applied to identification of chemical substances or to nondestructive analysis of unknown specimens. There have been reported a number of papers of NQR studies since the first paper due to Dehmelt appeared in 1950.

Since 1980 Japan Association for International Chemical Information (JAICI) and the NQRS data committee have collected and compiled the numerical data of NQRS frequency in the computer-readable form.^{1,2)} The data thus collected (designated as NQRSDB) have been loaded for online, TSS search through the computer center of Osaka University. As of May, 1991, the total amount of NQRSDB reached 10,155 records (ca 9 Mb) which includes all the NQR frequency data published since 1950 up to the end of 1989.

We have now developed a search system of NQRSDB on a personal computer (PC) to assist the researchers in the world for easier use of the database.

SPECIFICATIONS OF THE SYSTEM

Our primary objective was to develop software which allows the NQRSDB to be

*Faculty of Liberal Arts, Shinshu University, 3-1-1 Asahi, Matsumoto 390, Japan

**Japan Association for International Chemical Information (JAICI) Gakkai Center Bldg., 2-4-16 Yayoi, Bunkyo-ku, Tokyo 113, Japan

loaded and used on a personal computer by a researcher who has no prior knowledge of programming, database management, and online searching. The system then has to have a maximum flexibility regarding the hardware requirements and must run in a standalone fashion, i. e. independent of other software as far as possible. The specifications were thus like the following.

1. Target PC : IBM's PC/XT, AT, PS/2, NEC's PC9801 series.
2. Programming Language : Turbo C Version 2.0. Only ANSI C standard functions may be used.
3. Any commercial software other than MS-DOS and Turbo C may not be used, e. g. dBASE III. If NQRPC were supported by dBASE III, users would also have to purchase dBASE III and further the portability of the system will then be restricted.
4. We use the MS-DOS version of the latest NQRSDB edition as the database (DB).
5. NQRPC thus consists of the following three modules.
 - a) MAKE_INDEX_FILES : The preprocessors extracting the index files from DB in advance.
 - b) PC_SEARCH : The main module including search program.
 - c) SETUP () : Installation program.

DATABASE AND SEARCH ELEMENTS

DB contains the following data elements.¹⁾ (See Fig. 1.)

1. CAS Registry Number and Molecular Formula of the substance.
2. Substance Name.
3. Crystal Modification.
4. Resonance Data.
 - a) Method of measurement.
 - b) Nuclear mass number.
 - c) Temperature of measurement in the unit of K.
 - d) Resonance frequency in the unit of MHz.
5. Author Names.
6. Journal Data including CODEN, Journal name, volume number, publication year, and pages.
7. Keywords and Keyphrases.
8. Remarks.
9. Record Number (RN) at the last seven columns of each line.

These data elements are written in card image format (80 columns). The card pack thus obtained is sandwiched between a number-card and an end-card to make a logical record. DB consists of 10,155 logical records sorted in the ascending order of the record number.

The following data elements were chosen as the searchable items ; CAS Registry

NUMB	*014	Q000754
FORM	33997-12-5 C5Cl6Hg	Q000754
NAME	Mercury, chloro(1, 2, 3, 4, 5-pentachloro-2, 4-cyclopentadien-1-yl)-;	Q000754
MODF	. alpha. Phase;	Q000754
FREQ	1C 35 77. 38.7 38.44 36.49 36.31 36.22 36.12	2Q000754
FREQ	2C 35 77. 35.61 35.47	2Q000754
AUTH	West R; Wulfsberg G;	1Q000754
JRNL	JACSAT J Amer Chem Soc; 93; 16; 1971; 4085; 4086;	1Q000754
AUTH	Wulfsberg G; West R; Rao V N M;	2Q000754
JRNL	JORCAI J Organomet Chem; 86; 3; 1975; 303; 319;	2Q000754
KYWD	CONJUGATION; BOND NATURE; CHARGE TRANSFER;	Q000754
REMK	Seven lines in the range 37.06-36.26MHz were also reported;	1Q000754
REMK	Intermolecular coordination;	2Q000754
END		Q000754

Fig. 1. Typical record of the Database.

Number, Molecular Formula, Substance Name, NQR Frequency, Author Name, CODEN, Publication Year, and Keyword.

METHOD

1. Outline.

In order to reduce the searching load, the 8 index files (IDF) corresponding to the search items are extracted from DB by the respective preprocessors (See Table 1). The IDF stores the search items (e. g. Author Name) and their respective RNs which point at the address of the items on DB. Searching begins by entering a search key (e. g. Smith J A S). When the system receives the query key from keyboard, it sets a search number

Table 1. Search items, IDFs, and Preprocessors extracting IDFs from Database.

Search items	IDFs	Preprocessors
CAS Registry Number	REG. IDX	Reg. c
Molecular Formula	FOR. IND	For. c
Substance Name	NAM.IND	Nam. c
NQR Faequency	FRE. IND	Fre. c
Author Name	AUT. IND	Aut. c
CODEN	COD. IND	Cod. c
Publication Year	YEA. IND	Yea. c
Keyword	KEY. IND	Key. c

(SET No) at first. Then the system begins to search the key term on the IDF (Author index file in this case) and writes the RNs of the hit item (s) in the user's work file. The RNs thus obtained are sorted in the ascending order and duplicating numbers are eliminated. The set of such RNs are stored in the array whose top address is the pointer which is referred to SET No. The system finally displays the total number of hits on the screen.

2. MAKE_INDEX_FILE

The searching time depends on the search algorithm and the IDF format. We decided on the search method to be used, judging from the property of the search element and the number of the relevant RNs.

2.1 REG. IDX

Most of CAS Registry Numbers uniquely correspond to their respective RNs although some Registry Number is shared by a small number of RNs. Binary search was adopted as the algorithm of searching the element. Therefore, the IDF has a simple format sorted in ascending order of CAS Registry Number.

Example

000074-87-3 ; 000537

000074-88-4 ; 009027

000074-89-5 ; 008704

2.2 FOR. IDX

The relationship of the Molecular Formula and the RN is the same as that of CAS Registry Number. Both the algorithm and file format are the same as for REG. IDX. The record length of the file is normalized to that of the longest Molecular Formula.

In this case, truncations are possible by using \$ marks for masking in a query as in the following examples.

=FOR C10\$ [to retrieve all the C10, and C100 and larger compounds.]

=FOR \$Br\$ [to retrieve all the bromine compounds.]

2.3 NAM. IDX

Three kinds of truncations are possible in a query for Substance Name. When users access to this IDF, most of them will try both hands truncation. In such a case, the total number of hits can be very large. In this case, therefore, it is proper that the IDF is sorted in ascending order of RN and linear searching is adopted as the algorithm because the system can save time for sorting such a large number of RNs after having searched the key name.

2.4 AUT. IDX

There may be many RNs corresponding to one author name. The elements of the IDF are sorted in two directions, that is, the Author Names are sorted in alphabetic order (vertically) and the RNs are sorted in the horizontal direction for each Author Name.

The top position of each Author Name is found using binary search method and fseek () function.

Example

```

ABDULLIN R S ; 001142 002504 003421 005814 006167 006731.....q
ABDULLIN R S ; 011625 011762 011763 011764 011765 011766.....q
ABE Y ; 000528 003363 003608 005631 005632 008356 008708.....q
    
```

2.5 COD. IDX, YEA. IDX, and KEY. IDX

These IDF formats and the search methods are the same as AUT. IDX.

2.6 FRE. IDX

In this case, users need to specify the nuclear mass number (N) and the range of frequency values (the lowest one, F1 and the heighest, F2).

How to find the target data.

- i) The system finds the top address (by N) of the memory region in which the target nuclear data are stored.
- ii) It then moves the file pointer to that address.
- iii) The system judges whether the datum lies between F1 and F2 or not by comparing the top of frequency data with F1 and/or F2.
- iv) If the judgment was negative, it continues to compare F1 and/or F2 with the frequency data sorted in descending order one by one until getting a positive in the judgment.

To do the above procedure efficiently, a special format of index file was devised. The first 11 lines of the IDF record all the top addresses of memory blocks, each storing data of a nuclear species like in the following :

```

2 ; 11 7 ; 617 8 ; 673 9 ; 675 10 ; 679 11 ; 695.....q
    
```

This means that the frequency data of D (mass number 2) starts at the (11+1)-th line. Fig.

2	3	0.1405	0.0766	0.1322	0.1288	0.1184	0.1148	q000387
2	2	0.1122	0.1121	0.1035	0.1035	0.1016	0.1000	q000387
2	1	0.0974	0.0929	0.0904	0.0843			q000387
2	1	0.1871	0.0615	0.1781	0.1208			q000418
.....								
.....								

Fig. 2. Format of frequency data in FRE. IDX.

2. shows the data from the 12-th line.

3. PC_SEARCH

PC_SEARCH consists of DB (ca 9 Mb), 8 IDFs (ca 4 Mb), and search program, nqr ().

3.1 Starting and Exiting from PC_SEARCH

To start PC_SEARCH, users enter "NQR" at the DOS prompt.

To exit from the system, users enter "END" at the system prompt,=.

3.2 Execution of Search () function

Search () is called by a command, "SEAR". Fig. 3 shows the flow chart of search (). The main 18 commands, <?>-<END>, are shown in the left column in Fig. 3.

3.2.1 <?> command (help) lists the brief description of commands.

3.2.2 <KEY>-<FRE> Search commands

Users can use these commands with some operand (s) to get the information of NQR spectral data. To search for data reported by a specific author, users enter the Author Name as a query.

=AUT smith j a s

The system responds to the query like this.

Set No 1 266 Records

Some commands allow truncated query input.

Example of both hands truncation :

Search of all the Molecular Formula having oxygen atom (s).

=FOR \$O\$

Set No 2 5492 Records

Note that this truncation may also retrieve such other elements as Os.

Example of <FRE>

Search of all the frequency data of mass number 17.

=FRE 17

Set No 3 173 Records

A case of specifying frequency range.

=FRE 35 38.2-40.5

Set No 4 841 Records.

When the system received a query, it sets the SET No (set++), opens the corresponding IDF, and searches the terms that hit. When it finds the hit data, it stores the RNs in the work file, ((R+set)->work [i++]). It then displays the total number of hits, and closes the IDF at the end of a search.

3.2.3 <AND>, <OR>, and <NOT> Boolean logic operators

Conventional Boolean operators may be used only between the answer sets.

MODE ?; Type in ? for help

=AND 1 3

Set No 5 23 Records.

3.2.4 <SET> Set command

Users can recall what they have done so far in a search session by entering a command, SET.

MODE ?; Type in ? for help

=SET

Set No 1: AUT SMITH J A S=266 Records.

Set No 2: FOR \$O\$=5492 Records.

Set No 3: FRE 17=173 Records.

Set No 4: FRE 35 38.2-40.5=841 Records.

Set No 5: AND 1 3=23 Records.

3.2.5 <DIS> Display command

The results of a search can be displayed on the screen by specifying the answer set and format. There are predefined 8 different formats, A through H.

Format	To print data for
A	: FOR+NAM+MOD+FRE
B	: FOR+NAM+MOD+FRE+REM
C	: FOR+NAM+MOD+FRE+ATE+JRN
D	: FOR+NAM+MOD+FRE+AUT+JRN+REM
E	: All fields of the record
F	: FOR+NAM
G	: Record numbers only
H	: Registry numbers only

Users can also specify the record number range to display.

MODE ?; Type in ? for help

=DIS 5 D 20-22

3.2.6 <FIL> Opening and closing a log file

Users can open or close a log file at any MODE level, to record all or part of their interaction of the searching. Users can specify two parameters, the file name and the file size in unit of kbyte.

MODE ?; Type in ? for help

=FIL

Do you want to open a history file ?

If you want, please enter the file name,
otherwise hit 'return' only.

=B: June13 200

Having opened a log file, users can close it by entering another FIL command.

MODE ?; Type in ? for help

=FIL

History file, B: June13. hst, is now open.

If you want to close it, please enter 'c',
otherwise hit 'return' key only.

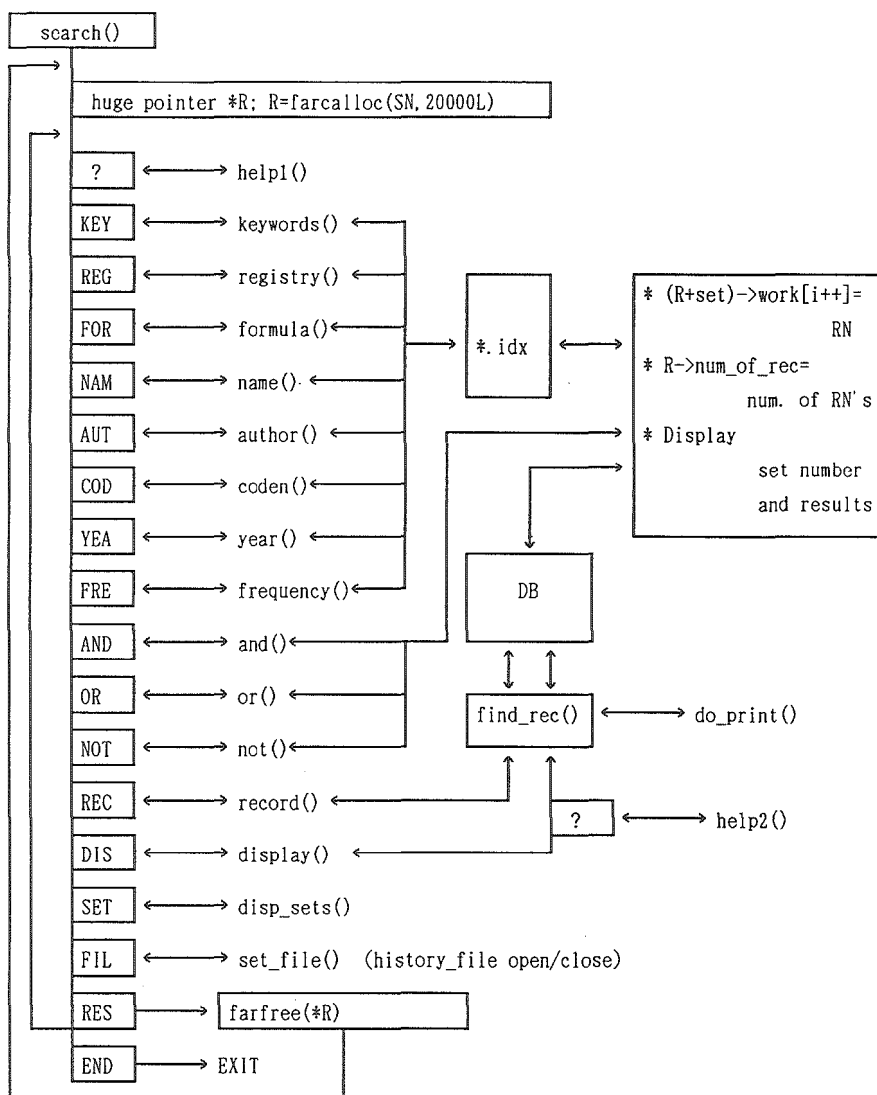


Fig. 3. Flow chart of search() function.

A log file can also be opened at the beginning of the session when users call PC_SEARCH.
 C: ¥> NQR Hb: June16 L50

In this case, "H" and "L" must be added at the top of the file name and its size, respectively. The log file in the name of June16 (size 50 kbyte) on the disk drive B remains open until users close it by using a FIL command or exit the system by entering END command. If the file size is not specified, the system automatically allocates 500 kb.

The system automatically sets the total number of answer sets at 20 unless users specify otherwise. Depending on the memory size of user's work area of the computer, he may want to limit the number to a smaller number; it can be done by using S option

at the beginning of the session as in the following.

C: ¥> NQR S16 [the maximum number of answer sets is now 16.]

The user may input all the parameters in a single line ; H, L, and S in any order :

C: ¥> NQR HB: June16 S16 L100

If the answer set number is exceeded, one cannot conduct any further searching although he can still use DIS, SET, FIL, RES, and END commands.

3.2.7 <RES> Reset command

This command will cancel all the current search session and bring the user back to the initial status. If a log file is kept open, it continues to record all the activities including the cancelled "current" session.

4. SETUP ()

The whole PC_SEARCH system (size of ca 14 Mb) was compressed and changed to the auto-decompressing file, NQRS. EXE, by using an archiver program, LHarc. EXE.³⁾ NQRS. EXE file (size of ca 2.5 Mb) was stored in 4 sheets of 2DD mini-floppy disks using a DOS command, BACKUP.

SETUP () creates a directory, <NQR>, in the hard disk drive specified by the user, re-stores NQRS. EXE in the <NQR>, and executes NQRS. EXE to expand into the original PC_SEARCH system.

CONCLUSION

All the source codes were compiled on NEC's PC9801/RA21. PC_SEARCH ran successfully on NEC's PC9801/RA21, IBM's PS/5530Z, PS/55note, PS/2, and several IBM compatible machines. The searching time is typically within 15 seconds under the condition of NEC's PC9801/RA21 and ICM HC130 (130 Mb hard disk, average seek time is 19 ms).

REFERENCES

1. H. Chihara, J. Mol. Struct., 83, 1-7 (1982).
2. H. Chihara and N. Nakamura, "Nuclear Quadrupole Resonance Spectroscopy data", LANDOLT-BÖRNSTEIN New series Group III vol. 20a, 20b, 20c, Springer-Verlag Berlin 1988.
3. H. Yoshizaki, C Magazine, vol. 3, No. 1, 59-68, (1991).