(10) Papier, L., "Reliability of Scientists in Supplying Titles: Implications for Permutation Indexing," *Aslib Proc.* **15**, 333-7 (1963).
(11) Resnick, A., "Relative Effectiveness of Document Titles and Abstracts for Determining Relevance of Documents," *Science* **134**, 1004-6 (1961).
(12) Saracevic, T., "Comparative Effects of Titles, Abstracts and Full Texts on Relevance Judgments," *Proc. Amer. Soc. Inform. Sci.* **6**, 293-9 (1969).
(13) Smith, P. P., "Increasing Recall of Keyword Indexes by Automatically Indexing Root Character Strings," *Ibid.*, **7**, 275-7 (1970).
(14) Tell, B. V., "Document Representation and Indexer Consistency; A Study of Indexing from Titles, Abstracts and Full Text using UDC and Keywords," *Ibid.*, **6**, 285-92 (1969).
(15) Tocatlian, J. J., "Are Titles of Chemical Papers Becoming More Informative?," *J. Amer. Soc. Inform. Sci.* **21**(5), 345-50 (1970).
(16) Windsor, D. A., "A Format for the Standardized Citation of Both Published and Unpublished Documents," *Ibid.*, **21**(3), 195-7 (1970).
(17) Zabriskie, K. H., and Farren, A., "The B. A. S. I. C. Index to *Biological Abstracts*," *Amer. J. Pharm. Ed.* **32**, 189-200 (1968).

Todai Scientific Information Retrieval (TSIR-1) System. I. Generation, Updating, and Listing of a Scientific Literature Data Base by Conversational Input

TAKEO YAMAMOTO,* TOMOKO KUMAI, KEN-ICHI NAKANO, CHIKATAMI IKEDA, TOSIYASU L. KUNII, HIDETOSI TAKAHASI, and SHIZUO FUJIWARA

The University of Tokyo, Hongo, Tokyo, Japan

Received August 30, 1971

A file structure for scientific literature data, STF, to be used in a scientific information retrieval system, TSIR-1, was developed by modifying the CAS Standard Distribution Format (SDF). The use of STF allows one to merge data records originating in CAS and those generated locally. A set of programs for the generation, updating, and listing of a scientific literature data base in STF from a TSS terminal in conversational mode is described.

For a scientist who wants to accumulate a data base of scientific literature for his own use and for the use of his scientific community, the CAS SDF[1, 2, 3] file structure has several attractive characteristics. It is a highly flexible format and is largely based on the natural language in its expression of the content of the literature. However, it still leaves much to be desired for generating new data records or for data records obtained from other sources than CAS as input to the data base.

We have been building a scientific information retrieval system (TSIR-1) using a HITAC 5020 TSS[4, 5, 6] of the Computer Centre, University of Tokyo. The TSIR-1 system is based on a data structure named STF (for *S*implified *T*odai *F*ormat, Todai being the abbreviated Japanese word for the University of Tokyo). STF is closely related to SDF, and its logical content is similar to that of SDF except for a number of modifications. Although more detailed description of STF will be published elsewhere, significant modifications follow.

In STF, all keywords corresponding to an article item are combined into one record element, and the element is included in the same logical record as the other data related to the item. Thus, each logical record corresponds to one article except for the case when the data are too great for one logical record (maximum record length: 3520 characters).

The text of the abstract and the full text of an article are given their own identification (ID) numbers and may be recorded in STF files.

Both Japanese (Romanized) and English entries are allowed for natural language input. The search programs and queries are expected to take care of the occurrence of the two languages in the retrieval phase.

The SERIAL NUMBER data element, ID number 0012 01, and the Temporary Abstract Number (TAN) data element, ID number 0054 01, in SDF[2] are modified to allow the use of these data elements for the file key in STF. The modification allows one to distinguish, if necessary, between the locally generated data records and records originating in CAS SDF files, after they have been merged into a unified data base. Thus, the compatibility as well as the distinguishability of the two kinds of records are obtained at the same time.

Only capital letters are used in the alphabet. This restriction was necessary because our input-output facilities do not include the lower-case alphabet.

In the rest of the present paper, a set of programs for the generation, updating, and listing of STF data as TSS disc files by conversational input will be described. The programs were written to help TSS users to generate and maintain their own literature files, without having a detailed knowledge of SDF or of STF.

The programs were written in PL/IW language,[5, 6] a subset of PL/I. They were compiled and executed by the HITAC 5020 TSS.

*To whom correspondence should be addressed at Department of Chemistry, Faculty of Science, the University of Tokyo, Hongo, Tokyo 113, Japan.

GENERATION OF THE FILE

The execution record of the program, GENERATE-STF, is given in Figures 1 and 2. In these figures as well as in subsequent ones, information input by the user is underlined.

In Figure 1, the system asks the format of the file. First, the number of authors in the first article item is asked. Then, the presence of several elements likely to appear in a typical scientific record is asked. The user replies by typing in 1 or 0 for the affirmative or the negative answer, respectively. Next, the system proceeds to ask if there are other elements the user wants. For these additional elements only the user is expected to supply the system with the corresponding ID numbers and modifiers applicable to STF.

In Figure 2, the system requests the input of data elements according to the format determined above. For data elements which are expected to be shorter than 68 characters (one input line), the system assumes that an input data element ends when the 'return' key or the 'line feed' key is pressed. For other data elements, 'equal' key followed by one of the above keys is assumed to be the end of the input string. The character count is taken, and the input is edited, if necessary, to form an STF data element. Finally, data which did not appear in the specification of the format, ABSTRACT NUMBER, is requested. This must be a four-character string which is uniquely assigned to the logical record in the user's file system—that is, not only in the present file but in all the files which are expected to be used together in a subsequent retrieval operation. The input data are transformed into the STF SERIAL NUMBER, ID number 0012 01, and into the STF TAN, ID number 0054 01, by connecting with strings 'XX00,' and 'XX000,' respectively.

The above data are used to form data element tags and data elements, and these are combined to form an STF logical record in a buffer area in the CPU. The remaining size of the buffer area is notified to the user in the LINE (80 characters) units. Then, it is asked whether more data input is desired. If the user answers in the affirmative, the system asks the number of authors of the next article item. As to the other kinds of data elements, the system assumes that the same format applies as in the first logical record.

```
$RUN(GENERATE-STF)



INPUT RECORD FORMAT
HOW MANY AUTHORS?
••• 3

KEYWORDS,1 OR 0?
••• 0

ARTICLE TITLE,1 OR 0?
••• 1

JOURNAL NAME,1 OR 0?
••• 1

CODEN,1 OR 0?
••• 0

VOLUME NUMBER,1 OR 0?
••• 1

PAGE,1 OR 0?
••• 1

YEAR,1 OR 0?
••• 1

ANY OTHER CATEGORY,1 OR 0?
••• 1

INPUT IDNO IN 4 DIGITS
••• 4096
INPUT IDMOD IN 2 DIGITS
••• 01

ANY OTHER CATEGORY,1 OR 0?
••• 0
```

Figure 1. The conversational generation of an STF file
The determination of the input format

```
INPUT AUTHOR NAMES,ONE BY ONE
••• FUJIWARA S

••• AOYAGI K

••• MIYAMAE T

INPUT TITLE,INPUT =-SIGN AT THE END

••• NUCLEIDIC MASS MEASUREMENT BY ION CYCLOTHON RESONANCE=

INPUT JOURNALNAME
••• BULL. CHEM. SOC. JAPAN

INPUT VOLUME NUMBER
••• 43

INPUT PAGE
••• 561

INPUT YEAR
••• 1970

INPUT DATA ELEMENT,IDNO= 4096---=-SIGN AT THE END

••• A NEW METHOD FOR MEASURING THE NUCLEAR MASS BY ION CYCLOTRON

••• RESONANCE SPECTROSCOPY IS DESCRIBED. THE METHOD IS APPLIED

••• TO THE DETERMINATION OF THE NUCLEAR MASS OF ARGON-40.

••• THE ACCURACY OF THE METHOD AND POSSIBLE SOURCES OF ERROR ARE

••• DISCUSSED.=

INPUT ABSTRACT NUMBER IN 4 CHARACTERS
••• 0027

LINES LEFT= 38
INPUT 1 IF NEXT DATA INPUT
INPUT 0 IF INPUT END AND FILE REGISTRATION WANTED

••• 0

INPUT FILENAME IN 4 DIGITS
••• BCSP
FILENAME= FILEBCSP

CPUTIME 005.85EC
```

Figure 2. The conversational generation of an STF file
The input of the content and the registration of the file

Then, input of the content is repeated as in the first item. (The messages requesting the input of data elements are typed out in abbreviated forms for the second time on.) If, on the other hand, the user answers in the negative, the system proceeds to write the disc file from the buffer area. The name of the file is generated by connecting the four characters given by the user with a fixed string, 'FILE'. The resultant eight-character name satisfies, as long as it is unique in the user's STF file system, all the necessary conditions for an STF file name. As the user needs to know the file name for future access to the file, he is notified of the resultant file name.

UPDATING THE FILE

STF files thus generated can not be updated by the usual UPDATE command[7] for a HITAC 5020 TSS file, because of their compact and complex structure. A program was written which can be used by the user to cancel one logical record, to replace one logical record for a new one, or to extend the file by several logical records. The execution record is given in Figure 3 for the replacement. First, the name of the file to be updated is requested. Then, the ABSTRACT NUMBER of the logical record to be updated is requested. As this is input in four characters and is processed in the system to generate an STF SERIAL NUMBER data element applicable only for the locally generated records, the user's updating a record originating in a CAS SDF file is blocked. For file extension, a dummy number '0000' must be given as the ABSTRACT NUMBER. The system then reads the appropriate STF file in from the TSS

```
$RUN(UPDATE-STF)
INPUT STF DISC FILE NAME

... FILEBCSF

INPUT RECORD SERIES NUMBER IN 4 DIGITS
INPUT 0000 IN 4 DIGITS FOR DATA EXTENSION
... 0027
INPUT 1 FOR DATA REPLACEMENT
INPUT 0 FOR DATA CANCELLASION
INPUT -1 FOR DATA EXTENSION

... 1

INPUT RECORD FORMAT
HOW MANY AUTHORS?
... 3

KEYWORDS,1 OR 0?
... 1

ARTICLE TITLE,1 OR 0?
... 1

JOURNAL NAME,1 OR 0?
... 0

CODEN,1 OR 0?
... 1

VOLUME NUMBER,1 OR 0?
... 1

PAGE,1 OR 0?
... 1

YEAR,1 OR 0?
... 1

ANY OTHER CATEGORY,1 OR 0?
... 0

INPUT KEYWORDS SEPARATED BY COMMAS.INPUT =-SIGN AT THE END

... ION CYCLOTRON RESONANCE SPECTROSCOPY, ARGON-40,

... ACCURACY=

INPUT AUTHOR NAMES,ONE BY ONE
... FUJIWARA S

... AOYAGI K

... MIYAMAE T

INPUT TITLE.INPUT =-SIGN AT THE END

... NUCLEIDIC MASS MEASUREMENT BY ION CYCLOTRON RESONANCE=

INPUT CODEN
... BCSJA8

INPUT VOLUME NUMBER
... 43

INPUT PAGE
... 561

INPUT YEAR
... 1970

INPUT ABSTRACT NUMBER IN 4 CHARACTERS

... 0027

LINES LEFT= 41
LAST RECORD DIRECTORY
- 1- RECORD IN FILE
LINES USED= 4
INPUT FILENAME IN 4 DIGITS
... BCSJ
FILENAME=FILEBCSJ
PLEASE DELETE- FILEBCSF -
CPUTIME 008.4SEC
```

Figure 3. The updating of an STF file record

```
$RUN(PRINT-STF)
INPUT STF DISC FILE NAME

... FILEBCSF

INPUT SERIES NUMBER IN 4 DIGITS

... 0027
00597001  0000A   FUJIWARA S
00597002  00008   AOYAGI K
00597003  00009   MIYAMAE T
005B7001  00035   NUCLEIDIC MASS MEASUREMENT BY ION CYCLOTRON RESONANC
                  E
005D7001  00016   BULL. CHEM. SOC. JAPAN
005F7201  00002   43
00617301  00003   561
005E7001  00006   000070
10007001  000F7   A NEW METHOD FOR MEASURING THE NUCLEAR MASS BY ION C
                  YCLOTRON RESONANCE SPECTROSCOPY IS DESCRIBED.  THE M
                  ETHOD IS APPLIED TO THE DETERMINATION OF THE NUCLEAR
                   MASS OF ARGON-40.  THE ACCURACY OF THE METHOD AND P
                  OSSIBLE SOURCES OF ERROR ARE DISCUSSED.
00127001  00008   XX000027
00547001  00009   XX0000027
INPUT 1 IF NEXT SERIES NO INPUT
INPUT 0 IF INPUT END
... 0

CPUTIME 002.7SEC


$RUN(PRINT-STF)
INPUT STF DISC FILE NAME

... FILEBCSJ

INPUT SERIES NUMBER IN 4 DIGITS

... 0027
00777001  00038   ION CYCLOTRON RESONANCE SPECTROSCOPY, ARGON-40, ACCU
                  RACY
00597001  0000A   FUJIWARA S
00597002  00008   AOYAGI K
00597003  00009   MIYAMAE T
005B7001  00035   NUCLEIDIC MASS MEASUREMENT BY ION CYCLOTRON RESONANC
                  E
00557001  00006   BCSJA8
005F7201  00002   43
00617301  00003   561
005E7001  00006   000070
00127001  00008   XX000027
00547001  00009   XX0000027
INPUT 1 IF NEXT SERIES NO INPUT
INPUT 0 IF INPUT END
... 0

CPUTIME 002.6SEC
```

Figure 4. The listing of the STF records before and after the updating

file. It then requests the user to specify the kind of work to be done: replacement, cancellation, or extension. The old file is then divided into logical records, the records are searched for the SERIES NUMBER record element, and the value of the element is compared with the one given by the user. Appropriate work as specified by the user is performed on the logical record thus accessed. The updated records are placed in a buffer area, and finally it is filed under a new file name generated by the same technique as described before. The new name is given to the user, and the system advises the user to delete the old file as it will otherwise be left in the TSS file.

LISTING OF THE FILE

Because of the complex structure of an STF file, a usual file PRINT command[7] yields an unintelligible terminal output. A program whose execution record is shown in Figure 4 asks the user to give the STF file name and the ABSTRACT NUMBER of the logical record to be displayed at the terminal. The STF file is read in, the appropriate logical record is assessed, and its record elements are displayed together with the data element tag and the total character counts.

DISCUSSION

Any local community of scientists should have its own data base on which a retrospective search may be performed by the member, preferably on an on-line basis. By 'local community' is meant such groups as a professional society in a country, a regional branch of the society, a college, a department of a university, and a research group of active scientists. The bulk of the data base may be generated conveniently by retrieving relevant data records by the current awareness mode of operation (generation of subfiles instead of bibliographic listings is needed) from some large-scale data bases such as CAS and MEDLARS tapes; however, some means of the members' supplementing the data base with their own data records must be available if it is going to be the group's central data base. The members cannot, and should not be expected to, know all of the intricate rules and techniques which enables the system to manage the inhomogeneous files thus generated. The system has to assist the users with a set of commands which:

Outputs the necessary instructions for the user
Edits the input data into forms suitable for storage
Protects the user against coincidental errors such as giv-

ing the same TAN to a private record as is used in a CAS file,[2] or giving an inappropriate file name

The present system, TSIR-1, is designed to be such a system for the scientific community at the University of Tokyo. In the system, STF records are generated by the on-line input described above, by a tape-to-tape conversion process of SDF tapes, and by the retrieval of large-scale STF files on the 'current awareness' basis. The programs described above were developed for use as commands in the system. As a typical Japanese user is not expected to be a good typist, the programs were made to work with as simple input from the user as is compatible with flexibility of its functions. For the same reasons, numbers '1' and '0', rather than 'YES', 'NO' or 'Y', 'N', were requested as the affirmative and the negative responses from the user.

The programs have been successfully used by several TSS users as their private programs, and the data files thus generated within their TSS files have been used for on-line generation of KWIC indexes of both English and Japanese title data elements. The programs will be registered as commands in the system in the near future.

ACKNOWLEDGMENT

The stimulating discussions of Haruo Hosoya and other members of the Chemical Information Seminar held at the Department of Chemistry, University of Tokyo, are gratefully acknowledged. Part of the work has been supported by a fund from the Ministry of Education of Japan, Tokutei-Kenkyu I, Showa 45, No. 99042.

LITERATURE CITED

(1) Chemical Abstracts Service, "Standard Distribution Format Technical Specifications," Columbus, Ohio, 1970.
(2) Chemical Abstracts Service, "Data Content Specifications for CA Condensates in Standard Distribution Format," Columbus, Ohio, 1970.
(3) Anzelmo, F. D., "A Data Storage Format for Information System Files," *IEEE Trans. Comput.* **C-20**(1), 39 (1971).
(4) Motobayashi, S., Masuda, T., and Takahashi, N., "The HITAC 5020 Time Sharing System," *Proc. ACM 24th Nat. Conf.*, p. 419, 1969.
(5) Nakata, I., and Hamada, H., "HITAC 5020 PL/1 Compiler," *Hitachi Rev.* **17**(11), 432 (1968).
(6) Central Research Laboratory, Hitachi Ltd., "PL/IW Gengo no Shiyo (Specifications of PL/IW Language)," 3rd ed., Tokyo, Japan, Jan. 1969 (in Japanese).
(7) Takahasi, H., Kunii, T. L., Saiga, N., and Nakano, K., "Nijigen Banchizuke Hoshiki ni Yoru 5020 TSS no Gaiyo (An Outline of 5020 TSS with Two-Dimensional Addressing)," *Center News* (The Computer Centre, Univ. of Tokyo) **2**(4), p. 6 (1970) (in Japanese).

Reference Literature to Thermodynamic Diagrams

A. L. HORVATH
18, Harlow Close, Thelwall, Nr. Warrington, England

Received March 21, 1971

A review of the sources of published thermodynamic diagrams which are frequently used by engineers in research and industry is presented.

Thermodynamic diagrams, such as pressure-enthalpy (p-H), temperature-entropy (T-S), enthalpy-entropy (H-S), volume-enthalpy (v-H), and temperature-enthalpy (T-H) are frequently used by engineers in design calculations, particularly for compressors, refrigerators, and power cycles (Rankine, Carnot, etc.). Several diagrams are available in various textbooks; however, a review of the published charts will assist engineers to a quick and easy selection according to the requirements. This article is based mainly on secondary sources (textbooks), which are available in most technical libraries. In most cases, the original diagram is presented in an enlarged form with an accordingly higher accuracy, which is always required by these types of calculations.

Tables I and II list single inorganic and organic substances arranged in alphabetical order by chemical formula, thus providing a quick method for finding a given compound in the tabulations. The numbers refer to the

Table I. Inorganic Compounds

Formula	Name	Refrigerant No.	References	Formula	Name	Refrigerant No.	References
A	Argon	740	1b,7b,18b,23c,36ace	Hg	Mercury	900	23c
—	Air	729	1b,7b,18b,23b,24ce,25a, 26a,36ace	K	Potassium	739	23c
				N_2	Nitrogen	728	1b,5a,7b,18b,23b,36ace, 39a
CO	Carbon monoxide	728A	5a,7b,13b,18ab,23c				
CO_2	Carbon dioxide	744	1a,5a,7b,18b,19a,23bc, 24a,25a,26a,37abc	NH_3	Ammonia	717	1a,5a,7b,9a,14a,18b,19a, 24a,25a,26a,39a
Cl_2	Chlorine	771	15ab,18a,28b	N_2O	Nitrous oxide	744A	23c,24a,25a,26a
H_2	Hydrogen (normal)	702n	1b,5a,18b,23bc	Na	Sodium	723	23c
H_2	Hydrogen (para)	702p	1bc,30b	Ne	Neon	720	1b
H_2O	Water	718	5a,9d,16c,17t,18bc,23c, 24c,26bc,27t,31ct,33t	O_2	Oxygen	732	1b,5a,18a,36ace
				SF_6	Sulfur hexafluoride	846	18ab
He	Helium	704	1bc,18bc,23b,39a	SO_2	Sulfur dioxide	764	5a,24a,25a,26a