LETTERS TO THE EDITOR

Dear Sir:

We should like to comment on the article "A Qualitative Comparison of WLN with RINGDOC" by Mitsuo Sasamoto, *et al.*, in *J. Chem. Doc.*, **13**, 206–11 (1973).

Whether a retrieval system can be considered satisfactory depends on several aspects (see, *e.g.*, Lancaster[1]), and we found it somewhat disappointing that in this paper some of these aspects were not considered.

1. Usually in a paper both recall and precision are estimated simultaneously. Although recall is often difficult to ascertain, this is easier for internal files, because these are usually smaller and better defined than external files. At least a comparison of the relative recall should have been made.

2. It is mentioned that the study concerned only the internal WLN and RINGCODE files, and we may assume that they are identical with respect to the chemical compounds. However, it is not clear whether all kinds of compounds are present. If some types of structures are missing or if there is a preference for certain types (not unusual in internal files), both the formulation of the search logic and the relevance of the answers would be affected.

3. Comparison of two systems implies optimal use of the possibilities of both systems. This rule has not been observed in the present case. We will demonstrate this on a few example questions that have a poor logic.

(a) Question 2a. In the WLN logic the dialdehyde and some of the monoaldehydes are excluded (VH4V. .). In the RINGCODE logic, 11/2 instead of 11/1 should have been used, or even better 11/2(OR)11/3(OR)11/4; also apart from 23/1, other punches in column 23 are relevant to the substructure given.

(b) In question 2b the RINGCODE logic excludes all substituents other than alkyl or ring carbon, and also methyl ether. The WLN-logic omits . .YQ1M. ., etc. These restrictions do not follow from the substructure shown.

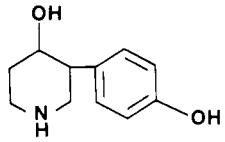
(c) *Re* question 3a: most of the problems encountered in the WLN search could easily have been solved by a better search logic. If we understand correctly the possibilities of the search logic, then the formulation (NOT)T(FB)MNV(FB)J and (NOT)T(FB)VNM(FB)J would have excluded most of the compounds of type a-1, and T(FB)N(FB)J*VM(OR)MV-*T(FB)N would have included the wanted a-2 type, plus, admittedly, a lot of noise. The RINGCODE logic as stated requires at least one carbon substituent, which is not indicated by the substructure; furthermore, the search logic used does not exclude carbamates, formamides, etc., which result from the overcoding used in external files. However, in internal files overcoding is seldom used and then these compounds can be excluded.

(d) *Re* question 3b: the RINGCODE logic excludes benzene rings as substituents on the benzene ring; 12/3 and 12/12(OR)12/2 added to the search logic would have raised the relevance.

(e) In the RINGCODE, 3/11 differentiates between phenazines and benzoquinoxalines. It was brought to our attention by Dr. Sasamoto that this would exclude systems substituted by naphthalene. This is true, but these could be included by another question, asking for (2/2 OR 2/3) AND 2/0 AND 2/11, etc.

(f) *Re* question 4a: in the RINGCODE logic benzene rings are excluded as substituents; the use of 10/0 and 10/5 would have been more appropriate than the use of 10/2 or 10/4 or 10/7 and could have raised the relevance.

(g) *Re* question 4b: if we assume that other substituents are allowed then the compound



will not be found with this RINGCODE logic. If other substituents are not allowed then, for instance, structures 4b-1 and 4b-2 can be excluded by using NOT 2/11, 2/0, 2/1, 2/2, 2/3 instead of AND 6/12.

(h) Compound 5a-2 should not have been found in the RINGCODE search using the stated logic because the enolization rule does not apply (two benzene rings and one alicyclic ring).

4. It was rather confusing to find some erroneously used WLN characters, *e.g.*, I instead of 1, spaces where they do not belong, i instead of I, etc. Also we should have preferred the word RINGCODE instead of RINGDOC in the title.

M. Osinga and A. van der Woude

GIST-BROCADES N.V.
Research & Development Department
Haarlem, Netherlands

Dear Sir:

We read Drs. Osinga and van der Woude's letter with much interest. Their comments were very instructive. This publication was our first experience to contribute in a foreign language. So we asked Dr. W. J. Wiswesser to revise our draft. We now feel the contribution could not fully discribe our intention owing to our poor knowledge of English.

We should like to answer their comments as follows.

1. As they pointed out, we must consider an omission as well as noise in information retrieval. However, we think that recall is almost 100% for both WLN and Ringcode in our internal files. Therefore, we represented the percentage for only relevance.

2. As they pointed out, our studies on WLN and Ringcode were done on a limited number of compounds in our internal files, but our experimental aim was to compare WLN with Ringcode. So we believe that there was no problem about the scope of our file.

3. (a) Ringcode: If all cases of carbonyl groups are considered, their logic is correct. Though our explanation was insufficient, our logic considered the ketones as an example. Also, the 0 of the V024 was a misprint; we should have used V 24.

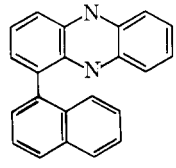
WLN: We did not consider the aldehyde compounds because of our limited number of compounds. Coding of punches 11/1 was a mistake. Also the reason we coded only 23/1 is as above mentioned.

(b) In WLN and Ringcode we considered the substituents of both NH– and O– to be alkyl; our explanation was insufficient. Generally, their comments are right.

(c) We agree. We retrieved with minimal search logic.

(d) This logic was set for alkyl groups and not for aralkyl groups.

(e) Punches 3/11 should not be "NOT" logic. If it were, we would not be able to retrieve the following compound:



We try not to use the "NOT" logic as much as possible for prevention of omissions.

(1) Lancaster, P. W., "Evaluation of the Medlars Demand Search Service," National Library Medicine, Washington, D. C., 1968.

(f) The details were insufficient. We put a limitation to some uncomplicated substituents other than benzene.

(g) In this case, we searched for piperidine ring having only one hydroxy, so that compound met our requirement. Their coding NOT 2/11, 2/0, 2/1, 2/2, 2/3, instead of 6/12, seems to be appropriate. We think it was enough to specify only punches 6/12 without a number of rings. If we specified "NOT" logic, we would have had to take account of all unnecessary cases. So the use of "NOT" logic must be minimized. We believe that punches 6/12 were suitable for this case.

Mitsuo Sasamoto

Information Center
Tanabe Seiyaku Co., Ltd.
Saitama, Japan

J. H. KUNEY APPOINTED NASA FACILITY DIRECTOR

In September, Joseph H. Kuney was appointed Director of the NASA Scientific and Technical Information Facility in College Park, Maryland. Operated for the NASA Scientific and Technical Information Office by Informatics Information Systems Company, the Facility furnishes data processing, information services, publishing, and other support for the NASA Scientific and Technical Information System.

Immediately prior to joining Informatics, Mr. Kuney was Manager of the Journal and Encyclopedia Department of Wiley-Interscience in New York. Prior to that, he served for more than 25 years in a variety of assignments with the publishing programs of the American Chemical Society.

NSF GRANT

Professor Martha E. Williams, Director of the Information Retrieval Research program at the University of Illinois, Urbana, Ill., has recently been awarded a $100,300 research grant from the National Science Foundation, Office of Science Information Service. The project, entitled "Data Base Mapping Model and Search Scheme to Facilitate Resource Sharing," will lay some of the ground work necessary for networking of machine-readable bibliographic and numeric data bases. A preliminary task will involve generating a "Data Base of Data Bases." The data base of data bases will be maintained on-line at the University of Illinois and will contain information about data bases such as originating organization, subject coverage, data element content, format, names of centers that process the data base, etc.

The purpose of this project is to test the feasibility of building linkages among data bases. These interconnections will be expressed as a data base mapping model. The model will be based on the relationships that exist between available data bases, conversion tables and algorithms, the information centers that are processing data bases, search software systems, and data base standards. The resulting map will contain descriptive data about bibliographic data bases as well as data-type data bases. A data base mapping system will permit identification of optimal and potential routes from data base to data base using intervening conversion tables or algorithms as needed.

NEW EDITION OF CASSI

Chemical Abstracts Service will issue a new cumulative edition of the CAS *Source Index* (CASSI) early in 1975. The new edition updates the information published in the 1969 edition and incorporates the changes and new listings contained in the CASSI supplements published from 1970 through 1974.

Compiled in cooperation with 399 libraries in 28 nations, CASSI is a comprehensive key to the source literature of the chemical sciences. The new cumulative edition provides complete bibliographic descriptions of some 24,000 serial titles and 10,700 monographs and conference proceedings volumes and indicates which of the participating libraries hold copies of these publications. Included are titles covered by *Chemical Abstracts* since 1907, additional titles covered by *Chemisches Zentralblatt* from 1830 to 1940 and Beilstein's *Handbuch der organischen Chemie* prior to 1907, and some 700 journals currently monitored by BioSciences Information Service of Biological Abstracts. Records for over 14,000 currently published journals have been completely updated for the new edition, and nearly 100,000 new library file locations have been added since the 1969 compilation.

Bibliographic data in CASSI include complete titles in the original language of publication along with English translations if the titles are other than English, French, German, or Spanish, International Standard Organization's standard abbreviations and ASTM CODEN for the titles, titles catalogued according to standard library practice, information on language, history and frequency of publication, and, for currently published titles, the names and addresses of publishers or sales agents. The new edition also indicates which of the listed journals are monitored by *Biological Abstracts* and *Engineering Index* and identifies journals included in the Original Article Tear Sheet service of the Institute for Scientific Information. Additional sections contain bibliographic and library holdings data on patents monitored by Chemical Abstracts Service, directories of participating libraries and publishers and sales agents, and a list of selected directories of microfilm publications.

Supplementary listings of new additions to or changes in the CASSI data base will continue to be issued quarterly as a subscription service. The first quarterly supplement to the new cumulative edition will appear in April 1975.

The new edition of the CAS *Source Index* also is available in computer-readable form. This service consists of a master file on magnetic tape, with new information and changes to previous data provided four times yearly in computer-readable form.

CAS ASSUMES CODEN RESPONSIBILITY

Effective January 1, 1975, Chemical Abstracts Service assumed responsibility for community-wide assignment and dissemination of ASTM CODEN for periodical and nonperiodical publications. The CODEN system was transferred to CAS by the American Society for Testing and Materials, which sponsored the system since 1955.

CODEN, which are used in lieu of full or abbreviated titles of publications in processing and storing bibliographic data in many computer-based information handling systems, are unique, unambiguous codes that identify publications. Each CODEN consists of five characters plus a check character that permits computer detection of errors in the code. Once assigned, a CODEN serves as a permanent identifier for a specific title. If the title of a serial publication is changed, a new CODEN is assigned. The original CODEN continues to identify the earlier title.

Requests for assignment of CODEN or inquiries about CODEN should be directed, after January 1, 1975, to:

Chemical Abstracts Service
CODEN Section—Department 32
The Ohio State University
Columbus, Ohio 43210