which published, at the end of 1972, a "Laboratory Guide" which included a list of new books on the subject of the environment.

One other service, the Ecology Placement Service, may be of interest. It publishes a monthly newsletter[33] listing jobs and graduate research opportunities in the environmental area.

BIBLIOGRAPHY

(1) Siehl, George H., "Our World—and Welcome to It! An Introduction to the Literature of the Environmental Crisis and Related Problems," *Library J.* **95,** 1443-47, April 15, 1970.
(2) "Ulrich's International Periodicals Directory, A Classified Guide to a Selected List of Current Periodicals, Foreign and Domestic," 12th ed., Vol. **1,** Scientific, Technical and Medical, N. Y., Bowker, 1967, pp. 54-5.
(3) "Ulrich's International Periodicals Directory, A Classified Guide to Current Periodicals, Foreign and Domestic," 14th ed., 1971-72, Vol. **1.** N. Y., Bowker, 1971, pp. 633-8.
(4) *Chemical Abstracts, Fifth Decennial Index,* Vol. **41-50,** 1947-56, Subjects, Washington, D. C., American Chemical Society, 1960.
(5) *Chemical Abstracts, Eighth Collective Index,* Vol. **66-75,** 1967-71, Subjects, Washington, D. C., American Chemical Society, 1972.
(6) American Chemical Society, Committee on Chemistry and Public Affairs, Subcommittee on Environmental Improvement, "Cleaning our Environment, the Chemical Basis for Action, A Report" Washington, D. C., Amer. Chem. Soc., 1969.
(7) *Environment Reporter,* Washington, D. C., Bureau of National Affairs.
(8) *Environmental Science Citation Index,* N. Y., Environment Information Center.
(9) *Toxicon,* Rockville, Md., Informatics, Inc.
(10) *Environ,* organized for the U. S. Environmental Protection Agency, Rockville, Md., Informatics, Inc.
(11) *Pollution Abstracts,* La Jolla, Calif.
(12) *Environment Information Access,* N. Y., Environmental Information Center.
(13) *Environment Index,* N. Y., Environment Information Center.
(14) *Environmental Patent Abstracts,* Princeton, N. Y., Princeton Microfilm Corp., Book Division.
(15) *Environmental Science and Technology,* Washington, D. C., American Chemical Society.
(16) *Chemosphere,* Elmsford, N. Y., Pergamon Press, Journals Division.
(17) *Critical Reviews in Environmental Control,* Cleveland Ohio, Chemical Rubber Co.
(18) *Science of the Total Environment,* Amsterdam, Holland, Elsevier Pub. Co.
(19) *Journal of Environmental Sciences,* Mt. Prospect, Ill., Institute of Environmental Sciences.
(20) "Handbook of Environmental Control," Cleveland, Ohio, Chemical Rubber Co.
(21) "Industrial Pollution Control Handbook," N. Y., McGraw Hill, 1971.
(22) "Environmental Measurement and Interpretation," N. Y., Van Nostrand Reinhold, 1972.
(23) "Pollution Analyzing and Monitoring Instruments," Park Ridge, N. J., Noyes Data Corp., 1972.
(24) McCoy, J. W., "Chemical Analysis of Industrial Water," Oxford, England, Robert Maxwell, 1970.
(25) "Chemical Markets in Water and Wastewater," N. Y., Environmental Science Services, 1972.
(26) "Important Chemical Reactions in Air Pollution Control," N. Y., American Institute of Chemical Engineers, 1971.
(27) "Environmental Control in the Organic and Petrochemical Industries," Park Ridge, N. J., Noyes Data Corp., 1972.
(28) "Environmental Control in the Inorganic Chemical Industry," Park Ridge, N. J., Noyes Data Corp., 1972.
(29) "Pollution Control Companies, U. S. A., 1972," Park Ridge, N. J., Noyes Data Corp., 1972.
(30) National Foundation for Environmental Controls," Directory of Environmental Information Sources," Boston, Cahners Pub. Co., 1972.
(31) Schildhauer, Carole, Comp., "Environmental Information Sources; Engineering and Industrial Applications, A Selected Annotated Bibliography," N. Y., Special Libraries Assn., 1972.
(32) "1972-73 Laboratory Guide," *Analytical Chemistry,* **44,** August, 1972, pp. 339LG-369LG.
(33) *Placement Bulletin,* St. Paul, Minn., Ecology Placement Service.

Evaluation of an SDI Service Based on the *Index Chemicus Registry System*

PETER LEGGATE,* B. N. ROSSITER[a] and J. F. B. ROWLAND[b]
Experimental Information Unit, University of Oxford, 7 Keble Road, Oxford OX1 3QL, U.K.

Received August 20, 1973

The *Index Chemicus Registry System* (ICRS) is the machine-readable equivalent of *Current Abstracts in Chemistry & Index Chemicus* (CAC & IC). In an earlier paper,[1] we described the development of an experimental selective-dissemination-of-information (SDI) service based on these tapes. A detailed description of the techniques of profile construction for searching a Wiswesser Line Notation (WLN) structure file is given in this earlier paper. The present paper describes the evaluation of the SDI service in terms both of quantitative measures of retrieval performance, coverage and currency, and also of user reactions to the service, as expressed in their replies to a questionnaire. Failure analysis techniques[2] were used to identify the reasons for retrieval failures and possible methods for improving retrieval performance. A fuller description of the evaluation has been published in report form.[3] More detailed discussions of the evaluation methodology, with particular reference to the nature of user relevance decisions, the methods for measuring recall, and

* To whom correspondence should be addressed.
[a] Present address: Computing Laboratory, The University, Newcastle upon Tyne, NE1 7RU, U.K.
[b] Present address: United Kingdom Chemical Information Service, The University, Nottingham, NG7 2RD, U.K.

Precision, recall, and coverage ratios have been estimated for an experimental SDI service based on the *Index Chemicus Registry System* (ICRS) machine-readable data base. The relative currency of the service was also measured. Failure analysis was used to identify the causes of recall and precision failures, and possible methods for improving retrieval performance are suggested. User reactions to the service were assessed by means of a questionnaire survey. The results of the evaluation are interpreted in relation both to substructure searching in general, and to the ICRS data base.

the validity of recall and precision as measures of retrieval performance from the users point of view, appear in the full report[3] and in other reports from the Experimental Information Unit.[4,5]

THE EXPERIMENTAL POPULATION

During the course of the experiment, search profiles were written for 183 research workers, of whom 101 volunteered to take part in the evaluation and to provide the required feedback. The evaluation population consisted of 78 research workers from 31 different university departments, 22 from four industrial organizations, and one user from a government research department. The industrial population was mainly from one large industrial organization and may, therefore, have been unrepresentative.

THE SDI SERVICE

As explained in the earlier paper,[1] it was our intention to offer participants a service which would satisfy as many of their current awareness requirements as were appropriate to the subject coverage of the data base. Participants were not asked to restrict their requirements to structural interests. We were, therefore, evaluating the use of ICRS tapes for the provision of a general-purpose SDI service rather than for answering specific, exactly-defined substructure search requests. In practice, though requirements were often diverse, the subject coverage of ICRS dictated that structural concepts were a dominant feature of the requirements of most users.

Both WLN and word terms (and also molecular formulas and instrumental data alerts) could be used in search profiles. Word terms were matched against the title, the subject index and the use profile records of the ICRS file. Our main interest was in the design of satisfactory WLN substructure search strategies.[1] However, even for a population with strong interests in chemical structures, word searches were an essential complement to the WLN file for several reasons:

(i) An interest in chemical structures is not necessarily exclusive. There is often a need for co-ordinate searches involving both structural and nonstructural concepts.

(ii) Some requirements which could normally be described as "structural" cannot be satisfied by any form of substructure search. Examples are groups of compounds defined partly by structural features, and partly by their biogenetic origins (e.g., terpenes, alkaloids) or by nonspecific stereochemical features (e.g., hindered phenols). A natural or controlled language subject index is essential in searches for such groups of compounds.

(iii) A word index can provide an alternative means of substructure searching to circumvent some of the limitations of a WLN structure file.

PROCEDURE FOR MEASURING RETRIEVAL PERFORMANCE AND COVERAGE

The retrieval performance of the service was measured by the following parameters: the output size, the recall ratio, and the precision ratio. A coverage ratio *relative* to *Chemical Abstracts* (CA) was also obtained.

The evaluation was concerned with the performance of the complete retrieval system, of which the user himself is a part, and all performance ratios are based on the user's relevance judgements. Values for recall and precision ratios are therefore influenced by several variables: the effectiveness of communication between the user and the system, the skill of the search strategists, the quality of the index languages, and the logical facilities[1] provided by the search program.

Relevance Judgements. All relevance judgements were based on either the digest appearing in the printed CAC & IC or on an abstract in *Chemical Abstracts*. Users were asked to classify references as RELEVANT (R), IMPORTANT (I), or IRRELEVANT. They were also asked to indicate any reference which CONTAINS STRUCTURE SOUGHT (S), irrespective of whether they considered the reference to be IMPORTANT, RELEVANT, or IRRELEVANT. The following definitions were offered in instruction sheets which were distributed to users:

IMPORTANT—a reference directly related to your research interests which you will read as soon as possible—i.e., it is essential that you know of it

RELEVANT—a reference which, although not important, is of some interest or relevance to you—i.e., you are glad that you were notified of it

CONTAINS STRUCTURES SOUGHT—at least one of the chemical formulas listed on the digest contains a structure (or substructure unit) for which you wanted a search made

The definitions of IMPORTANT and RELEVANT references were intended only as a general guide to users, and it was recognized that each user would impose his own personal interpretation on these categories. We surmised that the additional category, CONTAINS STRUCTURES SOUGHT, was susceptible to a more objective interpretation. Our purpose in including this category was to provide a measure of the theoretical effectiveness of the substructure search, which is dependent only on the quality of the search strategy and the characteristics of the WLN index, and is not influenced by the fact that a user's requirements may not be exactly definable in structural terms.

In practice, we suspected that this category was also subject to personal interpretation and, in a few cases, to misinterpretation owing to ambiguity in the instructions given to users. Nevertheless the effect of these misinterpretations would be to underestimate the number of references in the category, and this would not affect the conclusions which we drew from the experimental data (see below).

The Precision Sample. The precision ratio is defined as:

$$\frac{\text{No. of } \textit{relevant} \text{ references } \textit{retrieved} \text{ by the search}}{\text{Total no. of references (relevant or irrelevant) retrieved}}$$

The precision sample comprised either two or three monthly outputs from the SDI service. For these outputs, participants were asked to record all IMPORTANT, RELEVANT, and CONTAINS STRUCTURES SOUGHT references on special forms which were sent with the output. Copies of the printed CAC & IC were provided to the user so

Table I. Mean Values for the Output Size, Precision Ratios, and Recall Ratios

Relevance Category	Output Size, Refs./Month	% Precision	% Recall (CAC & IC Sample)
	42.2 ± 6.4 (97)[a]	—	—
IMPORTANT + RELEVANT	—	31.9 ± 2.4 (96)	63.9 ± 3.1 (86)
IMPORTANT ONLY	—	8.8 ± 1.1 (94)	71.4 ± 4.1 (63)
CONTAINS STRUCTURE...	—	46.6 ± 3.5 (91)	62.7 ± 3.2 (75)

[a] For outputs from the issues which supplied the precision samples.

Table II. Output Sizes, and Precision and Recall Ratios for Different User Groups (I + R References)

User Group	Output Size Refs./Month	% Precision	% Recall (CAC & IC Sample)
All users	42.2 ± 6.4 (97)	31.9 ± 2.4 (96)	63.9 ± 3.1 (86)
Academic	44.8 ± 6.9 (76)	34.1 ± 2.8 (76)	66.6 ± 3.2 (68)
Industrial	17.0 ± 3.1 (20)	23.4 ± 3.9 (19)	52.6 ± 8.1 (17)
JFBR	44.6 ± 8.2 (60)	33.6 ± 3.3 (60)	68.0 ± 3.7 (54)
BNR	38.2 ± 10.3 (37)	28.9 ± 3.5 (36)	57.0 ± 5.1 (32)
BNR/Academic	45.6 ± 12.3 (16)	36.2 ± 5.9 (16)	61.1 ± 6.3 (14)
WLN-MF	35.1 ± 5.2 (47)	32.8 ± 3.3 (47)	61.3 ± 4.2 (43)
WLN/word	48.8 ± 11.4 (50)	30.9 ± 3.5 (49)	66.4 ± 4.4 (43)

that he could consult the digests of the retrieved references before making his relevance decisions.

The Recall and Coverage Samples. The recall ratio is defined as:

$$\frac{\text{No. of } \textit{relevant} \text{ references } \textit{retrieved} \text{ by the search}}{\text{No. of } \textit{relevant} \text{ references in the data base}}$$

The coverage ratio is defined as:

$$\frac{\text{No. of } \textit{relevant} \text{ references (from source X) in the data base}}{\text{No. of } \textit{relevant} \text{ references in source X}}$$

The recall ratio *as defined above* is a characteristic of the retrieval service and is independent of the source of the sample. Estimation of this ratio requires a sample which is effectively random with respect to relevant references in the data base. In contrast, the coverage ratio is a relative measure, dependent on the source (X) of the coverage sample. Two recall samples and one coverage sample were obtained.

The CAC & IC Sample. Users undertook a search of one or two months issues of the printed CAC & IC before they received their computer printouts corresponding to these issues. IMPORTANT, RELEVANT, and CONTAINS STRUCTURES SOUGHT references were listed on forms similar to those used for the precision feedback. These references provided a recall sample.

The CA Sample. Users identified a sample of relevant references by searching one or more sections from the odd-numbered issues (biochemistry and organic chemistry) of *Chemical Abstracts*. The user was asked to select the section or sections which he considered to be most appropriate to his research interests. For example, a user interested in cyclohexadienones chose to scan sections 24 (alicyclic compounds) and 31 (alkaloids). In this case, users were not asked to identify CONTAINS STRUCTURES SOUGHT items. The complete set of references found in this search provided a coverage sample. References in the sample which also appeared in the ICRS data base provided a recall sample.

Apart from the questionnaire replies, the search of *Chemical Abstracts* was the last type of feedback required from the users, and was also the most laborious task given to them. As a result, CA samples were returned by only 39 participants. Six others undertook a search but reported no RELEVANT or IMPORTANT references.

Overall Recall. We have adopted the term overall recall for the following ratio:

$$\frac{\text{No. of } \textit{relevant} \text{ references (from source X) retrieved by the search}}{\text{No. of } \textit{relevant} \text{ references in source X}} = \text{recall ratio} \times \text{coverage ratio (relative to source X)}$$

Though it is desirable to differentiate between coverage and recall performance from the viewpoint of the system designer, the user is primarily concerned with the proportion of missed references, irrespective of whether these are designated as retrieval or coverage failures. Overall recall can be considered as a measure of retrieval performance as seen by the user. Like the coverage ratio, it is a relative measure.

Table III. Mean Recall, Coverage and Overall Recall Ratios from the CA Sample (I + R References)

Parameter	All Items	Serial Items Only
% Recall, CAC & IC sample	—	69.7 ± 4.7 (31)
% Recall, CA sample [1]	—	68.0 ± 5.5 (31)
% Coverage, CA sample [2]	38.5 ± 4.8 (39)	49.4 ± 5.6 (38)
% Overall recall [1] × [2][a]	24.5 ± 3.8 (31)	29.4 ± 4.2 (31)

[a] Computed by obtaining the product of recall and coverage ratios for each profile, then averaging the products.

Table IV. The Proportion of Output and the % Precision Attributed to Different Term Types (WLN, Text)

Term Type	% of Total Output	% Precision
Text only[a] (titles, subject index, use profiles)	26.5 ± 3.7	39.4 ± 4.9 (48)
Subject index or use profile exclusively[b]	4.5 ± 1.1	40.9 ± 7.1 (31)
WLN only[a]	60.5 ± 4.3	35.5 ± 3.1 (75)
Text or WLN[c]	3.3 ± 0.9	61.2 ± 8.2 (20)
Text and WLN[d]	4.3 ± 1.4	43.9 ± 9.1 (15)

[a] The reference would be retrieved by this term even if no other term types were included in the profile. These categories include references retrieved by two term types acting independently and also counted in the Text *or* WLN group.

[b] Restricted to references which would *not* have been retrieved by a search of titles alone.

[c] References which would have been retrieved by either term type even if the other had been absent.

[d] Retrieval required both a Text *and* a WLN term.

Statistical Methods. Mean values of the performance ratios were obtained by first computing ratios for each user and then averaging these ratios. All ratios are expressed as percentages. The standard errors, which are quoted for all mean values, define the probability that, if the experiment were repeated with a different but equally representative population, the same estimate would be obtained. In Tables I to IV, the number of users contributing data to a mean value is given in parenthesis after mean value and its standard error—e.g., 51.5 ± 3.0 (61).

A 0/0 ratio was always treated as meaningless and excluded from the calculation of mean values.

Wilcoxon's U-statistic[6] was used to test for significant differences between two mean values—e.g., mean values for two subsets of the user population, or two estimates for the same parameter obtained by different methods for the same users. This statistic is nonparametric and is therefore suitable for use with parameters which do not exhibit a normal distribution, such as the performance measures for an information retrieval system.

EXPERIMENTAL VALUES FOR PERFORMANCE PARAMETERS

Table I gives the mean output size, and the precision and recall ratios for three categories of reference: IMPORTANT and RELEVANT (I + R), IMPORTANT only (I), and CONTAINS STRUCTURES SOUGHT (S) references. The values of these parameters for individual users varied widely about the mean, as is shown by the standard deviations: ±63 (output size), ±24% (precision), and ±29% (recall). Thus, 28 profiles had an output size of less than 10 items/month, and 9 profiles were retrieving more than 100. Similarly, precision ratios varied from less than 10% (11 profiles) to greater than 80% (5 profiles).

Recall ratios for individual users showed no correlation with either output sizes or precision ratios—i.e., large output sizes were *not* associated with high recall but (presumably) represent a particular type of user requirement.

Influence on Output Size of Changes in File Size. The mean output size of profiles was influenced by major changes in the ICRS coverage policy[7] which were implemented in the middle of the evaluation experiment. Section B of ICRS (and of CAC & IC), which included articles which were of chemical interest but which did not refer to new structures, was abandoned, and at the same time, the number of source journals covered was reduced by approximately half. As a result of these changes, the file size fell from 45,961 articles in 1970 to 13,807 in 1971. Approximately 70% of output size and precision data were based on outputs from the 1970 file, and 30% on outputs from the 1971 file. For the 76 academic users, the mean output size during November 1970 was 44.6 references per month compared with 31.9 in January 1971. Though the file size had been reduced to only one-third of its original size, the mean output size dropped by only 28%. This reflects the fact that most of the output was being retrieved by WLN terms whereas most of the reduction in file size resulted from the removal of section B, which contributed no WLN records to the structure file. We did not have comparable precision figures for 1970 and 1971, but there is no reason to believe that changes in file size materially affected the mean precision of the profile.

Performance Ratios for Different Relevance Categories. As was found in Lancaster's evaluation of MEDLARS,[2] and in evaluations of *Chemical Titles*[8] and *BA Previews*[5] undertaken by the Experimental Information Unit, the recall ratio for IMPORTANT references (71.4%) is significantly higher than that for all relevant references (I + R) (63.9%). These figures suggest that it may be a general characteristic of SDI services that they are somewhat better at retrieving important items than the "merely relevant" ones.

Almost half (46.6%) of the references retrieved were judged by the user to contain the structures in which he was interested, even though the reference was not necessarily relevant (I or R) to him. Thus, the technical effectiveness of the profiles in retrieving specified structural fragments is substantially better than is suggested by the precision ratio for relevant (I + R) items (31.9%).

Participants in the experiment were not asked to restrict themselves to concepts of an exclusively structural nature. However, we found that 83% of IMPORTANT retrievals were also judged by the user as CONTAINS STRUCTURES SOUGHT. This gives support to our assertion that structural concepts were the dominant feature of most users' requirements.

Performance for Different Groups of Users. Table II shows performance data (I + R references only) for different subsets of the population, defined by the following characteristics:

(i) The user's employment category: industrial or academic

(ii) The liaison scientist responsible for the profile: JFBR (J. F. B. Rowland) or BNR (B. N. Rossiter). All profiles for industrial users were written by BNR; profiles for university users were divided between the two liaison scientists

(iii) The term types used in the profile:

(a) WLN-MF profiles in which a WLN or molecular formula term was *always* essential for the retrieval of a reference. This definition does not exclude the retrieval by a combination of a WLN (or MF) term *and* a word term.

(b) WLN/word profiles in which at least *some* references could have been retrieved by word terms alone.

The figures given in Table II show that the % precision, the % recall and the output size for academic users were all significantly higher than for industrial users. These differences cannot be attributed to the liaison scientist (BNR) who wrote all the profiles for industrial users. The two sets of profiles written for academic users by the two liaison scientists showed no significant differences in performance.

A further illustration of the different characteristics of academic and industrial users is that the profiles of the former retrieved 13.7 ± 2.8 relevant (I + R) references per issue, compared with 3.3 ± 0.7 for the industrial users. It appears that our industrial users demanded a smaller output and had more specific requirements than their academic counterparts, and profile strategies were less successful in satisfying such requirements. Though these results support our own impressions regarding the requirements of industrial users, we have already pointed out that this sample of users may be unrepresentative. We cannot be certain whether our data are characteristic of all industrial users or only of those in one research laboratory. The observed differences are indicative rather than definitive.

Recall and Coverage Estimates from the CA Sample. Table III gives the performance ratios obtained from the CA samples. These samples provided recall ratios for only 31 users. For comparison purposes, the mean recall ratio computed from the CAC & IC sample for the same 31 users is also tabulated.

There is no significant difference between the two estimates of recall. If we discount the possibility of both samples being equally biased, the identity of the two estimates provides evidence that the searches of either CA or CAC & IC give an effectively random sample of relevant items in the ICRS data base, and hence both samples will provide valid estimates of the recall ratio.[2,4]

The relatively low estimates for the coverage ratio relative to *Chemical Abstracts* (38.5% of all items) and hence for the overall recall ratio (24.5% of all items) are to be expected in view of the article selection rules for ICRS (articles referring to *new* compounds, reactions or syntheses selected from 103 key source journals).

Of the articles not covered by ICRS, 32.1% were from nonserial publications (e.g., government reports), 44.5% were from journals which are not in the 1971 CAC & IC source list, and the remaining 23.4% were articles which were not selected for inclusion in the data base though they were from journals in the source list.

Article content was examined for a small sample of 59 relevant articles, from serial publications, which were not covered by ICRS. In selecting this sample, preference was given to articles which had been judged IMPORTANT rather than RELEVANT, if the former were available.

We judged that 41 articles (70%) in this sample were within the stated subject field of the data base—organic and pharmaceutical chemistry. Of these 41 articles, nine referred to *new* compounds or syntheses as defined by ISI. None of these nine appeared in journals which are covered by ICRS and hence there was no evidence of indexer error in article selection. The remaining 32 articles (including 4 review papers) were concerned with the synthesis, reactions, or properties of organic compounds which were not *new*.

We concluded that the coverage policy is reasonably successful in relation to its declared objective of covering new compounds, but less so when considered in relation to articles which the user judges as relevant.

Output Size and Precision Attributable to Different Term Types. Table IV gives mean values for the proportion of references retrieved by different term types and the % precision achieved by each term type. (Figures for instrumental data alert and molecular formula terms were not computed). The value of subject index words and use profiles in augmenting an author's title is indicated by the fact that 17% (4.5/26.5) of all text retrievals would not have been retrieved but for these enrichments. The precision ratios for references retrieved by text and by WLN terms respectively are not significantly different. On the grounds that word terms are, in general, less specific than WLN terms, we would have expected a much lower precision for text retrievals. The rather low proportion (26.5%) of items retrieved by text terms and the unexpectedly high precision of such items, is indicative of our highly selective use of word terms. It is likely that significantly greater use could be made of word terms without substantially reducing precision.

Articles which were associated with both a "relevant" word term and a "relevant" WLN term exhibited very high precision (61.2%) but this conclusion is only of academic interest in view of the small proportion of articles which appear in the category (4.3%).

Comparison with Chemical Titles (CT). Table V shows that the retrieval performance of the ICRS SDI service for academic users compares favorably with that of a *Chemical Titles* service which was supplied to a population of university chemists in another study[8] undertaken by the Experimental Information Unit.

In view of the greater sophistication of the ICRS indexes, more dramatic differences in performance might be expected. A possible explanation is that differences in the retrieval performance of the two indexes are obscured by the substantial proportion of retrieval failures attributable to inadequate user/system communication.

Article and Compound Retrieval. The evaluation measured the retrieval performance for individual articles, not for individual compounds. For example, the criterion of a *miss* is that the profile failed to retrieve the article. However, each document has an average of 10.9 encoded compounds and this provides substantial redundant retrieval power. A profile may fail to match with three relevant compounds but succeed on the fourth and, as the article is retrieved, no failure is recorded. Conversely, if the profile fails to match with all four compounds, only one failure is recorded. It is probable that recall, and possibly precision, would have been lower if we had measured performance in relation to compound retrieval.

FAILURE ANALYSIS[2]

A retrieval failure is objectively defined as (a) an article which was judged as relevant by the user but not retrieved by the service (a recall failure), or (b) an article which was judged irrelevant by the user but was retrieved by the service (a precision failure). An analysis of samples of individual recall and precision failures was undertaken to determine why a failure occurred, to identify the most frequent causes of failure, and to suggest methods of preventing such failures in a future service. The ultimate objective of this failure analysis was to identify practical means for enhancing retrieval performance.

Failures were classified into four broad categories corresponding to the main subsystems which comprise an information retrieval service: the communication process, the data base, the search strategy, and the computer search system. These four categories were further subdivided as required. It must be emphasized that we viewed failure analysis as an empirical exercise in which the definitions of our failure categories are determined by the characteristics of the particular service, and the viewpoint of the investigator. The meanings we attach to terms such as communication failure, and to the other failure categories, may not be universally agreed, but are defined by the examples and explanatory notes given for each category.

The depth of the analysis was deliberately uneven. More attention was paid to WLN than to word search failures. Also, strategy failures were examined in somewhat greater detail than any other group. The design of the search strategy, unlike that of the data base itself, was directly under our control: we were responsible for the construction of all profiles. We therefore had both the incentive for a detailed analysis and access to the necessary information.

Procedures for Failure Analysis. Samples for use in failure analysis were obtained by selecting at random a maximum of five recall and four precision failures for each profile. In some cases, the profile did not provide the required number of failures and samples were restricted to a total of 357 precision failures from 93 profiles, and 252 recall failures from 67 profiles.

Each liaison scientist classified failures for the profiles for which he had been responsible, and his classification was then checked by the other liaison scientist. Disagreements were resolved by discussion. The following data were used in the analysis:

(i) The user's written statement of requirements (if available), the liaison scientist's notes, any correspondence with the user, and any other information about the user's interests. As profile strategies had been negotiated in collaboration with the user, the profile itself sometimes provided evidence of the user's interests which was not explicit in the statement of requirements

(ii) The search profile

(iii) The ICRS index records

(iv) The CAC & IC abstract

Where necessary the full document was also consulted.

The analyses of both recall and precision failures were partly hierarchical. The first step in either analysis was the identification of communication failures by comparing the contents of an article with the user's requirements *as we understood them.* If a recall failure did not match the user's requirements, or if a precision failure did do so, the failure was attributed to the communication process and was not considered as a candidate for any of the other failure categories.

Classification of the remaining failures as being due to the search strategy, to the data base, or to the search system was often self-evident. However, in some cases, we had difficulty in deciding whether a failure should be attributed to the search strategy, or to inherent limitations of the index language. In all cases of doubt, a failure was attributed to the search strategy. Dual classification was

Table V. Comparison of ICRS and CT SDI Services

	ICRS	CT
% Recall	66.6 ± 3.2	59.8 ± 1.9
% Precision	34.1 ± 2.8	23.8 ± 1.1
Output size (refs/month)	44.8 ± 6.9	118.3

allowed in this and other cases but was adopted for only 7.1% of recall failures and 0.7% of precision failures.

Data base failures may result either from inherent characteristics of the index language, or from an indexer error. This distinction is clear-cut for the WLN index, the use of which is dictated by exact rules, but not for natural-language subject word indexes. In the latter case, we made no attempt to discriminate between the index language and the indexing process.[2]

THE ANALYSIS OF RECALL FAILURES

Of the 252 articles in the recall failure sample, 88% referred to relevant structural concepts, and 12% to relevant nonstructural concepts. The recall failure categories are listed in Table VI, which gives the number and the percentage of recall failures assigned to each category and the number of profiles which contributed at least one article to a category.

The first stage in the analysis was the identification of articles which were not retrieved because of major omissions from the profile, which could be classified as either communication or strategy failures but could not be specifically associated with either the WLN or the word searches. These failures are described under the heading "Recall Failures, General." The remaining articles (72.2%) in the sample were analyzed in relation to the WLN and/or the word searches. Articles referring to structural concepts were analyzed in relation to both the WLN and the word search. Those referring to nonstructural concepts were considered only in relation to the word search.

Table VI. Recall Failures Analysis[a]

	No. of Failures	% of Failures	No. of Profiles in Which Failure Occurred
Recall failures, general			
The Communication Process	**44**	**17.5**	**24**
The Search Strategy	**26**	**10.3**	**12**
Recall failures, the WLN search			
The Communication Process	**53**	**21.0**	**27**
(i) Omission of relevant structures	41	16.3	24
(ii) Exclusion of relevant structures	12	4.8	9
The Data Base	**60**	**23.8**	**38**
(i) Compound selection rules	49	19.4	32
(ii) The WLN index	9	3.6	8
(iii) WLN coding errors	4	1.6	2
The Search Strategy	**35**	**13.9**	**18**
(i) Omission of alternative notation symbols	11	4.4	5
(ii) Omission of alternative notation paths	21	8.3	12
(iii) Over-restrictive logic	6	2.4	4
The Search System	**13**	**5.2**	**10**
Recall failures, the word search			
The Communication Process	**8**	**3.2**	**6**
The Data Base	**53**	**21.0**	**29**
(i) Concept completely absent	33	13.1	21
(ii) Trivial or unusual names	17	6.7	11
(iii) Data base (key-punching) errors	3	1.2	3
The Search Strategy	**131**	**52.0**	**59**
(i) Concepts not present in the profile	90	35.7	42
(ii) Omission of related terms	35	13.9	19
(iii) Restrictive logic	6	2.4	3

[a] Note: Total figures for the primary failure categories, corresponding to the main subsystems of a retrieval service, are given in bold type. As a result of dual classification of a few failures, the sum of the figures for subdivisions of these categories (shown in normal type) is sometimes greater than the total figure for the category.

The proportion of the recall failure sample placed in each category is indicated in parenthesis after the category heading.

Recall Failures, General. *The Communication Process* (17.5%). This category was reserved for missed articles which referred to a subject area or structural type which was quite distinct from the user's known interests. Other missed articles which represented only a marginal extension of a user's interest were classified as communication failures in relation to the WLN or word searches (see below). Inevitably, in some cases it was difficult to separate this category both from the other groups of communication failures, and from the general search strategy failures discussed below.

The majority of general communication failures were articles which, though not included in the user's stated interests, were recognized as being related to those interests. Also included in this category were a few documents which, to the outside observer, bore no obvious relationship to the user's known interests.

EXAMPLE 1: A missed article concerned with isoquinolines for a user who was interested in quinolines.

EXAMPLE 2: An article describing the oxidation of isoquinoline by ferrate ion was reported as relevant by a user whose interests were thought to be: ketoamides, imidazolidinones, trifluoromethylanilines, and guanyl- or amidinoureas.

The Search Strategy (10.3%). This group of recall failures were articles which referred to topics which were explicitly mentioned in the user's requirements, but which the profile made no attempt to cover either by a word or a WLN search. Fifteen of these failures appeared to be the result of a straightforward error of omission by the liaison scientist.

EXAMPLE 3: A user was interested in "enamines" (*inter alia*), but the profile contains no terms corresponding to this concept.

The remaining eleven failures arose because the liaison scientist considered that the user's stated interests would produce a prohibitively large output and therefore wrote a profile which was more restrictive than the user's requirements.

EXAMPLE 4: The user was interested in "all N-heteroaromatics." An article on pyrimidines was missed because the search was restricted to selected N-heteroaromatic structures only.

There were two possible reasons for failures of the types illustrated in Examples 3 and 4:

(a) Insufficient time had been spent in defining the user's requirements.

(b) The user has an interest in "some papers of special interest" within a very broad field. This is a perfectly legitimate requirement but impossible to satisfy with any normal index. In an analysis from a different viewpoint, this might be classified as an unavoidable communication failure because the user was unable to specify what he meant by some papers of special interest.

Recall Failures, the WLN Search. *The Communication Process* (21.0%). Construction of a satisfactory WLN profile requires as an exact a definition as possible of the structures which are of interest to the user. This may lead to communication failures resulting from an over-specific definition of the user's requirements. In contrast to communication failures described under the heading "Recall Failures, General," a majority of such failures could have been prevented by minor profile amendments which would not have increased output size significantly. It is probable that many potential failures of this type were avoided by the liaison scientists' normal practice of writ-

ing a profile which was somewhat broader than a user's stated interests. The broad profile often covered structures which *might* be of interest to the user, even though they were not specified in his requirements. Communication failures resulted both from accidental omission (16.3%) and, more rarely, from explicit exclusions (4.8%) of relevant structures from the user's stated requirements.

EXAMPLE 5 (omission of relevant structures): The user was interested in nonbenzenoid aromatic structures and supplied a list of the relevant structures. He omitted to mention azulene which was the subject of the missed document.

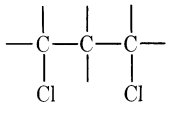
EXAMPLE 6 (exclusion of relevant structures): The user was interested in quinones and explicitly stated that only carbocyclic quinones would be of interest. He subsequently reported a relevant document which referred to a heterocyclic quinone (2,3,6-triazanaphthoquinone-1,4).

The Data Base (23.8%). The great majority of data base failures was due to inherent characteristics of the article selection rules, or of the WLN index language. Only five failures could be attributed to errors by the indexers.

(i) The compound selection rules (19.4%): ISI's compound selection rules stipulate that only *new* compounds should be registered on the WLN structure file. The compound is considered to be *new* if a paper gives no reference to a previous preparation. There were 49 recall failures in which a relevant structure appeared as a structural diagram in the printed CAC & IC but was not selected for inclusion on the structure file. An examination of 34 of the original documents showed that in 33 cases the ISI selection rules had been correctly applied and in only one case had an indexer error resulted in a compound not being registered even though it appeared to be new according to ISI's definition. In the 33 documents in which the selection rules had been correctly applied, the relevant compound(s) was a reactant in 19 cases, a product in 14, an uncharacterized intermediate in one, and a compound which could not be prepared by a particular method in one.

Though it is obviously impractical to code all compounds mentioned in a paper, it is of interest that in 34 (69%) of the 49 failures some mention of the relevant compound appeared in the title. These failures might have been avoided by extending the article selection rules to include any compound referred to in the title of the paper, even if it is not new.

(ii) The WLN Index (3.6%): some types of substructure requirement can never be adequately satisfied by a WLN search. For example, it would be impossible to devise a WLN search strategy for the fragment:



in which none of the groups to be attached to the unsatisfied valencies are defined. Only two recall failures were attributed to the inherent limitations of the notation. However, recall failures were attributed to the search strategy rather than to the WLN index language, if there was even the slightest possibility of a reasonable strategy.

Certain types of compound (e.g., metal complexes, inorganic compounds) are not coded in WLN on the structure file. Seven articles were not retireved because there were no WLN records for the relevant compounds and we were not able to compensate for this by the use of the molecular formula file.

(iii) WLN Coding Errors (1.6%): only four failures resulted from indexer or clerical errors in the WLN coding.

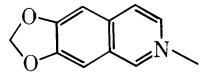
The Search Strategy (13.9%). Recall failures attributable to the search strategy were of three types:

(i) Omission of alternative notation symbols (4.4%)

EXAMPLE 7: A profile, which was designed to search for organic peroxides, included the WLN term OO, but not the term OQ. The missed document referred to a compound which contained the substructure CH—O—O—H, in which the peroxide group is coded as OQ.

(ii) Omission of alternative notation paths (8.3%): The majority of these omissions (17 out of 21) were associated with searches for a ring structure which is embedded within a larger ring system. Embedded ring[1] requirements always present problems for a WLN search and, even if the problems are soluble, the profile strategy will usually be complex. The relatively small number of failures of this type was therefore encouraging.

EXAMPLE 8: The requirement was for structures containing the isoquinoline ring skeleton. The profile searched for a variety of ring systems which contain this structure by the use of the terms T66 CN, *or* T E6 B666 CN, *or* T D6 B666 *followed by* -DN, together with obvious variations such as T66 CM, *or* T D6 B666 *followed by* -JVN. The missed document referred to a compound (I):



(I)

which is coded as T C566 DO FO KN EHJ, and was not retrieved.

(iii) Over-Restrictive Logic (2.4%): Six articles were not retrieved because of either the unpredicted effect of *not* (or *ignore*) terms, or the use of over-restrictive contextual logic. The latter[1] allows the co-ordination of two (or more) WLN fragments with the further stipulation that the two fragments must not be separated by more than x characters in the notation. In four instances a document would have been retrieved if a larger value of x had been used.

The Search System (5.2%). No failures were attributed to the inherent limitations of the search logic though these may have contributed to some WLN search strategy failures. Of the 13 failures which were attributed to the search system, eight were assumed to be due to errors in the experimental search programs, as no other explanation for the retrieval failure could be found. The remaining five failures resulted from mistakes in profile coding which were not corrected in subsequent error checks. Even the relatively small number of search system failures which were observed is higher than would be countenanced in an operational service.

Recall Failures, The Word Search. Of the 184 missed documents which were analyzed in relation to the word search, 159 (86%) referred to relevant structural concepts.

The Communication Process (3.2%). These were articles which referred to relevant nonstructural concepts which were not included in the user's requirements, as we understood them.

The Data Base (21.0%). Some types of substructure search are a recognised part of profiles for searching title records in data bases such as *Chemical Titles* and *BA Previews*[9]—e.g., *OL, *OLS to retrieve alcohols; *AZ* to retrieve (*inter alia*) a variety of nitrogen heterocycles. A nomenclature index, such as that now available from Chemical Abstract Services, would probably provide better substructure retrieval facilities than titles, but this possibility has not yet been fully explored. Nevertheless one would expect that many structural fragments could never be adequately represented in a title and subject word index. The number of articles (50, 19.8%) which we judged could not have been retrieved by a search of the title or subject index records was suprisingly small. These recall failures were divided into two categories:

(i) The relevant concept, though present in the printed digest or the graphic abstract of CAC & IC, was completely absent from the title and the subject index.

EXAMPLE 9: The user was interested in carbohydrates, and the missed document contained a compound of this group. The title was "New Synthesis of Bicyclic Phos-

phites." The subject index words, which were added by the ISI indexers, were: PHOSPHORUS ACIDS; CYCLIC ESTERS; NEW SYN, TRIOXAPHOSPHABICYCLOOCTANES; NEW SYN, BICYCLIC PHOSPHITES. None of these words indicate that the paper refers to carbohydrate structures.

(ii) The relevant structural concept was inadequately represented (from the search strategists viewpoint) by trivial, specific, or unusual names.

EXAMPLE 10: The user has a very broad requirement for articles on terpene chemistry. The title of a missed document was: "Iridoids. 10-0-betaglucosylaucubine from *linaria vulgaris*." The subject index words consisted solely of permutations of the significant words in this title. Iridoid is the name of a particular class of terpenes. Though the profile contained the search term TERPEN*, it would have been impossible to add all the trivial names for terpenes.

Seven of the failures in the category typified by Example 10 were also classified as search strategy failures. This dual classification recognizes that, in searching a natural language index, it is the responsibility of a search strategist to make some allowance for the variations and vagaries of natural language indexes.

We did not attempt to distinguish between data base failures which were due to inherent limitations of word indexes, and those in which the indexer might reasonably have been expected to compensate for the deficiencies of the author's title. However, in the course of the analysis we made two general observations:

(i) In some instances, most of the subject index words are a repetition of words which appear in the title and hence provide little additional information.

(ii) The most likely direction for improving subject index words is by more generous assignment of generic names for groups of compounds.

The Search Strategy (52.0%). We intended that profiles should make full use of all the term types provided by the data base, but our primary interest was in the design of WLN searches, and the attention paid to word terms suffered accordingly; 34 profiles did not contain any word terms. In our judgement, over half of all recall failures were, *in principle*, retrievable by a search of the text records—i.e., the relevant concept was represented in some form in the title word, subject index word, or use profile records. In some borderline cases, it is unlikely that a reasonable strategy would have included the necessary terms—e.g., the trivial names for all members of a large group of compounds. However, both the examples given below illustrate *avoidable* errors in the search strategy.

(i) Concepts not present in the profile (35.7%): These 90 recall failures occurred in profiles in which there had been no attempt to construct a word profile covering the relevant concepts.

EXAMPLE 11: The user was interested in all articles referring to the chemistry of benzoquinones, but no word terms were included in the profile. The missed article referred to phenyl-1,4-benzoquinone-2-sulfonate and the word BENZOQUINONES appeared in the title. The search term *QUINON* should have been included in the profile and would undoubtedly have improved the recall performance.

(ii) Omission of Related Terms (13.1%)

EXAMPLE 12: The requirement was for santonin and its derivatives. The word SANTENOLIDES appeared in the title of the missed article. The article was not retrieved because the profile contained the term SANTONIN* but made no allowance for related terms, such as SANTENOLIDES, which could be covered by the term SANT* without serious loss of precision.

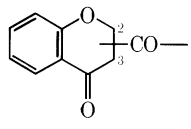
(iii) Restrictive Logic (2.4%): The profile included a term which matched with a relevant article but the article was not retrieved because the search strategy required (unnecessarily) that a second concept should be present in the index records for the article.

THE ANALYSIS OF PRECISION FAILURES

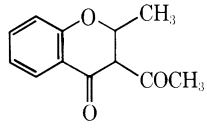
Responsibility for the retrieval of articles which were judged irrelevant by the user could always be clearly attributed to a particular term type or combination of term types. The duplicate analysis which was necessary for recall failures was not, therefore, required for precision failures. Table VII summarises the results of our analysis of 357 precision failures, 80% of which were retrieved as a result of matches with index records referring to *structural* concepts. The proportion of the precision failure sample placed in each category is indicated in parenthesis after the category heading.

The Communication Process (48.2%). Almost half of our sample of irrelevant articles appeared to match the users' requirements, as we understood them, and we had no basis for deciding why the article was judged irrelevant (to provide a clearer picture of why these failures occur, it would have been desirable to have asked the user why he judged the reference irrelevant, as was done in Lancaster'[2] evaluation of MEDLARS).

EXAMPLE 13: The user's requirement was for chromone-2 and chromone-3-carboxylic acids and their derivatives (II).



(II)



(III)

An article referring to compound (III) was judged by the user to be irrelevant.

EXAMPLE 14: The requirement was for articles on the toxicity of chemical compounds. The search term TOXICITY retrieved the title: "acute oral toxicity of dimethoate," but the article was considered irrelevant by the user.

Table VII. Precision Failure Analysis

	No. of Failures	% of Failures	No. of Profiles in Which Failure Occurred
The Communication Process	**172**	**48.2**	**70**
Precision Failures, the WLN Search			
The Search Strategy	**110**	**30.8**	**51**
False Co-ordinations	39	10.9	19
Lack of Specificity in the Search Logic	64	17.9	35
Degeneracy of WLN Symbols	7	2.0	6
The Data Base	**14**	**3.9**	**9**
WLN Index Language	8	2.2	4
Coding Errors	6	1.7	5
The Search System	**7**	**2.0**	**6**
Precision Failures, the Word Search	**45**	**12.6**	**25**
Precision Failures, the WLN *and* the Word Search	**5**	**1.4**	**2**
Precision Failures, the Author Search	**6**	**1.7**	**3**

We believe that some, perhaps many, of such communication failures are inevitable in that they arise because:

(i) The user is unable to provide a more precise translation of an intuitive understanding of his own information needs into a logical statement of requirements, or

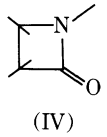
(ii) He is unwilling to do so, because he had decided (probably wisely) that, beyond a certain point, he will not delegate to the retrieval system the responsibility of discriminating between relevant and irrelevant reference. He elects to receive all references on a particular topic, knowing that some will be irrelevant, and reserves the right to make his own final selection.

Some communication failures will occur, for the reasons discussed above, in most retrieval services but their incidence in the ICRS SDI service is probably accentuated by the exhaustivity of the WLN compound index. Though a user may have expressed an interest in all articles referring to a give substructure, he may in fact only be interested in the substructure when it is the main topic of the article. However, the WLN search may retrieve articles which are judged irrelevant because the substructure is incidental to the main subject matter of the paper.

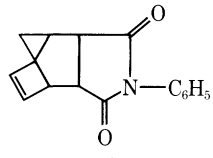
Precision Failures, The WLN Search. *The Search Strategy* (30.8%). With the exception of seven irrelevant articles (2.0%) which were retrieved because we had failed to allow for the degeneracy of some WLN symbols, precision failures associated with the WLN search strategy were attributed either to false co-ordinations between two WLN terms, or to a general lack of specificity in the search logic. Generous criteria were used in assigning precision failures to these two categories. An analysis undertaken from a different viewpoint would probably have attributed some of these failures to inherent limitations of the WLN index or to limitations of the search logic.

(i) False Co-ordinations (10.9%): These failures occurred in profiles in which the search strategy required the co-ordination (*and* logic) of two WLN symbol strings representing two structural fragments. Both symbols achieved correct matches within the same notation, but the corresponding fragments did not have the required structural relationship.

EXAMPLE 15: The requirement is for compounds containing the β-lactam structure (IV).



(IV)

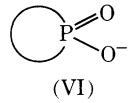


(V)

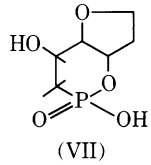
The profile searches for T *followed within 8 characters* by 4 *followed* within 25 characters by NV *followed within 25 characters* by T *followed within 7 characters* by J. The notation which is matched is T3 D5554 1A K EVNVJUTJ FR representing the structure (V).

(ii) Lack of Specificity in the Search Logic (17.9%): The search either does not specify all the features of a relevant structure, or else it does not exclude all the possible irrelevant environments for the required structure.

EXAMPLE 16: The requirement is for phosphinic acid derivatives including ring structures such as (VI).



(VI)



(VII)

The profile searches for: T *followed within 15 characters by* OP *followed within 15 characters by* J *and, in the same notation* [a space] *followed within 2 characters by* O. _A match is achieved with the notation: T56 BO FOPTJ GQ_GO IQ representing structure (VII) which is a phosphonate, not a phosphinate.

EXAMPLE 17: The requirement is for various compounds which contain a *partly saturated* phenanthrene carbon skeleton. The profile searches for L B666 *followed by* _ EO1 and matches with the notation L B666J EO1 . . . for a compound containing the fully unsaturated phenanthrene nucleus. The retrieval of such compounds could easily have been prevented by the inclusion of an *ignore* term: L B666J. This failure occurred because we chose to use a broad search strategy for a specific requirement, whereas the failure shown in Example 16 results from difficulties in searching WLN.

The Data Base (3.9%). Failures attributable to the data base occurred either because of coding errors by the ISI indexers (1.7%) or because we were forced to use imprecise molecular formula terms in searching for structures which are not coded on the WLN file.

The Search System (2.0%). Seven irrelevant articles were retrieved because the experimental search programs did not operate according to specifications.

Precision Failures, The Word Search (12.6%). The proportion of irrelevant articles retrieved by word terms was small and the causes of these failures were not analyzed in detail. The most frequent cause of word search failures was lack of specificity of the search logic as illustrated in the following example.

EXAMPLE 18: The requirement was for articles concerned with aminopyrimidines. The profile contained the term PYRIMIDIN* which was not linked to the "amino" concept, and which retrieved an article on 2,4-dithiopyrimidine.

Precision Failures, The WLN and The Word Search (1.4%). Five documents in the precision failure sample were retrieved by a co-ordinated search for a WLN *and* a word term. The articles were irrelevant because the two concepts were not related in the required manner.

FAILURE ANALYSIS—CONCLUSIONS

The attribution of a retrieval failure to a particular part of the retrieval mechanism does not necessarily imply that the failure could have been avoided by reasonable changes in that part of the system. A proportion of retrieval failures are unavoidable for all practical purposes, notably some of those associated with the communication process or with inherent limitations of a WLN or a natural language index. It will also be recognized that some recall failures can only be prevented at the expense of creating precision failures, and vice versa. A further limitation on real improvements in the retrieval system arises from the interdependence of the different subsystems. Improvements in one subsystem may only result in shifting responsibility for a failure to another part of the retrieval mechanism. For example, improvements in the communication process may result in a more complete statement of the user's interests, but retrieval failures will then result because we are unable to devise a search strategy to cope with these more comprehensive requirements. However, the analysis did reveal several areas in which improvements in retrieval performance are possible.

The Communication Process. This was responsible for more precision (48%) and more recall failures (38.5%) than any of the other three main subsystems. Some of these failures were of the type common to all retrieval services, but others are probably characteristic of a substructure search system in which the capability to undertake a very precise search makes special demands on the accuracy of communication between user and system. Some communication failures are inevitable. The achievement of perfect communication between the user and either the machine or the human search strategist is probably an unattainable ideal. For example, the user may deliberately

refuse to delegate the final selection of relevant material to the search profile (or even to a human searcher). He requests a search which is wider than his actual needs. However, in view of the high proportion of communication failures, the communication process almost certainly offers substantial scope for improvement in retrieval performance. Greater care needs to be taken in identifying a user's requirements and in making sure that the user understands the system he is using.

The WLN Search. It was difficult to draw an exact dividing line between unavoidable failures owing to inherent characteristics of the WLN, and avoidable failures owing to an inadequate search strategy. We elected to assign failures to the search strategy whenever there was the slightest possibility of devising an adequate strategy. As a result, no precision failures and only 3.6% of recall failures were attributed to the WLN index. Undoubtedly, the limitations of WLN and, to a lesser extent, limitations of the search logic were contributing factors with many "search strategy failures." Even so, only 17.5% of recall failures and 33% of precision failures were classified as being due either to the WLN search strategy or the WLN index. At least some of these were avoidable including, for example, the 4.4% of recall failures which occurred because the search strategist failed to allow for alternative notation symbols for a substructure. We concluded that the intrinsic limitations of the WLN index were not a serious source of retrieval failures. As expected, almost half of all recall failures which were attributed to the search strategy were concerned with the retrieval of embedded ring systems.

Improvements in the search logic might have had some effect in reducing the number of substructure search failures. Though the modified UKCIS search programs provided almost all the essential logical facilities for searching a WLN file, the coding of a profile was often a cumbersome task and in some cases this may have discouraged implementation of an optimum search strategy.

The largest group of failures attributed to the data base (19.4%) were recall failures arising from the compound selection rules. The registration of all compounds mentioned in a paper is not feasible and detailed recommendations regarding the data base itself were outside the scope of the analysis. However, it was observed that 69% of failures of this type would have been prevented by registering any compound which appeared in the title of the article, irrespective of whether or not it was new.

The Word Search. Word terms were much under-used in the search profile and only 13% of precision failures were attributed to the word search. The greater use of word terms would undoubtedly have prevented a significant proportion of the recall failures which resulted from the inherent limitations of the WLN index language, from inadequate WLN search strategies, and from the compound selection rules. We found that 52% of all recall failures were, in principle, retrievable by a word search. This is a surprisingly high proportion as one would expect that many structural concepts would never be adequately represented in a word index. The efficacy of a word search might be further increased by greater use by ISI of the generic names for groups of compounds in the subject index.

CURRENCY

Currency comparisons were based on a sample of 32 issues of 15 journals which were covered by CAC & IC during the first nine months of 1970. The countries of origin of the 15 journals were: United States (3 journals), USSR (3), United Kingdom (2), Germany (2), Japan (2), Austria (1), Czechoslovakia (1) and Switzerland (1). The contents of a journal issue may be spread over several issues of *Chemical Abstracts*. Comparisons of the currencies of CAC & IC and *Chemical Abstracts* were therefore based on three papers selected from each of the 32 journal issues. Table VIII shows a scale of relative currencies calculated from both the nominal publication dates and the accession dates recorded by the Radcliffe Science Library, University of Oxford, for the two secondary publications. The base line for the comparison was the library accession dates for the primary journal.

The two secondary publications have identical currencies in terms of their publication dates but the airmail distribution of CAC & IC confers a significant currency advantage within the United Kingdom. The computer outputs from our experimental service were received by the user five weeks later than the accession date of the printed CAC & IC. In a fully operational service, this time-lag could probably be reduced by two weeks.

THE USER'S VIEW

Quantitative measures of service performance must always be considered in conjunction with the users' opinions of the service. Eighty participants replied to a final questionnaire which was distributed at the end of the experiment. In response to the question: Is the ability to look for compounds by their structure of importance to you?, 62% of users replied that it was of major importance, and a further 32% that it was of minor importance. However, 78% of the population also said that a structure search required supplementation by a search of title records.

Users were asked to indicate what they considered to be the major advantages and disadvantages of the ICRS service from a list of possibilities provided in the questionnaire. They were instructed: you may tick none, one or more than one in each list. Their replies are summarized below.

Advantages	
(a) A convenient and time-saving way of keeping up with the literature	41%
(b) Finds the important references central to your interests	39%
(c) Finds background references that you would otherwise have missed	51%
(d) Enables you to search for chemical structures	54%
(e) Tells you of references in obscure or foreign journals	49%
(f) Tells you of papers quickly after their first publication	17%
(g) Other: please give details	7%
Disadvantages	
(a) Covers too few journals	16%
(b) Too many of the references printed out are irrelevant	59%
(c) Too many relevant references are missed	17%
(d) The computer printout gives too little information about the papers	34%
(e) Because it uses Wiswesser notation, the profile is not under your direct control	26%
(f) Does not tell you of references soon enough after their publication	20%
(g) Other: please give details	17%

Participants were also asked: could a service such as this replace your regular scanning of abstract journals (such as Chemical Abstracts) or title lists (such as Cur-

Table VIII. Estimates of Relative Currencies

	Delay, Weeks
Primary journal (accession date)	0
CAC & IC (publication date)	8
Chemical Abstracts (publication date)	8
CAC & IC (accession date)	9
Chemical Abstracts (accession date)	14
Arrival of computer output	14

rent Contents)? A positive reply was given by 34% (27 users) of the population.

The most frequently mentioned disadvantage was the retrieval of irrelevant material. It is possible that the user's intolerance of low precision may have been increased both by the need to consult the printed CAC & IC to decide on the relevance of an article, and by the unattractive appearance of the output which was sent to users. There was a strong correlation between a user mentioning this disadvantage and the output size and precision figures for his profile. The mean output size for the 47 users who mentioned irrelevant material as a disadvantage was 53 references per month, compared with 27 references per month for those who did not do so.

However, of the 27 users who had considered the services to be a possible replacement for their manual searching, 15 mentioned the retrieval of irrelevant material as a disadvantage. This strongly suggests that the user considers irrelevancy as an annoyance rather than a critical disadvantage.

In spite of the relatively low estimates obtained for the coverage and overall recall ratios relative to *Chemical Abstracts,* only 25 users mentioned either missed references or inadequate coverage (or both) as a disadvantage and 39 users considered that one of the advantages of the service was the retrieval of references from foreign or obscure journals. In our view, these responses should not necessarily be interpreted as meaning that recall and coverage performance are of minor importance to the user. When the user scans his output, recall successes and precision failures are self-evident, but recall or coverage failures are not. Furthermore, his judgements are comparative in that his awareness of missed references will depend on the effectiveness of his other search methods. Nevertheless, even with these qualifications, it does appear that the user is less concerned with the coverage of ICRS than we would have predicted on the basis of our quantitative measures.

The only information given in the search outputs was the authors, the title, and the bibliographic citation of a retrieved article, and also any non-WLN search terms responsible for the retrieval of an article. In our view, some form of structural representation, in a form easily recognizable to the user, is a desirable or even an essential feature of a substructive retrieval system. Somewhat surprisingly, therefore, only 34% of users criticized the lack of information in the printout, and only one user added the comment that only an output containing the complete CAC & IC abstract could provide an acceptable service. It is possible that the supply of free copies of CAC & IC during the evaluation compensated for the lack of information in the printout, and moderated user criticism of this deficiency.

From their general comments in the questionnaire, the users' over-all opinion of the experimental service appeared to be qualified approval which is best summed up by the comment of one user: "I have a slight feeling of frustration as I am now aware of the potential usefulness of structure searching, but feel that this cannot be realized until its efficiency is great enough to enable it to replace manual scanning."

CONCLUSIONS

We were concerned with assessing the use of the ICRS data base as the basis for a general-purpose SDI service to a user population whose primary interests were in chemical structures. Some of our conclusions are specific to this data base or to a WLN structure file, but others may have a more general significance for substructure retrieval systems.

In an earlier paper,[1] different types of substructure search were classified according to the problems they presented in the construction of satisfactory WLN search strategies. We expressed the view that we had been able to construct satisfactory WLN searches for the majority of participants. The evaluation of the SDI service has largely confirmed this view. Profiles retrieved an average of 64% of relevant articles in the data base at an estimated mean precision of 32%. The known drawbacks of a WLN structure file were only a minority cause of retrieval failures. A significant proportion of the retrieval failures which were caused by inadequate search strategy design or by the WLN index would probably have been prevented by much greater use of text terms. Concentration of our interest on the WLN index resulted in substantial under-use of the associated bibliographic file of ICRS.

However, satisfactory WLN profiles often required the use of a multiplicity of search terms and of complex logic and their construction required much more of a professional search strategists' time than would (say) a natural language profile. Some simplification in profile construction could be achieved by improvements in the search logic. However, profile complexity was often inherent in the type of requirements presented to the system and in the character of WLN. The complexity of user requirements arises because we were offering a general-purpose SDI service. It is probable that simpler profiles would result if we had made use of the ICRS data base in an alternative manner to provide a specialized current awareness service designed to handle only specific structural 'queries' while another service with broader coverage but less sophisticated indexes caters for the users' more general current awareness needs. It is probable that the ANSA[10] service now being offered by ISI is intended to be used in this manner.

The results of our study of the coverage policy for the ICRS data base were inconclusive. Scanning of 102 journals ensures the coverage of the majority of articles concerned with *new* compounds. However, the data base covered only 49% of relevant articles from serial publications which were located by the user in *Chemical Abstracts.* Many of the articles not covered were concerned with "old" compounds and this throws doubt on the assumption that novelty is a suitable criterion in terms of the user's requirements. On the other hand only a minority of users criticized coverage in their questionnaire replies.

The experiment provided two insights which are applicable to any substructure search service, irrespective of the character of the structure file.

First, searches of a structure file can often be more precise than searches of word files. To take full advantage of the potentially greater precision offered by a substructure search, an exact definition of a user's requirements is necessary. As a result, failures of the retrieval system owing to inadequate communication are likely to be more frequent than for other types of service. These failures are likely to be accentuated if the user is unfamiliar with the method used for representing chemical structures in the structure file.

Second, a conventional subject index, covering nonstructural and structural concepts, should be available for use in conjunction with the structure file. This index can serve several functions. It can be used in searches for purely nonstructural concepts, in searches for groups of compounds which cannot be exclusively defined in structural terms (e.g., alkaloids), and in searches for structure concepts as a means of circumventing the limitations of a particular structure file.

ACKNOWLEDGMENT

We thank the Institute for Scientific Information for their help and interest, and the research workers who comprised our user population. The project was funded by

a grant from the Office for Scientific and Technical Information (UK). The grant holder was Sir Ewart Jones, F.R.S., Professor of Organic Chemistry at the University of Oxford. Notwithstanding these acknowledgments the authors take sole responsibility for the contents of the paper.

LITERATURE CITED

(1) Crowe, J. E., Leggate, P., Rossiter, B. N., and Rowland, J. F. B., "The Searching of Wiswesser Line Notation by Means of a Character-Matching Serial Search," *J. Chem. Doc.* **13,** 85-92 (1973).
(2) Lancaster, F. W., "MEDLARS: Report on the Evaluation of its Operating Efficiency," *Amer. Doc.* **20,** 119-42 (1969).
(3) Crowe, J. E., Leggate, P., Rossiter, B. N., and Rowland, J. F. B., "The Development and Evaluation of a Current Awareness Service Based on the *Index Chemicus Registry System,*" The Experimental Information Unit, University of Oxford, June 1973. Report to the Office for Scientific and Technical Information (U.K.).
(4) Leggate, P., "The Evaluation of Operational Current Awareness Services. A Discussion of Practical and Theoretical Problems of Experimental Design," The Experimental Information Unit, University of Oxford, September 1971. OSTI Rept. No. 5111.
(5) Leggate, P., Smith, B., Stow, J., Williams, M. I., "The *BA Previews* Project: the Development and Evaluation of a Mechanised SDI Service for Biologists," The Experimental Information Unit, University of Oxford, January 1973. OSTI Rept. No. 5139.
(6) Wilcoxon, F., "Individual Comparisons by Ranking Methods," *Biometrics Bull.* **1** (1), 80-3 (1945); and "Probability Tables for Individual Comparisons by Ranking Methods," *Biometrics* **3** (3), 119-22 (1947).
(7) Garfield, E., Editorial in *Current Abstracts of Chemistry and Index Chemicus* **40** (1) (6 January 1971).
(8) Corfield, M. G., Firth, R. J., Fraser, G., Hartley, D., Leggate, P., Norgett, M. M., Riley, J., Rossiter, B. N., "The Liaison Scientist Experiment: A Study of the Provision of Mechanised Current Awareness Services to University Chemists," The Experimental Information Unit, University of Oxford, May 1973. OSTI Rept. No. **5169.**
(9) Rubinstein, R. I., and Qazi, A., "Alternatives to Searching Semantic Surrogates of Chemical Structures," *J. Chem. Doc.* **11,** 110-16 (1971).
(10) Garfield, E., "Introducing ANSA—ISI's Automatic New Structure Alert—A Compound-Retrieval Service for People more interested in Compounds than Retrieval," *Current Contents* **13** (19), 5-10 (9 May 1973).

Use of the Sequence Rule for Indexing Functional Groups in Organic Compounds

PAUL F. HUDRLIK
School of Chemistry, Rutgers University, The State University of New Jersey, New Brunswick, N. J. 08903

Received September 26, 1973

A new method of indexing functional groups in organic compounds is described, utilizing the Cahn-Ingold-Prelog sequence rule. Functional carbon atoms are first classified by "functionality," a measure of the oxidation state, then ordered by means of a modified sequence rule. Substructure searching and other applications are discussed.

Searches for organic compounds containing a particular subunit or substructure are frequently required of organic chemists. Conventional indexes are not well suited for such searches.[1-11] This paper describes a new method of indexing functional groups in organic compounds, based on an index of functional carbon atoms, which should facilitate such searches. The ordering of the index is based on atomic number and oxidation state, thus related functional groups are near one another and the system is easy for the average chemist to use.

BACKGROUND

An organic chemist usually looks for information concerning (1) a specific compound or compounds, (2) a class of related compounds—i.e., compounds containing a given subunit or substructure, or (3) a chemical reaction.[8]

Searching for information about a specific compound is usually not difficult with conventional indexes. Particularly with formula indexes, a given compound will have a unique, predictable place. (In formula indexes, where a number of compounds have a common molecular formula, the addition of a structural formula would be desirable.[8] The use of linear notations[3] or structure codes[1,3,12] for this purpose has been suggested.) Searching for information about substructures or about reactions (other than "name" reactions) is much more difficult.

Since many substructures contain functional groups, and since functional groups are intimately involved in most chemical reactions, the development of a logical index of functional groups may ultimately facilitate reaction searching as well as substructure searching.

A system for classifying or indexing functional groups should satisfy three criteria:[2]

(1) Any functional group should be listed in one unique, predictable place

(2) Related functional groups should be located near one another, as much as possible

(3) The system should be easy to understand and use

Traditional indexes are easy to use but do not meet the first two criteria. Conventional formula indexes and ring indexes are completely unsuitable for functional group