

S45B

# BULLETIN



BRITISH SOCIETY  
OF RHEOLOGY

INFORMATION RETRIEVAL

SPECIAL ISSUE

*Hon. Editor:*

Dr. D. W. HADLEY,  
J. J. Thomson Physical Laboratory,  
Whiteknights,  
Reading,  
Berks.

1968

No. 2

# CONTENTS

Page

|                                                                                             |               |     |        |     |    |
|---------------------------------------------------------------------------------------------|---------------|-----|--------|-----|----|
| "Why Bother?" (Editorial),                                                                  | ...           | ... | ...    | ... | 1  |
| List of Contributors                                                                        | ...           | ... | ...    | ... | 2  |
| "Information Retrieval, and What to Do About It."                                           |               |     | Part 1 | ... | 4  |
| Wilfred Ashworth.                                                                           |               |     | Part 2 | ... | 8  |
| "A Simple Numerical System of Describing and Classifying<br>Rheological Papers."            | ...           | ... | ...    | ... | 12 |
| Joint B.S.R.—I.Chem.E. Working Party.                                                       |               |     |        |     |    |
| "Information Retrieval for the Chemical Engineer."                                          |               |     | ...    | ... | 24 |
| D. W. Hadley,                                                                               |               |     |        |     |    |
| "Some Thoughts on the Preparation of a Thesaurus for Rheological<br>Information Retrieval." | ...           | ... | ...    | ... | 29 |
| D. C-H. Cheng.                                                                              |               |     |        |     |    |
| "Retrieval Reviewed."                                                                       | Edwin Tal.    | ... | ...    | ... | 35 |
| "The Retrieval of Chemical Structure Information."                                          |               |     | ...    | ... | 40 |
| W. J. Wiswesser and A. J. Barnard Jr.                                                       |               |     |        |     |    |
| "N.L.L. to the Rescue,"                                                                     | D. W. Hadley. | ... | ...    | ... | 53 |
| "Classification, Information Retrieval and Rheology"                                        |               |     | ...    | ... | 59 |
| Albert Yanez.                                                                               |               |     |        |     |    |
| "A Commercial Miscellany."                                                                  | ...           | ... | ...    | ... | 66 |
| "Guides for the Preparation of Scientific Papers and Abstracts."                            |               |     | ...    | ... | 69 |

## WHY BOTHER?

The problems of information retrieval have been with us for many years and will continue to trouble us in the future. Several official bodies and large Scientific Institutions, in Britain, U.S.A. and other countries, have initiated enquiries into ways of coping with the ever-increasing volume of scientific information—so why should a small society like B.S.R. devote an enlarged issue of its Bulletin entirely to this single topic?

One of the first fruits of the joint B.S.R.-I. Chem. E. Working Party on Non-Newtonian Flow has been a classification scheme for fluids, to be applied in *Rheology Abstracts*. Almost simultaneously with the announcement of this scheme came notice from I.Chem.E. of the unveiling of its "Keyword Retrieval System," several features of which were obviously of interest to rheologists. Hence here were two articles for the Bulletin.

But what was the point of two articles unless they could be placed in context? Several professional journals are devoted to information science, but these are generally read only by specialists and not by the working scientist—who nevertheless is in need of help if he is to benefit fully from already published information. What was required was a plain man's guide to information retrieval, and who better to provide such a guide than Wilfred Ashworth? By this stage it was clear that there were the makings of a symposium on information retrieval, and the present volume is the result.

Dr. Tal agreed (he insists "was forced" would be a more appropriate term) to comment on the retrieval systems already described, and Dr. Cheng prepared some thoughts on a rheological word list. Neither of these articles is claimed to be definitive, but both have the object of making the reader think for himself about how he could tackle his personal problems of retrieving rheological information.

The problems described so far have been those in which small-scale retrieval systems could be used. Computer-assisted retrieval is, however, becoming increasingly important, and it is hoped that an invited lecture on this topic will be presented at one of the Society's future Conferences (and subsequently reprinted in the Bulletin). The use of computers in information retrieval has been pioneered in U.S.A., and accordingly the assistance of the (U.S.) Society of Rheology was requested to recommend an author for an article describing how computers could assist in retrieval processes. We have been extremely fortunate in obtaining a contribution written by two such experts as Drs. Wiswesser and Barnard. Their paper on Retrieval of Chemical Structure Information is important not only as the first full review of the subject to be published in Britain, but as an example of the manner in which retrieval techniques can be applied to extremely complex situations.

Most of us have frequent recourse to interlibrary loan services, and more often than not the requested journal arrives within a few days

from the National Lending Library at Boston Spa, Yorkshire. What is N.L.L. and how does it operate? The Editor recently attended a short course at Boston Spa, and has attempted to answer not only the questions above but also to pass on some of the useful information he acquired during his visit.

Providing the scientist with the information he requires without it being necessary for him to search through the whole of the literature forms the basis of several commercially operated schemes. Three systems with dissimilar objectives are described briefly, not because each is considered to be exceptionally good of its kind, but as examples of the systems available—and because advertisements for the schemes arrived on the Editor's desk at an appropriate time.

Retrieving information from scientific publications would be greatly simplified if all papers were written clearly and logically, with adequate abstracts. This symposium ends therefore with a report on Guides for the Preparation of Scientific Papers and Abstracts approved at a Conference under the joint sponsorship of Unesco and I.U.P.A.P. Prospective authors who read and take note of this report will earn the gratitude of those of us concerned in the preparation of *Rheology Abstracts*.

Sincere thanks are due to the authors, only two of whom are members of the Society, for devoting so much valuable time to the production of their articles. In addition I am grateful to those members of B.S.R., my colleagues, and Officers of The Institution of Chemical Engineers, The Society of Rheology (U.S.A.), Groupe Français de Rhéologie and The Institute of Physics and the Physical Society, without whose assistance this volume could not have been compiled.

Any criticisms and adverse comments should be addressed to the Editor, and not to the contributors—who have been most patient and forbearing in submitting to modifications of their original manuscripts. Personally I view with favour the idea of occasional enlarged issues of the Bulletin devoted to a single *important* theme, and hope that my successors as Editor will be of the same opinion. Your views may be completely different—so please write if you have any constructive criticisms or ideas. Who knows, some year you might yourself be offered the post of Honorary Editor.

D. W. H.

## THE CONTRIBUTORS

The classification scheme for fluids is the first published report of the Joint Working Party on Non-Newtonian Flow of the British Society of Rheology and the Institution of Chemical Engineers. Any enquiries or comments on the scheme should be sent to Dr. J. M. Smith, Dept. of Chemical Engineering, University College of Swansea, South Wales.

Wilfred Ashworth, who is Librarian and Information Officer at I.C.I. Fibres Ltd., Pontypool, is one of the leading experts in information science. He is a member of the Council of Aslib, and General Editor of "The Handbook of Special Librarianship and Information Work."

Dennis Hadley has edited (and written for) B.S.R. Bulletin since mid-1964. Four years with I.C.I. ("Britain's largest university staff training school") changed his specialisation from high energy nuclear physics to the mechanical properties of polymers, and he now lectures in Chemical Physics at Reading University. His ambition is to survive a full period of editorship and then resume a comparatively normal life.

David Cheng, of Warren Spring Laboratory, Stevenage, has become well known to all members of the Society both for his contributions at conferences and his articles for the Bulletin. He is interested in all aspects of classification, and was joint author of the B.S.R. Membership Survey published in 1967.

Edwin Tal claims to be a professional layabout, with particular interests in the non-constructive aspects of art and science. He is at present engaged on a Work Study programme, which he defines as "appearing to achieve the maximum amount of work compatible with the minimum amount of study." His hobbies cover the whole of experience, except for those parts concerned with his paid employment.

William J. Wiswesser, of Crops Division, Fort Detrick, Maryland, U.S.A., is of course primarily known as the inventor of the Wiswesser Line Notation, which has recently become widely used in the chemical classification retrieval field—but, as is made clear in his joint article with Dr. Barnard, his interests in the subject of chemical structure information retrieval are ever-widening. It is no coincidence that the headquarters of the Chemical Notation Association are situated in Dr. Wiswesser's home town of Reading, Pennsylvania.

A. J. Barnard, Jr., is Manager, Research Analytical and Information Services, J. T. Baker Chemical Co., Phillipsburg, New Jersey, U.S.A. The extent to which he has been responsible for putting his Company in the forefront of those making use of chemical structure information can be judged from the article in which he has co-operated with Dr. Wiswesser. He is Co-editor of the monograph series "Chelates in Analytical Chemistry," serves on the editorial advisory board of "The Micro Chemical Journal" and was from 1951 to 1957 Editor of the "Chemist-Analyst."

Belated thanks to M. Albert Yanez of Centre de Documentation de l'Armement, Paris, for supplying a salutary warning on the dangers inherent in classification and information retrieval. We are grateful that despite unexpected additional duties he nevertheless made time to contribute to this symposium.)

# INFORMATION RETRIEVAL AND WHAT TO DO ABOUT IT

## Part I

Wilfred Ashworth (I.C.I. Fibres Ltd.)

There are three ways of finding out something which is already known; rediscovery; asking someone who knows; or consulting a store of recorded knowledge. Strictly the last two are one, using biological and physical recording media respectively, but for convenience physical records are separated and their consultation has been given the current "with it" name *information retrieval* though, of course, librarians have been doing it for years without benefit of a special term.

Asking someone who knows is chancy, though effective if the right person is asked the right question at the right time. A danger is that the questioner, impressed by the standing and experience of the person consulted, will too readily take "no" for an answer. He fails to realise how limited in content is each biological store and, because little is known about the human retrieval process, may not interrogate correctly. However, at any given time there is some information which is available only as a memory record, and cases can arise in which the very posing of a new question starts thought-processes leading to an answer which, though derived from information in the store, was not recorded there. At present this highly-desirable feature of concocting a reply is found only in systems with human intervention at the retrieval stage, but clearly should be part of any ideal system.

In spite of all that has been written on the subject in the last few years there are only three methods of information retrieval and they are by no means new. Let us assume a perfect store containing every known fact and concept. They will not be discrete items, however, being usually embedded in "documents" where this term is used with the widest interpretation. It is possible to:

- (1) read through every document in conjunction with each question posed, trying to find an answer;
- (2) pre-arrange the documents in some order. They can be named and alphabetically arranged (indexed) or logically arranged according to their subject (classified). The principle is unaltered if a code or symbol is interposed for convenience of operation, or if the arrangement is in the form of a separate list or catalogue. In retrieval it is necessary to decide where a question would have been stored had it been an item of information and at that place in the pre-arranged order the answer should be found;
- (3) pre-decide what topics are mentioned in each document. Each topic is then named by a keyword or "descriptor" and each descriptor is tagged with the serial number of every document in which reference is made to the appropriate topic.

Retrieval involves stating the problem to be solved in terms of a combination or co-ordination of the descriptors employed, and then finding which items in the store appear against every one of these relevant descriptors. This method is the inverse of (2), so a store arranged under descriptors is sometimes called an inverted index.

The retrieval process can be refined by a search strategy in which certain topics are deliberately excluded or in which the presence of chosen descriptors is given a weighting according to their importance in a scale of relevance so that a particular document is only retrieved if specified by a question involving a combination of descriptors totalling more than a predetermined threshold weight. Alternatively, similar weighting can be used to rank the documents retrieved in order of relevance to a particular enquiry.

These three methods are fundamental and, in practice, are used in conjunction with a variety of helpful devices for storage, consultation and matching, such as index cards, edge-punched, body-punched or optical coincidence cards, tabulators, collators, or a computer. It is important not to confuse the apparatus with the method by assuming, for example, that computer retrieval methods are basically new.

Information retrieval is effective if the right, or a useful, answer is found, but is only efficient if the cost of providing a reply is either less than that of finding out in another way or represents a saving when compared with the loss which would be incurred by not having the problem solved.

This all sounds very easy but its practice bristles with difficulties. Quite apart from the troubles arising because ideas have to be communicated by means of imprecise words, problems are rarely simple and in a real-life situation there may be a number of valid, acceptable answers to each question. Part of the available information on a topic is far more often required than all the information, a fact which old-school librarians, so much given to producing lengthy bibliographies, have yet to learn. Further, it often happens that an enquirer does not really know what he wants until he finds it. This is because innovation often results from the correlation of previously unrelated knowledge, when a new idea is sparked-off by a chance discovery in the literature, or when techniques from one area are found to be applicable to quite different fields. For these reasons the various methods described prove not to be self-sufficient, some problems yielding more readily to a technique based on logical arrangement, while others are resolved only by deep co-ordinate indexing methods and others by browsing.

Many books have been written on the problems involved in the analysis and description of documents so that their content can be retrieved without significant loss, on language control, and on file organisation and coding, and much attention has been focused on the potentialities of all or part of the storage and retrieval processes being carried out by machines.

At the present time important progress has been made towards the possibility of improving the subtlety and flexibility of classification by the development of the techniques of facet analysis. These techniques provide rules which enable more ready recognition of the best terms to be used in the classification of a subject so that they represent it faithfully and are capable of being grouped and arranged in a satisfactory order for optimum practical use. On the other hand, for extraction of the complex and detailed information in scientific literature, especially technical reports, many information units have abandoned classification in favour of post-co-ordinate indexing, usually applied with optical coincidence cards. The system is deceptively simple because the problems of control of language are not immediately obvious, and can be used by untrained workers for efficient and rapid retrieval provided that the right conditions are met.

Attempts to replace intellectual effort by the use of machines for analysis of the content of documents at the input stage have so far proved abortive and experimental attention has been diverted from automatic indexing and abstracting projects based on statistical analysis of text towards attempts to recognise semantic patterns. Alternatively, there are proposals for eliminating indexing entirely in favour of processes where the searcher postulates words or phrases which are probably to be found in a document on the wanted subject. A computer is then set to read entire documents searching for the appropriate words or phrases—a reversion to method (1) outlined above. Only two mechanical indexing methods have so far been used on any scale. One is keyword-in-context which automatically prints out entries under every significant, or key, word found in titles of articles. This can be done quickly and relatively cheaply, and though quite unsophisticated provides a better-than-nothing approach where human indexing would be impracticable. The other is the dual dictionary in which descriptors and their associated document numbers are printed side-by-side for easy comparison and combination in pairs.

Computers have also been used to retrieve documents by co-ordination of descriptors. Here the method is identical to co-ordinate indexing operated by more simple apparatus, having the added advantages of speed and the possibility of making a more complex search (using Boolean search strategy and perhaps making multiple searches simultaneously), and the disadvantage of greatly increased cost as equipment and methods stand at present. A promising extension is to combine the process of storage of new information and the comparison of each item with known users' interests, providing a print-out addressed selectively to those users. This selective dissemination of information is an old service in the new guise of mechanical operation, but may be able to offer adequate personal current-awareness as an inexpensive by-product of the preparation of abstracting publications or of systems in which information is stored on magnetic tape or in other machine-readable form.

In this brief outline of how information retrieval works it has been necessary to cut many corners to present the fundamentals in simple language. Those who wish to know more could profitably read one of the books listed below.

The second part of this article will attempt to show what individual scientists can do about information retrieval both personally and by using existing services.

#### References

- <sup>1</sup> B. C. VICKERY. *"Faceted classification, a guide to construction and use of special schemes."* Aslib, 1960.
- <sup>2</sup> J. R. SHARP. *"Some fundamentals of information retrieval."* Andre Deutsch, 1965.
- <sup>3</sup> B. C. VICKERY. *"On retrieval system theory."* 2nd ed., 1965. Butterworth.

# INFORMATION RETRIEVAL AND WHAT TO DO ABOUT IT

## Part II

Wilfred Ashworth (I.C.I. Fibres Ltd.)

The writers of those articles which bewail the information explosion and the impossibility of any individual reading all that appears in even a limited field of interest make the unwarranted assumption that comprehensiveness is necessary and desirable, and ignore the fact that it is relatively easy for anyone to obtain as much information as he has time or opportunity to use. The trick is to select the most profitably-usable information while spending the least amount of effort and time on selection.

Sources of information must first be located. For those with access to a good library or information service this is easy because the librarian will provide guidance, but for those who are not so fortunate there are quite a number of books on how to find out, some dealing with specific fields of interest. Any public librarian or a good book-seller should be able to advise on these, but amongst the best for the scientist or engineer is one by R. J. P. Carey,<sup>6</sup> and for the chemist R. T. Bottle<sup>7</sup> or C. R. Burman,<sup>8</sup> while G. Chandler<sup>9</sup> covers sources of information in general.

Every scientist will find he needs to read a few periodicals thoroughly, not only for their articles but also for news and personalities, new products, conferences and meetings. A larger group of periodicals, about three times as many, will be found to repay the time spent on quickly scanning through them. Trial and error will soon show which periodicals fall into which groups. It is good practice also to dip from time to time into a few periodicals well outside the range of one's normal interests. Browsing is a gamble, but a little information flutter adds excitement to the game and the prizes are there to be won even though the odds are long. Current awareness services (abstracts, title-lists of current papers and library bulletins) make it possible to cover a wide range of material otherwise inaccessible because of lack of time or of ready availability. These publications are usually broadly grouped under sub-topics or may be limited to a small subject-area, so that consultation need not take up too much time. Complete familiarity with a journal speeds scanning considerably so it is advisable to shop around until one's own "best buy" is found, and then to stick to it. It cannot be too firmly stressed that the mental state of scanning is not at all conducive to remembering, so immediate note should be made of everything which *interests* the scanner even though at the time he may see no possible use for it. The mind which has been intrigued by an idea is pre-disposed to be able to find an application if the right opportunity arises, but if no note is made then the information may not be recoverable when needed. A full bibliographical reference must accompany every noted idea or article. This reference consists of the name and

initials of the author, title of the article, complete title of the periodical in which it appears, volume number, issue number and date, and all the page numbers on which any part of the article is found. In the case of pamphlets or books, the issuing body or publisher is needed, and for reports any serial number as well.

How many such notes are made depends on the character of the individual and on his field of interest. If the number is not too great then it may be satisfactory enough to write them in a special book as the ideas are located and to rely on memory to find them again when they are needed. An obvious simple improvement is to use a loose-leaf book so that the notes can be grouped under a few appropriate headings. However, the number of articles worth noting can easily grow and face the scientist with the need to make a more organised approach to their retrieval. It is worth checking, before starting to make a personal index, that no library or abstracting service already adequately indexes the material in which one is interested. Some people find a straightforward alphabetical index quite adequate for the number of entries concerned and a simple card index the best storage medium, but there are snags in choosing headings which are co-extensive with the subject matter of documents without being cumbersome. A few helpful hints will be found in a chapter of a book by C. D. Needham<sup>9</sup> but for those who fall to the undoubted fascination of subject-indexing E. J. Coates<sup>10</sup> has all the answers. An ingenious "ramifying" keyword indexing system, based on rather different principles and suitable for personal indexes, is described by J. E. Holmstrom.<sup>10</sup> It is easy to use once the underlying idea has been grasped.

While generally satisfactory for retrieval of whole articles or broad topics the foregoing methods do not easily provide the facilities needed to pick out all the separate ideas in complex papers. For this purpose classification in depth or co-ordinate indexing would be more appropriate. It is possible to use existing classification schemes or published lists of descriptors such as those of the Engineers Joint Council or the American Institute of Chemical Engineers, but only a fraction of their available classes or terms will be needed so it is probably better, and it is certainly cheaper, to create a home-made system. Choice of equipment is largely a matter of personal preference. Edge-notched cards are inexpensive and very effective in small-scale systems where relatively few cards have to be sorted through. Cards for the Uniterm system are even cheaper and this method also works well on a small scale. Optical coincidence cards are quickest to use when making a search, and permit complex combinations of topics to be made with utmost simplicity, but they cost a little more and this system does not adapt so readily to small-scale operation although it has been used quite effectively. All these systems are described in easy language in reference.<sup>2</sup> However, anyone thinking of making a scheme can learn much from the way in which others have tackled the job. Two excellent articles by A. E. Cawkell<sup>11</sup> (edge-notched cards), and R. M. A. McClelland and W. W. Mapleson<sup>12</sup> (optical coincidence

cards) illustrate the right approach to the construction of retrieval systems.

The methods so far described will enable scientists to keep abreast of the more important new knowledge, but this is not the whole picture. There will still be need for retrospective searches when new projects are planned or unexpected problems arise in the course of research. Colleagues may be able to help, but the balance of probability is that answers will be found more certainly from literature. Here the precise formulation of questions is important, though frequently neglected. A librarian called upon for assistance will need not only to know as exactly as possible what is required but also the level of information applicable to the problem, the enquirer's previous knowledge of the subject, whether all available information or only a representative paper is wanted, and how far back in time the search should be carried. If the question cannot be exactly specified, as when a person knows what he wants to do but has little idea how to do it, then the problem must be talked over freely with the librarian so that he is aware of the full background to the enquiry and is thus able to assess the potential value of any information found.

Most scientists will have to carry out some searches for themselves so the following limits may be helpful. Abstracting and indexing services are not perfect and it cannot be taken for granted that a search through any one set will reveal all available papers. If there are overlapping services and complete information is sought then all must be tried. It is profitable to spend some time on making oneself familiar with the construction of indexes to abstracting journals and trying to predict, from experience in use, the type of headings the indexer is likely to choose for different topics. Papers on the same subject may, because of human failures in indexing, not always be put under the same terms in the subject index. Once a relevant paper has been located a check of the author index may reveal other papers by the same author(s) and on the same subject, although not indexed as such. Lists of references appended to useful articles found often quote further useful papers. Citation indexes make use of this known fact. In searching these the enquirer begins with a known relevant paper and from the index learns of other articles which have cited that paper as a reference. Citation indexes are new and expensive and, as yet, not widely available in this country but there is no doubt that they can lead to some information which conventional indexes would not have uncovered, and are therefore complementary to them.

There have in recent years been several interesting investigations of the ways in which scientists find information. That by J. Martyn<sup>12</sup> is typical and shows that scientists and engineers view the finding of information in a very different way from that of librarians, setting in almost reverse order of value the various possible approaches to the discovery of known facts. To read one of these surveys and to recognise one's own shortcomings could be a salutary exercise.

## References

- <sup>4</sup> R. J. P. CAREY. *Finding and using technical information*. Edward Arnold, 1966.
- <sup>5</sup> R. T. BOTTLE. *Use of the chemical literature*. Butterworths, 1962.
- <sup>6</sup> C. R. BURMAN. *How to find out in chemistry*. Pergamon Press, 1965.
- <sup>7</sup> G. CHANDLER. *How to find out. A guide to sources of information for all*. Pergamon Press, 1963.
- <sup>8</sup> C. D. NEEDHAM. *Organising knowledge in libraries*. Andre Deutsch, 1964.
- <sup>9</sup> E. J. COATES. *Subject catalogues: headings and structure*. Library Association, 1960.
- <sup>10</sup> J. E. HOLMSTROM. *Facts, files and action. Part 2. Filing, indexing and circulation*. Chapman and Hall, 1953.
- <sup>11</sup> A. E. CAWKELL. Classification and retrieval of technical information.  
 1. Considerations leading to the use of notched cards.  
 2. The notched card system in practice. *Wireless World*, vol. 68, no. 8 and 9, August and September, 1962, pp. 352-7, 432-4.
- <sup>12</sup> R. M. A. McCLELLAND and W. W. MAPLESON.  
 The use of feature cards and their application in the indexing of anaesthetic literature. *Proceedings of the Royal Society of Medicine*, vol. 58, March 1965, pp. 203-8.
- <sup>13</sup> J. MARTYN. An examination of citation indexes. *Aslib Proceedings*, vol. 17, no. 6, June 1965, pp. 184-196.
- <sup>14</sup> J. MARTYN. *Report of an investigation on literature searching by research scientists*. Aslib, 1964.

## A SIMPLE NUMERICAL SYSTEM OF DESCRIBING AND CLASSIFYING RHEOLOGICAL PAPERS

### Introduction

In 1964 the Institution of Chemical Engineers and the British Society of Rheology set up a joint working party, to study research needs in the field of non-Newtonian fluids.

The Working Party began by trying to assess the existing state of knowledge in various parts of the field. It soon became clear that a great deal of specialised information was available, but that it was dispersed through thousands of published papers which have never been organised or systematically classified.

Some method of describing and classifying papers on non-Newtonian fluids was obviously needed, to simplify the task of reviewing the extent of existing information on various topics. The Working Party therefore developed such a system for its own specialised purpose, but it soon became apparent that the system would be of use in other connections also.

This paper describes the system in its present form, which has been reached after two years of development and one year of use by the Working Party. Further modification will probably take place in the light of further experience by other users, but meanwhile the system is sufficiently well-established to bring it to the notice of a wider public.

### Basis of the System

The system is based upon a formalised question-and-answer technique. It would be possible to find out everything about a paper only by asking an unlimited number of questions, but if the questions are well chosen a small number of questions will provide most of the really essential information. The following 10 questions form the basis of the present system:

- (A) What type of paper is it (theoretical, experimental, etc.)? (5)
- (B) In what language is it written? (6)
- (C) What type of fluid was involved? (9)
- (D) What were its rheological properties? (7)
- (E) What kind of flow was involved (laminar, turbulent, etc.)? (5)
- (F) What were the boundary conditions? (4)
- (G) What other variables and effects were important? (11)
- (H) What kind of theoretical treatment was followed (5)
- (J) What sort of apparatus was used? (4)
- (K) What industrial process was involved? (13)

If each question had to be answered in detail, a very large number of possible answers would be necessary. By accepting incomplete answers, which give the essential information while omitting the details, the number of possible answers has been reduced drastically.

The numbers in brackets after each of the above questions indicates the number of answers to each question permitted by the present system.

The number of permissible answers per question varies from four (Questions F and J) to 13 (Question K), with a grand total of 69 answers. These answers are set out in Appendix 1, where they are listed in the left-hand column, and are numbered in sequence from 1 to 69. Each answer is expressed as either a single word or a short phrase. It has therefore been necessary to add a further column which helps to explain what is meant by the terse "answer" in the left-hand column; this right-hand column gives examples of the kind of topic included in each "answer," but it is not intended to be exhaustive.

Thus, for example, to Question F, "What were the boundary conditions?" there are only four possible answers. Of these, No. 34 is "Non-uniform pipe." The right-hand column explains that "non-uniform" is to be interpreted as meaning either with a varying cross-section or with sharp bends, and that this includes entry, exit, orifices, jets, Venturi meters, and valves.

To assist a classifier to find his way through the system a cross-index has been prepared, which appears in Appendix 2. This gives in alphabetical order all the terms appearing in both columns of Appendix 1, with the relevant numbers. Some additional terms have also been included in this alphabetical list. The classifier can see at a glance by looking under "B" in this appendix, for example, that ball mills are included in No. 62, bearings in No. 35, bends in No. 34, Bingham plastics in No. 24, and so on.

### Operation of the System

As an example of the system in operation, it will be applied to the paper, "Diffusion in slurries and in non-Newtonian fluids," by S. B. Clough and others. (*J. Amer. Inst. Chem. Engrs.*, 1962, 8, 346.) This would be classified as follows:

- (A) 2, (B) 6, (C) 13, 16, (D) 22, (E) 28, (F) 33, (G) 40, 45, 47, (H) 50, (J) 53, (K)—.

The letters in brackets are added for the sake of convenience, but are not strictly necessary. This series of numbers yields the following information about the paper:

2. Primarily a theoretical paper
6. Written in English.
- 13, 16 Concerned with polymer solutions and with liquid/solid systems (and no other numbers from 21-27)
22. The only rheological property of these fluids considered in the paper is the variation of viscosity with rate of shear
28. The fluids are considered to be undergoing laminar flow . . .
33. . . . along a uniform pipe or passage
40. Concentration is an important variable

45. The fluids are changing their nature as they flow
47. Wall effects are important
50. The behaviour of the fluids is discussed in terms of molecular theory
53. The paper is not wholly theoretical, since some experiments are performed with a mechanical apparatus  
(No numbers from 57 to 69.) The paper is not concerned with any industrial process.

Thus this simple series of numbers gives a fair indication of the contents of the paper, and would need to be supplemented by only a brief abstract to give a complete picture.

These numbers can be transferred to a specially designed punch-card, shown in Appendix 3. The reference of the original paper, and, if necessary, of an abstract appearing in another journal, can also be recorded on this card by punching the appropriate holes. Filing information in this way is quicker than copying details on to ordinary filing cards, and is much more useful for retrieval purposes.

### Advantages of the System

The system was developed in the first place to provide a method of recording information about published papers, so that all the papers dealing with various aspects of rheology could be located quickly. With the aid of the punched cards it is the work of only a few moments to locate all the papers dealing with, for example, "theoretical studies of viscoelastic effects in lubrication," or, "birefringence experiments on time-dependent fluids in accelerating flow," or with any other field in which the user may become interested.

An additional advantage is that it should prove of assistance to authors of papers and abstractors. Published abstracts, whether prepared by the original authors or by separate abstractors, tend to be very variable in the amount of useful information they contain. Abstractors using the present system will be led to give a uniform amount of essential information in terms of code numbers. This will both simplify the abstractors' task, and assist the readers.

### Adoption in "Rheology Abstracts"

The British Society of Rheology has decided to adopt this system for use in its journal, *Rheology Abstracts*. As from the first issue in 1968, every abstract dealing with fluids will give the relevant classification numbers from the system. Punch-cards will be obtainable from the British Society of Rheology, so that readers of *Rheology Abstracts* wishing to file information on cards will be able to do so. If the demand is sufficient, it may ultimately be possible to supply ready punched cards along with each issue of *Rheology Abstracts*.

Work on the development of a parallel scheme of classification for papers dealing with solids is proceeding. It is intended that this also will be adopted in *Rheology Abstracts* when it is completed.

The present adoption of the system in *Rheology Abstracts* is on a provisional basis, so that its merits can be assessed in use. Comments from readers and users of the system will be welcomed. These will enable the Society to review the system after it has been in use for a trial period, and so to make any necessary alterations before it is finally adopted in a permanent form.

## APPENDIX 1

### System of Classification of Papers on Non-Newtonian Flow (Developed by the Joint Working Party of the Institution of Chemical Engineers and the British Society of Rheology)

| CLASSIFICATION                         | EXAMPLES OF TOPICS INCLUDED<br>(AND DEFINITIONS)                                                                                                      |
|----------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| A. Type of Paper                       |                                                                                                                                                       |
| 1. Review                              | Emphasis on review of previously published work. General and educational papers.                                                                      |
| 2. Theoretical                         | Emphasis on theory.                                                                                                                                   |
| 3. Experimental                        | Emphasis on results of experiments.                                                                                                                   |
| 4. Instrumentation                     | Emphasis on apparatus or experimental techniques.                                                                                                     |
| 5. Data                                | Values of physical properties suitable for reference. Methods of processing, correlating and predicting data.                                         |
| B. Language                            |                                                                                                                                                       |
| 6. English                             |                                                                                                                                                       |
| 7. French                              |                                                                                                                                                       |
| 8. German                              |                                                                                                                                                       |
| 9. Russian                             |                                                                                                                                                       |
| 10. Other                              |                                                                                                                                                       |
| 11. English summary                    |                                                                                                                                                       |
| C. Type of Fluid                       |                                                                                                                                                       |
| 12. Simple fluid                       | Single-phase non-polymeric liquids. Solutions of non-polymeric substances. Gases, supercritical fluids.                                               |
| 13. Polymer solution                   | Viscosity-index-improved mineral oils.                                                                                                                |
| 14. Polymer melt                       | Liquid elastomers, silicone fluids.                                                                                                                   |
| 15. Liquid/liquid                      | (i.e. two-phase liquid/liquid systems.) Emulsions.                                                                                                    |
| 16. Liquid/solid                       | Slurries, pastes, suspensions other than colloidal suspensions.                                                                                       |
| 17. Colloid                            | Gels, sols, greases.                                                                                                                                  |
| 18. Gas/solid                          | Fluidised beds.                                                                                                                                       |
| 19. Gas/liquid                         | So-called "two-phase flow," foams, bubbles in liquids.                                                                                                |
| 20. Biological fluid                   |                                                                                                                                                       |
| D. Rheological Properties              |                                                                                                                                                       |
| 21. Newtonian                          | Approximately Newtonian.                                                                                                                              |
| 22. Shear-rate dependence of viscosity | Pseudoplasticity (shear thinning), dilatancy (shear thickening), structural viscosity, "generalised Newtonian liquids," power law, stress dependence. |
| 23. Time dependence of viscosity       | Thixotropy, rheopexy (negative thixotropy), false-body, hysteresis, work-hardening.                                                                   |
| 24. Yield stress                       | Bingham plastics, false-body.                                                                                                                         |

|                                  |                                                                                                                                                                                                                                                               |
|----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 25. Viscoelasticity              | Memory, relaxation, retardation, equations of state (constitutive equations), spectrum, normal stress, Weissenberg effect, complex modulus, dynamic viscosity, compliance.<br>Maxwell liquids and many other idealised liquids named after their originators. |
| 26. Anisotropy                   | (i.e. different properties in different directions.)<br>Orientation of molecules or particles, birefringence.                                                                                                                                                 |
| 27. Fracture                     | Melt-fracture, shear-fracture, break-up, fatigue.                                                                                                                                                                                                             |
| E. Flow Character                |                                                                                                                                                                                                                                                               |
| 28. Laminar                      | Streamline flow.                                                                                                                                                                                                                                              |
| 29. Complex                      | Laminar/turbulent transition, critical Reynolds number, instability, Taylor vortices, secondary flow, combinations of two or more simple flows.                                                                                                               |
| 30. Turbulent                    | Friction factor, drag reduction.                                                                                                                                                                                                                              |
| 31. Sinusoidal                   | Oscillatory rheometry, alternating shear.                                                                                                                                                                                                                     |
| 32. Accelerating                 | Decelerating flow, water hammer, non-sinusoidal oscillations, transient flow.                                                                                                                                                                                 |
| F. Boundary Geometry             |                                                                                                                                                                                                                                                               |
| 33. Uniform pipe                 | (i.e. without sharp bends and with a constant cross-section, which may be circular or non-circular.)<br>Ducts, channels, capillaries, passages.                                                                                                               |
| 34. Non-uniform pipe             | (i.e. with sharp bends or with a varying cross-section, which may be circular or non-circular.)<br>Entry, exit, orifices, jets, Venturi meters, valves, nozzles.                                                                                              |
| 35. Moving boundary              | Rotational viscometers and rheometers, rotational mixers and mills, bearings, compression tests, penetration tests. Flow round single particles or past solid objects.                                                                                        |
| 36. Granular boundary            | Sedimentation, erosion, porous diffusion, filtration, associated particles.                                                                                                                                                                                   |
| G. Variables and Effects         |                                                                                                                                                                                                                                                               |
| 37. Pressure                     | Vacuum.                                                                                                                                                                                                                                                       |
| 38. Temperature                  | Temperature gradient, heat transfer, viscosity index, isentropic, adiabatic.<br>Constant volume.                                                                                                                                                              |
| 39. Density                      | Concentration gradient, particle migration.                                                                                                                                                                                                                   |
| 40. Concentration                | Einstein equation, Einstein coefficient.                                                                                                                                                                                                                      |
| 41. Structure                    | Particle size, shape and orientation.<br>Molecular arrangement, cross-linking, Brownian motion.                                                                                                                                                               |
| 42. Molecular weight             |                                                                                                                                                                                                                                                               |
| 43. Surface rheology             | Surface viscosity, interfacial stresses, surface tension, very thin films, adsorption, adhesion.                                                                                                                                                              |
| 44. Electrochemical effects      | Zeta potential, electrophoresis, dipole moment, polar, non-polar.                                                                                                                                                                                             |
| 45. Change of nature during flow | Chemical change, phase change, polymerisation, degradation, mass transfer, irreversible breakdown under shear.                                                                                                                                                |
| 46. Ageing                       | Long-term changes, curing, oxidation, evaporation.                                                                                                                                                                                                            |
| 47. Boundary effects             | Surface roughness, slip, wall effects.                                                                                                                                                                                                                        |

|                                 |                                                                                                |
|---------------------------------|------------------------------------------------------------------------------------------------|
| <b>H. Theoretical Treatment</b> |                                                                                                |
| 48. Continuum mechanics         | Fluid mechanics, boundary layer theory.                                                        |
| 49. Thermodynamics              | Statistical mechanics.                                                                         |
| 50. Molecular behaviour         | Liquid state theory, kinetic theory of fluids, qualitative explanation of molecular behaviour. |
| 51. Analogies                   | Springs, dashpots, mechanical analogies, electrical analogies.                                 |
| 52. Dimensional analysis        | Non-dimensional parameters, similitude theory.                                                 |
| <b>J. Apparatus</b>             |                                                                                                |
| 53. Mechanical                  | Conventional viscometers and rheometers, including capillary type.                             |
| 54. Electromagnetic             |                                                                                                |
| 55. Ultrasonic                  | Acoustic, piezo-electric, magnetostriction.                                                    |
| 56. Optical                     | Birefringence, schlieren, flow visualisation, tracers.                                         |
| <b>K. Industrial Process</b>    |                                                                                                |
| 57. Adhesion                    | Polymerisation.                                                                                |
| 58. Chemical processing         | (i.e. forced flow through a die, Drawing.)                                                     |
| 59. Extrusion                   | Cooling, melting, freezing.                                                                    |
| 60. Heating                     | Friction, wear, tribology.                                                                     |
| 61. Lubrication                 | Milling, mastication, grinding, emulsification, dispersion.                                    |
| 62. Mixing                      | Filtration, sedimentation, centrifuging.                                                       |
| 63. Separation                  | Pressing, casting, diecasting, vacuum forming, injection moulding, sintering.                  |
| 64. Moulding                    | Pumping, channel flow.                                                                         |
| 65. Pipe flow                   | Calendering, doctoring.                                                                        |
| 66. Rolling                     | Fibre manufacture.                                                                             |
| 67. Spinning                    | Atomising.                                                                                     |
| 68. Spraying                    | Printing.                                                                                      |
| 69. Spreading                   |                                                                                                |

## APPENDIX II

### Cross Index

|        |                                      |
|--------|--------------------------------------|
| 32     | ACCELERATING FLOW                    |
| 55     | Acoustic apparatus                   |
| 57     | ADHESION (as an industrial process)  |
| 43     | Adhesion (as a phenomenon)           |
| 38     | Adiabatic change                     |
| 43     | Adsorption                           |
| 46     | AGEING                               |
| 31     | Alternating shear                    |
| 51     | ANALOGIES                            |
| 26     | ANISOTROPY                           |
| 68     | Atomising                            |
| 62     | Ball mills                           |
| 35, 61 | Bearings                             |
| 34     | Bends                                |
| 24     | Bingham plastics                     |
| 20     | BIOLOGICAL FLUIDS                    |
| 26     | Birefringence                        |
| 56     | Birefringence measurement apparatus  |
| 47     | BOUNDARY EFFECTS                     |
| 48     | Boundary layer theory                |
| 45     | Breakdown under shear (irreversible) |
| 27     | Break-up                             |
| 41     | Brownian motion                      |

|        |                                         |
|--------|-----------------------------------------|
| 19     | Bubbles in liquids                      |
| 66     | Calendering                             |
| 33     | Capillaries                             |
| 53     | Capillary viscometry                    |
| 29, 34 | Cascade flow                            |
| 64     | Casting (as an industrial process)      |
| 63     | Centrifuging (as an industrial process) |
| 45     | CHANGE OF NATURE during flow            |
| 33     | Channels                                |
| 65     | Channel flow (as an industrial process) |
| 44     | Charged particles                       |
| 45     | Chemical change                         |
| 58     | CHEMICAL PROCESSING                     |
| 33     | Coiled pipes or tubes                   |
| 17     | COLLOIDS                                |
| 29     | Combined rotation and axial flow        |
| 29     | COMPLEX FLOW                            |
| 25     | Complex modulus                         |
| 25     | Compliance                              |
| 35     | Compression tests                       |
| 40     | CONCENTRATION (as a variable)           |
| 40     | Concentration gradient                  |
| 35, 53 | Cone-and-plate viscometers              |
| 53     | Coni-cylindrical viscometers            |
| 39     | Constant volume (experiments at)        |
| 25, 48 | Constitutive equations                  |
| 48     | CONTINUUM MECHANICS                     |
| 34     | Convergence (flow through)              |
| 60     | Cooling (as an industrial process)      |
| 5      | Correlation of data                     |
| 47     | Couette correction                      |
| 35     | Couette flow                            |
| 29     | Critical Reynolds number                |
| 41     | Cross-linking (of polymer molecules)    |
| 46     | Curing                                  |
| 35     | Cylinders, flow round                   |
| 31, 35 | Cylindrical container, oscillating      |
| 51     | Dashpots (used in an analogy)           |
| 5      | DATA                                    |
| 32     | Decelerating flow                       |
| 46     | Degradation with ageing                 |
| 45     | Degradation with shearing               |
| 39     | DENSITY (as a variable)                 |
| 34, 59 | Die, flow through                       |
| 64     | Diecasting                              |
| 36, 40 | Diffusion                               |
| 22     | Dilatancy                               |
| 52     | DIMENSIONAL ANALYSIS                    |
| 44     | Dipole moments                          |
| 31, 35 | Disc, oscillating                       |
| 34     | Discharge coefficients                  |
| 62     | Dispersion (as an industrial process)   |
| 34     | Divergence (flow through)               |
| 66     | Doctoring                               |
| 30     | Drag reduction                          |
| 59     | Drawing                                 |
| 35     | Drops, liquid, motion of single         |
| 36     | Drops, liquid, motion of associated     |
| 33     | Ducts                                   |
| 25     | Dynamic viscosity                       |
| 30     | Eddies                                  |
| 1      | Educational papers                      |

|        |                                                           |
|--------|-----------------------------------------------------------|
| 40     | Einstein equation (coefficient)                           |
| 25     | Elastico-viscous fluids                                   |
| 14     | Elastomers (liquid)                                       |
| 51     | Electrical analogies                                      |
| 44     | <b>ELECTROCHEMICAL EFFECTS</b>                            |
| 54     | <b>ELECTROMAGNETIC APPARATUS</b>                          |
| 44     | Electrophoresis                                           |
| 62     | Emulsification                                            |
| 15     | Emulsions                                                 |
| 34     | End effects (Engineering sense, i.e. at pipe entry, etc.) |
| 47     | End effects (Rheological sense, i.e. at walls, etc.)      |
| 34     | Enlargement (of a pipe, etc.)                             |
| 34     | Entry                                                     |
| 48     | Equations of motion                                       |
| 25, 48 | Equations of state                                        |
| 36     | Erosion                                                   |
| 46     | Evaporation (as a phenomenon)                             |
| 60     | Evaporation (as an industrial process)                    |
| 34     | Exit                                                      |
| 34     | Expansion (of a pipe, etc.)                               |
| 3      | <b>EXPERIMENTAL</b>                                       |
| 32     | Explosions (flow due to)                                  |
| 59     | <b>EXTRUSION</b> (as an industrial process)               |
| 53     | Extrusion rheometry                                       |
| 23, 24 | False-body                                                |
| 27     | Fatigue (in fluid systems)                                |
| 67     | Fibre spinning                                            |
| 43     | Films, thin                                               |
| 36     | Filtration (as a phenomenon)                              |
| 63     | Filtration (as an industrial process)                     |
| 41, 44 | Flocculation                                              |
| 56     | Flow visualisation                                        |
| 48     | Fluid mechanics                                           |
| 18     | Fluidised beds                                            |
| 19     | Foams                                                     |
| 27     | <b>FRACTURE</b> (of liquids)                              |
| 60     | Freezing (as an industrial process)                       |
| 61     | Friction (industrial application)                         |
| 30     | Friction factor (in turbulent pipe flow)                  |
| 12     | Gases                                                     |
| 19     | <b>GAS/LIQUID</b> systems                                 |
| 18     | <b>GAS/SOLID</b> systems                                  |
| 17     | Gels                                                      |
| 22     | "Generalised Newtonian" fluids                            |
| 36     | <b>GRANULAR BOUNDARY</b>                                  |
| 62     | Grinding                                                  |
| 17     | Greases                                                   |
| 38     | Heat transfer                                             |
| 60     | <b>HEATING</b> (as an industrial process)                 |
| 33     | Helical pipes or conduits                                 |
| 23     | Hysteresis                                                |
| 35     | Indentation                                               |
| 29, 32 | Inertial effects in fluids                                |
| 64     | Injection moulding                                        |
| 34     | Inlet region                                              |
| 29     | Instability                                               |
| 4      | <b>INSTRUMENTATION</b>                                    |
| 43     | Interfacial stresses                                      |
| 21     | Intrinsic viscosity                                       |
| 38     | Isentropic change                                         |
| 38     | Isothermal conditions                                     |
| 34     | Jets                                                      |

|        |                                                    |
|--------|----------------------------------------------------|
| 50     | Kinetic theory of fluids                           |
| 28     | LAMINAR FLOW                                       |
| 29     | Laminar/turbulent transition                       |
| 15     | LIQUID/LIQUID systems (two-phase)                  |
| 12     | Liquids, simple (i.e. non-polymeric, single-phase) |
| 16     | LIQUID/SOLID systems                               |
| 50     | Liquid state theory                                |
| 46     | Long-term changes                                  |
| 61     | LUBRICATION (as an industrial process)             |
| 43     | Lubrication (fundamentals)                         |
| 55     | Magnetostriction apparatus                         |
| 45     | Mass transfer                                      |
| 62     | Mastication                                        |
| 25     | Maxwell liquid                                     |
| 51     | Mechanical analogies                               |
| 53     | MECHANICAL APPARATUS                               |
| 27     | Melt-fracture                                      |
| 60     | Melting (as an industrial process)                 |
| 25     | Memory (material with)                             |
| 40     | Migration of particles                             |
| 35, 62 | Mills                                              |
| 35, 62 | Mixers                                             |
| 62     | MIXING (as an industrial process)                  |
| 41     | Molecular arrangement                              |
| 50     | MOLECULAR BEHAVIOUR                                |
| 42     | MOLECULAR WEIGHT                                   |
| 64     | MOULDING (as an industrial process)                |
| 35     | MOVING BOUNDARY                                    |
| 45     | Nature, change of, during flow                     |
| 21     | NEWTONIAN behaviour (exact or approximate)         |
| 52     | Non-dimensional parameters                         |
| 38     | Non-isothermal conditions                          |
| 44     | Non-polar (molecules)                              |
| 32     | Non-sinusoidal oscillations                        |
| 34     | NON-UNIFORM PIPE                                   |
| 25     | Normal stress                                      |
| 34     | Nozzles                                            |
| 56     | OPTICAL APPARATUS                                  |
| 26     | Orientation (of molecules or particles)            |
| 34     | Orifices                                           |
| 34     | Orifice-plate flow meters                          |
| 31, 32 | Oscillating flow                                   |
| 46     | Oxidation                                          |
| 35     | Parallel plate viscometry                          |
| 52     | Parameters (non-dimensional)                       |
| 40, 41 | Particle interaction                               |
| 40     | Particle migration                                 |
| 41     | Particle orientation                               |
| 41     | Particle shape                                     |
| 35     | Particle, single, flow round                       |
| 41     | Particle size                                      |
| 33     | Passages                                           |
| 16     | Pastes                                             |
| 53     | Penetrometers                                      |
| 35     | Penetration tests                                  |
| 36     | Perforated plates                                  |
| 45     | Phase change                                       |
| 55     | Piezo-electric apparatus                           |
| 33, 34 | PIPE (see tube)                                    |
| 65     | PIPE FLOW (as an industrial process)               |
| 33     | Planes, parallel, flow between                     |
| 33     | Plate, inclined, gravity flow                      |

|        |                                                 |
|--------|-------------------------------------------------|
| 29     | Plates, parallel, flow between                  |
| 33     | Poiseuille flow                                 |
| 44     | Polar (molecules)                               |
| 14     | POLYMER MELTS                                   |
| 13     | POLYMER SOLUTIONS                               |
| 45     | Polymerisation (as a phenomenon)                |
| 58     | Polymerisation (as an industrial process)       |
| 36     | Porous diffusion                                |
| 22     | Power law                                       |
| 5      | Prediction of data                              |
| 64     | Pressing (as an industrial process)             |
| 37     | PRESSURE (as a variable)                        |
| 37     | Pressure drop                                   |
| 37     | Pressure measurement                            |
| 37     | Pressure, viscometry at high                    |
| 69     | Printing (as an industrial process)             |
| 5      | Processing of data                              |
| 22     | Pseudoplasticity                                |
| 22     | Pseudoplastics                                  |
| 32     | Pulsating flow                                  |
| 65     | Pumping (as an industrial process)              |
| 36     | Relative motion of multiple particles to fluids |
| 35     | Relative motion of single particles to fluids   |
| 25     | Relaxation                                      |
| 25     | Retardation                                     |
| 1      | Reviews                                         |
| 33, 53 | Rheometers, capillary                           |
| 35, 53 | Rheometers, rotational                          |
| 23     | Rheopexy                                        |
| 35, 62 | Roller grinders                                 |
| 35, 62 | Roller mills                                    |
| 66     | ROLLING (as an industrial process)              |
| 35, 62 | Rotational mixers and mills                     |
| 35, 53 | Rotational viscometers and rheometers           |
| 47     | Roughness                                       |
| 56     | Schlieren photography                           |
| 29     | Secondary flow                                  |
| 36, 63 | Sedimentation                                   |
| 63     | SEPARATION (as an industrial process)           |
| 27     | Shear fracture                                  |
| 22     | SHEAR-RATE DEPENDENCE (of viscosity)            |
| 22     | Shear-thickening                                |
| 22     | Shear-thinning                                  |
| 14     | Silicone fluids                                 |
| 52     | Similitude theory                               |
| 12     | SIMPLE FLUID                                    |
| 28     | Simple shear                                    |
| 64     | Sintering (as an industrial process)            |
| 31     | SINUSOIDAL FLOW                                 |
| 47     | Slip at a wall                                  |
| 16     | Slurries                                        |
| 25     | Spectrum (of relaxation or retardation times)   |
| 17     | Sols                                            |
| 12, 13 | Solutions                                       |
| 36     | Sphere clusters, motion of                      |
| 35     | Sphere, falling through a fluid                 |
| 35     | Sphere, flow past                               |
| 67     | SPINNING (as an industrial process)             |
| 68     | Spray drying                                    |
| 68     | SPRAYING (as an industrial process)             |
| 69     | SPREADING (as an industrial process)            |
| 51     | Springs (used in an analogy)                    |

|        |                                                                  |
|--------|------------------------------------------------------------------|
| 49     | Statistical mechanics                                            |
| 28, 35 | Stoke's law flow                                                 |
| 28     | Streamline flow                                                  |
| 22     | Stress dependence (of viscosity)                                 |
| 41     | STRUCTURE (molecular or particulate)                             |
| 22     | Structural viscosity                                             |
| 12     | Supercritical fluids                                             |
| 43, 44 | Surface active agents                                            |
| 43     | SURFACE RHEOLOGY                                                 |
| 47     | Surface roughness                                                |
| 43     | Surface tension                                                  |
| 43     | Surface viscosity                                                |
| 16     | Suspensions (non-colloidal)                                      |
| 29     | Taylor vortices                                                  |
| 38     | TEMPERATURE (as a variable)                                      |
| 38     | Temperature gradient                                             |
| 38     | Temperature, viscometry at extreme                               |
| 2      | THEORETICAL                                                      |
| 49     | THERMODYNAMICS                                                   |
| 23     | Thixotropy                                                       |
| 23     | TIME DEPENDENCE (of viscosity)                                   |
| 53     | Torsional tests                                                  |
| 56     | Tracers (use of, to visualise flow)                              |
| 32     | Transient flow                                                   |
| 29     | Transitional flow (laminar/turbulent)                            |
| 61     | Tribology                                                        |
| 33     | Tubes, circular or non-circular, straight                        |
| 34     | Tubes, circular or non-circular, bent                            |
| 30     | Turbulence damping                                               |
| 30     | TURBULENT FLOW                                                   |
| 19     | Two-phase flow (engineering sense, gas-liquid mixtures in pipes) |
| 55     | ULTRASONIC APPARATUS                                             |
| 33     | UNIFORM PIPE                                                     |
| 37     | Vacuum                                                           |
| 64     | Vacuum forming                                                   |
| 34     | Valves                                                           |
| 34     | Varying cross-section                                            |
| 34     | Venturi meters                                                   |
| 31, 32 | Vibrational flow                                                 |
| 33, 53 | Viscometers, capillary                                           |
| 35, 53 | Viscometers, rotational                                          |
| 38     | Viscosity index                                                  |
| 13     | Viscosity-index-improved oils                                    |
| 25     | VISCOELASTICITY                                                  |
| 38     | Viscous heating                                                  |
| 21     | Viscous liquids                                                  |
| 56     | Visualisation (of flow patterns)                                 |
| 29, 30 | Vortices                                                         |
| 47     | Wall layer                                                       |
| 47     | Wall effects                                                     |
| 32     | Water hammer                                                     |
| 61     | Wear                                                             |
| 34     | Weirs                                                            |
| 25     | Weissenberg effect                                               |
| 23     | Work hardening                                                   |
| 24     | YIELD STRESS                                                     |
| 44     | Zeta potential                                                   |

NOTE: Entries in capitals correspond to main headings in the Classification System.



## INFORMATION RETRIEVAL FOR THE CHEMICAL ENGINEER

D. W. Hadley (University of Reading)

### Introduction

The at least exponential, sometimes seemingly explosive, growth in the volume of scientific and technical literature affects the chemical engineer as much as his colleagues in other disciplines. The problem of coping with the sheer bulk, of containing, of reducing to manageable proportions, of storing, and, above all, of retrieving scientific information is one of the most pressing problems of the day. Its solution is made all the more urgent by the estimate that in the U.K. alone at least £6 million a year,<sup>1</sup> and possibly between five and ten times more,<sup>2</sup> is wasted by the duplication of research and development work because published information was not discovered in time. In a third of the cases timely discovery of relevant information would have influenced the course of the work.

Merely scanning the titles of possibly relevant documents may be an almost full time project—and in many cases the presence of the information required is not indicated by the title alone. Some form of organised retrieval system is essential, and for several years the I.Chem.E. has been considering a convenient system which could be used by small or medium sized organisations, and also by the individual engineer.

In U.S.A. the problem of chemical engineering information retrieval was tackled by the American Institute of Chemical Engineers, and a system of co-ordinate indexing based on a word list (or "thesaurus") was evolved. At the invitation of I.Chem.E., Mr. Bart Holm, of Du Pont, who had been active in setting up the A.I.Ch.E. system, gave a series of lectures in London in 1964.<sup>3</sup> The interest aroused led to the setting up of a Study Group, which subsequently became the I.Chem.E. Committee on Engineering Information Retrieval, under the Chairmanship of Mr. R. Snel (Senior Information Officer, Shell Petroleum). Other members of the Committee were: Dr. W. E. Batten (Head, Central Tech. Inf. Unit, I.C.I.); Dr. G. M. Dyson (Consultant); Dr. I. L. Hepner (Editor, *Chem.Proc.Eng.*); Prof. P. M. C. Lacey (Exeter Univ.); Mr. J. O. S. MacDonald (Editor, *Brit.Chem.Eng.*); Prof. R. W. H. Sargent (I.C.); and Dr. G. A. Somerfield (O.S.T.I.).

The Committee had as its objective "the setting up of a co-ordinate indexing system and the preparation of a word list as an aid in the timely acquisition of chemical engineering information." It undertook also to act in an advisory capacity to the Council of Engineering Institutions.

The system selected was the keyword method described in this article. Two papers describing the underlying philosophy were presented in May 1966,<sup>4</sup> and the system was officially unveiled and

demonstrated at a meeting entitled "Information Retrieval: Validation of the Keyword Method for Chemical Engineers" held in London before an invited audience during November 1967.

In congratulating the Committee on its achievement, the President of I.Chem.E., Mr. C. S. Windebank, thanked particularly those of its members who were not members of I.Chem.E. The blend of chemical engineers and information scientists was vital to the success of the project.

Mr. Windebank and Mr. Snel then discussed the finding of the Committee and described the bases of the keyword system. It was felt that what was required was a relatively cheap, simple, scheme, suitable for use by engineers in both research and production, which would be available quickly. A complex, computer controlled, system with individual user outputs would probably be more efficient, but might take 10 years to develop; not only would this system be prohibitively expensive for smaller companies, it might also be obsolescent by the time development was complete.

### **Bases of the I.Chem.E. System**

The initial decision was to adopt a co-ordinate indexing system, with retrieval through keywords. Basic to the method is the use of normal words, the natural language of communication, both for indexing and retrieving information. The words, however, are specially selected; in fact KEYWORDS, particularly relevant to the subject under consideration. Appropriate documents are first abstracted, and entered on document cards with the keywords underlined. A corresponding set of keyword cards is then constructed, each card listing the documents mentioning that particular keyword.

For retrieval purposes about three keywords are selected, and the corresponding keyword cards are examined to discover which abstracts contain *all* the keywords. The relevant document cards are then retrieved for further study.

Co-ordinate indexing, a technique invented in 1921, has become increasingly popular during the past 10-15 years. Each index card (in the present case each keyword card) is divided into a numbered grid, with each number corresponding to a document card. Whenever a document card mentions a particular keyword a hole is punched in the numbered space appropriate to that document. All keyword cards ought therefore to contain a number of holes, the co-ordinates of which denote relevant documents. (A keyword card containing no, or only very few, holes could imply that the keyword was not appropriate for that particular index.) The retrieval of relevant documents may be achieved by superposing the selected keyword cards and holding them up to the light to discover which holes are common to all cards i.e. which abstracts mention all the keywords. The system makes use of what are correctly known as optical coincidence cards, but popular terms such as "peek-a-boo" are often used, even in publications.

An advantage of keyword systems is that they require no pre-grouping of subjects; no division and sub-division into sections; and no coding and decoding. The method relies on the vocabulary used for expressing ideas in a particular field, and thus the efficacy of any system depends on the appositeness of the selected word list.

### The Thesaurus

The A.I.Ch.E. have already established a co-ordinate indexing system based on a thesaurus of chemical engineering terms. It might at first appear that the only sensible course would be to adopt the U.S. system, particularly when several U.S. journals give keywords at the head of each article; but further considerations show that the system is not really appropriate for the typical British chemical engineer. There are two principal difficulties: the large size of the thesaurus introduces complexity and adds to the operating cost; and there are considerable differences in nomenclature between the U.K. and U.S.A.

The second point is perhaps not of major importance, since most British engineers become acquainted with U.S. terminology. Even so, misunderstanding could arise over the different meanings of "gas": whilst a "measuring cylinder" becomes a "graduate" after crossing the Atlantic.

The A.I.Ch.E. system uses a basic thesaurus of about 7000 keywords, additional to which are a further 3000 terms that lead-in the searcher to the appropriate keyword, and a system of 10 role indicators,\* which adds about 30% to the total number of keywords. A computer is required for the satisfactory management of such a large scale system, thereby preventing its use except by the largest companies, taking into consideration the almost general unwillingness to spend much money on information science. One must always balance cost against the waste resulting from the use of an inadequate system. The aim of the I.Chem.E. Committee was to achieve a relatively simple and inexpensive system, appropriate for the required purpose, and only marginally less comprehensive than one of considerably greater complexity.

Considerations to be borne in mind when compiling a thesaurus are that it should be convenient for the user, who is assumed to be a chemical engineer (at the same time, in view of the increasing co-operation between all engineers, its use should not be inconvenient for other disciplines); its vocabulary should facilitate the preparation

---

\* Role indicators define the role of the keyword in the document . . . e.g. ethylene-1 = ethylene as a raw material or feed; ethylene-2 = ethylene as a product or byproduct; ethylene-3 = ethylene as a contaminant or undesirable component, etc. etc. In a sense role indicators serve the same purpose as inflexions in an inflexional language. However, for each role a separate keyword card must be set up, consequently the number of keyword cards is multiplied by the number of roles assigned to each keyword. They introduce an added complexity and there is a lot of controversy regarding their usefulness. The I.Chem.E. Keyword List contains no role indicators.

and retrieval of document cards by the eventual user; and it should be open-ended, capable of being up-dated, or modified to suit the individual user. The completed system should assist an indexer, help the searcher to phrase his questions correctly, and if used by an author might aid in achieving clarity of expression.

The thesaurus evolved by the Committee consists of 650 keywords embodying the everyday terms in chemical engineering, together with a further 1150 lead-in words, which are synonyms or associated words that are convenient to relate to the keywords. The same basic keyword is used to cover all derivatives; e.g. "store" is used both as verb and noun, and includes storage, storing, etc.

An attempt has been made to avoid the difficulties associated with chemical names by using the main chemical keywords and a fragmentation method for breaking down more complex compounds into their component parts.

It is felt that the thesaurus is large enough to meet most demands, and yet is small enough to be readily memorised and simple to use.

### Details of System

Associated with the thesaurus, which is the software of the system, one needs the relevant hardware.

Abstracts, with keywords underlined, are typed onto sequentially numbered document cards, whose ordering should not be disturbed. Keyword cards are translucent semi-rigid sheets 15"  $\times$  8½", arranged in alphabetic order of keyword. The sheets, with cardboard separators, can be placed in a trough fitting into a desk drawer. Each sheet will accommodate 5000 holes, arranged in 50 blocks of 100, and thus a single sheet can refer to a store of 5000 documents. A simple modification will permit indexing of a larger store. Ready numbered keyword sheets may cost 7d. or more each, but plain sheets at 2½d. each can be used in conjunction with a special co-ordinate punch (at £74) and a transparent numbered grid. Simple manual punches, without co-ordinate facilities could be used for small systems.

A benefit obtained through the use of translucent keyword sheets is seen during the retrieval process. The selected sheets are held up to the light, and documents containing *all* the keywords are indicated by superposed holes, whose co-ordinates can be found using a transparent grid. In addition, documents containing all but one of the selected keywords, and hence having coincident holes in all but one sheet, will be indicated by diffuse spots of light. Reference to these latter documents might be worth while. Although a given document might contain as many as 20 keywords, usually three keywords are sufficient for retrieval.

The rapidity and ease of retrieval from a prepared store of 1500 document cards was demonstrated convincingly at the November 1967 meeting. Several documents were retrieved which would have been missed if the search had been based on title only. The store had been prepared before the demonstration, but its preparation was

explained in some detail. It is in abstracting and keywording that expertise is required. A non-professional abstractor would probably lack the background for performing these processes effectively, but could, from a knowledge of the system and the keyword list, add to or alter the keywords for optimal retrieval.

Before being released the system had been tested in conjunction with three industrial companies, a university and Aslib. Both recall (the ratio of the number of relevant articles retrieved to the number of relevant articles in the store), and precision (the ratio of relevant items retrieved to total number retrieved) compared favourably with other retrieval systems, e.g. Uniterm. Using the demonstration store, each card of which had been indexed for about 20 keywords, in order to cover all possible aspects of each abstract, the recall was greater than 90%; a figure of about 70% is usual for the more normal indexing by 8-15 keywords.

Although the Keyword System is relatively simple to use, some expertise is required. The Information Retrieval Committee has therefore insisted that the aims of the I.Chem.E. would not be furthered by the issue of the word list alone without instruction in its use. A minimum of one day's instruction is the required provision, although it is thought that three to five days would be required for efficient training. Members of I.Chem.E. wishing to install the system on a personal basis can purchase a thesaurus, together with one day's instructions, for £10. Commercial organisations will pay £25 per day for a word list and instruction. The cost of equipment, including keyword and index sheets, is estimated to lie between £75 and £300, depending on the type of punch and filing arrangements used. Further details of the system can be obtained from: The Information Officer, the Institution of Chemical Engineers, 16 Belgrave Square, London, S.W.1.

*(This review could not have been written without the assistance of The Institution of Chemical Engineers, and particularly that of the Information Officer, Mr. B. E. Evans. Errors of fact are, however, the sole responsibility of the author, and should not be taken as implying the adoption by the Institution of a slipshod approach to the problems of information retrieval.)*

#### References

- <sup>1</sup> J. Martyn, *New Scientist*, 6th February, 1964, 21, 338.
- <sup>2</sup> L. Wilson, *Address to the Parliamentary and Scientific Committee*, 15th November, 1966, 14.
- <sup>3</sup> *Chemical Engineer*, December 1964.
- <sup>4</sup> *Chemical Engineer*, January 1967, CE 22.

# **SOME THOUGHTS ON THE PREPARATION OF A THESAURUS FOR RHEOLOGICAL INFORMATION RETRIEVAL**

D. C.-H. Cheng

Warren Spring Laboratory, Stevenage, Herts.

## **Introduction**

The general problems of information retrieval are discussed in several articles that appear in the present issue of the Bulletin. It appears that, among the different schemes that are in use, the one based upon the keyword method has proved to be the most popular. Such a scheme has been developed by the Committee on Engineering Information Retrieval of the Institution of Chemical Engineers and is also described in this Bulletin. If a similar system is adopted for the retrieval of rheological information all that remains to be done would be the preparation of a thesaurus of rheological terms.

The thesaurus of chemical engineering terms prepared by the Committee on Engineering Information Retrieval contains 650 keywords with a further 1150 lead-in words. For rheology, a smaller number would perhaps be sufficient. The list drawn up by the Joint Working Party on the Flow of Non-Newtonian Fluids of the Institution of Chemical Engineers and the British Society of Rheology (B.S.R. Bulletin No. 3, 1967, p. 2, and more details appear in the present Bulletin) contains 69 keywords and may, of course, be used for information retrieval. But, because these have been designed primarily for the classification and not the retrieval of rheological information and because it represents a rather coarse grid, the retrieval scheme based on these keywords cannot be comprehensively selective.

This note is therefore concerned with the preparation of a thesaurus of rheological terms for information retrieval.

## **An Analysis of Work in Rheology**

The thesaurus of rheological terms can be prepared through the classification of rheological information. It is considered that an analysis of rheological work is thus necessary in order to discover what exactly are the subjects to be classified.

A work (the term is used here to indicate any investigation projected, in progress or completed; unpublished or published; in the form of manuscripts, reports, papers, pamphlets or books) in rheology may be concerned with rheology as such or with industrial processes or other applications in which rheological principles are involved. Each piece of work must deal with some particular material or group of materials which can be identified by the chemical name, by the group or class, by the physical state or even by the type of rheological properties they possess. The work will either be concerned

with the rheological properties of the materials and/or their rheological behaviour and/or the rheological processing whether industrial or otherwise.\* For each of these topics, the approach may be theoretical or experimental. In the latter case, new forms of apparatus, experimental techniques or ways of interpreting data may be discussed. In either approach, the results may have been obtained for their own sake, or they may have been put to some use.

The work need not be directly concerned with rheology or its application but may be closely connected with them. Examples of this work can be on the subjects of management, teaching and classification of rheology and its application.

In order to identify the content of a work as well as its purpose, a whole series of questions elucidating the above points and other details would have to be asked and answered. A list of such questions is given below. The exact meaning to be attached to the terms used will become clear when sample answers are given in the next section.

1. What is the material studied?
  - 1.1 What is its chemical name?
  - 1.2 What is the class to which it belongs?
  - 1.3 What is its physical state?
  - 1.4 Of what rheological type is its property?
  - 1.5 What is the rheological model considered?
2. Is rheological property studied?

If so:—

  - 2.1 What rheological property is considered?
    - 2.1.1 What are the variables?
  - 2.2 Is theory of materials considered?

If so:—

    - 2.2.1 What use is made of the results?
  - 2.3 Is rheometry considered?

If so:—

    - 2.3.1 What apparatus is described?

---

\* The phrases "rheological property," "rheological behaviour" and "rheological process" are used here with specific meanings. "Rheological property" denotes the constitutive relationship between stress and strain tensors, their derivatives and other parameters of state which relate to the behaviour of an element of the material in question. This implies that the material is homogeneous or can be so treated, but it need not of course be isotropic. "Rheological behaviour" denotes the behaviour of a macroscopic mass of the material when the boundary conditions, geometric, dynamic and kinematic, are prescribed; this usually implies the behaviour in some idealised situation. "Rheological process" denotes rheological behaviour in more complex and realistic situations. The situation may be complex, not because the boundary geometry and deformation or flow is complicated but because the behaviour of multiphase systems are involved. Examples of these are two-phase gas-liquid flow and the flow of fast settling particle suspensions. Such situations may sometimes be approximated by a simple one, which then becomes an example of rheological behaviour, but often no such approximation is possible.

- 2.3.2 What method of data interpretation is considered?
  - 2.3.3 Is (worthwhile) data obtained?
  - 2.3.4 What use is made of the results?
3. Is rheological behaviour studied?  
If so:—
- 3.1 What geometry is involved?
  - 3.2 What other boundary conditions are involved?
  - 3.3 What is the flow character?
  - 3.4 What quantities are considered?
    - 3.4.1 What are the variables?
  - 3.5 Is the work theoretical?  
If so:—
    - 3.5.1 What mathematical technique is used?
    - 3.5.2 What use is made of the results?
  - 3.6 Is the work experimental?  
If so:—
    - 3.6.1 What apparatus is described?
    - 3.6.2 What method of data interpretation is considered?
    - 3.6.3 Is (worthwhile) data obtained?
    - 3.6.4 What use is made of the results?
4. Is a rheological process (industrial or otherwise) studied?  
If so:—
- 4.1 What process or application is considered?
  - 4.2 What equipment is involved?
  - 4.3 What quantities are considered?
    - 4.3.1 What are the variables?
  - 4.4 Is the work theoretical?  
If so:—
    - 4.4.1 What technique is used?
    - 4.4.2 What use is made of the results?
  - 4.5 Is the work experimental?  
If so:—
    - 4.5.1 What apparatus is described?
    - 4.5.2 What method of data interpretation is described?
    - 4.5.3 What use is made of the results?
5. If the answer to questions 2, 3 and 4 is no, specify the topic.

### Examples of Rheological Subjects

The answers to the questions given above serve to identify a work in rheology and could form the basis for the preparation of the thesaurus. It is appropriate here to give some examples of the answers expected to each question to illustrate the sort of keywords that would appear in the thesaurus.

- 1.1 Chemical name of material:
  - Polystyrene
  - Cheese
- 1.2 Class of material:
  - Polymers
  - Foodstuffs
- 1.3 Physical state of material:
  - Melts
  - Solutions
  - Suspensions
  - Emulsions
- 1.4 Rheological type:
  - Viscoelastic
  - Thixotropic
- 1.5 Rheological model:
  - Power law fluid
  - Reiner-Ericksen fluid
  - Material with square box relaxation time spectrum
- 2.1 Rheological property:
  - Viscosity
  - ? — Normal stress
  - Relaxation time spectrum
- 2.1.1 Variable:
  - Electric field
  - Temperature
  - Molecular weight
  - Composition
- 2.2 Theory of material:
  - Absolute kinetic rate theory
  - Casson theory
  - Oldroyd (1950) theory
- 2.2.1 Use of the result of theory of material:
  - Interpretation of rheological data
  - Interpretation of molecular behaviour
- 2.3 Rheometry:
  - 2.3.1 Rheometer:
    - Rheogoniometer
    - Creep apparatus
  - 2.3.2 Method of data interpretation:
    - Graphical method
    - Explicit formulae
  - 2.3.4 Use of rheological data:
    - Interpretation of molecular behaviour
    - Quality control
    - Design of industrial plant
- 3.1 & 3.2 Geometric, dynamic and kinematic boundary conditions:
  - Rising bubble
  - Impulsive motion of wall

- 3.3 Deformation of flow character:
  - Secondary flow
  - Transitional flow
- 3.4 Dynamic and kinematic quantities:
  - Velocity profile
  - Power consumption
  - 3.4.1 Variables: (See 2.1.1)
- 3.5 Theoretical
  - 3.5.1 Mathematical technique:
    - Numerical method
    - Perturbation method
    - Variational principle
  - 3.5.2 Use of theoretical results:
    - Application to rheological processes
    - Interpretation of rheometry
- 3.6 Experimental
  - 3.6.1 Experimental apparatus:
    - Anemometer
    - Tracer technique
  - 3.6.2 Method of data interpretation:
    - Statistical
    - Electronic circuitry
  - 3.6.4 Use of experimental results:
    - Comparison with theory
    - Application to rheological processes
- 4.1 Rheological process:
  - Mixing
  - Blood flow
- 4.2 Equipment:
  - Z-blade mixer
  - Medullary cavity of bone
- 4.3 Dynamic and kinematic variables: (See 3.4)
  - 4.3.1 Variables: (see 2.1.1)
- 4.4 Theoretical
  - 4.4.1 Theoretical method:
    - Dimensional Analysis
    - Modelling
  - 4.4.2 Use of theoretical results:
    - Scale-up procedure
    - Process control
- 4.5 Experimental
  - 4.5.1 Experimental apparatus: (See 3.6.1)
  - 4.5.2 Method of data interpretation: (See 3.6.2)
  - 4.5.3 Use of experimental results:
    - Comparison with theory
- 5. Other topics:
  - Teaching
  - Management
  - Classification
  - History of rheology

## The Preparation of the Thesaurus

The thesaurus of rheological terms can readily be prepared once a comprehensive list of rheological subjects, such as those given above, has been drawn up. The division of the rheological terms in the thesaurus into *keywords* and *lead-in words*, as used by the Institution of Chemical Engineers, could be followed with profit.

It is clear that the rheological knowledge of any one individual will be rather limited and this article does not therefore attempt to give a complete list of keywords for the whole of rheology. Such a task is considered to be best sponsored by the British Society of Rheology whose Research Committee may wish to consider setting up a Subcommittee to carry out this task as part of their programme in trying to stimulate rheological research.

(Crown Copyright Reserved.)

## RETRIEVAL REVIEWED

The human brain is a biological retrieval system used by all of us; but it is rarely a convenient system when used by itself, or supplemented only by a series of references scrawled on odd scraps of paper or jotted down haphazardly in a notebook [I should know; it is the method that I have used for many years!] To make optimal utilisation of the exponential growth in the scientific literature one needs an organised index of references from which the required information, and a minimum of redundant material, can be extracted quickly and easily. The organisation may range from a box of file cards grouped into a few broad subdivisions to a large computer-based storage system. Most of us will need something between these two extremes, but whatever system of storage is used one must ask the store the right questions in order to get sensible answers. If we are very fortunate someone else will have prepared the store, and in a few cases one of Mr. Ashworth's good librarians, or somebody at N.L.L., may have time to help us ask the right questions. In the majority of cases, however, a personal file system will be needed, and even if some references are available ready indexed the user will still need to classify many entries, and will always be responsible for what is retrieved.

Any system is better than none at all (to quote Babbage: "Errors using inadequate data are much less than those using no data at all"); but no system will be perfect. Comprehensiveness is difficult to achieve without complexity; and what you can get out of a system will always be limited by the amount of useful material fed into it. The B.S.R.—I.Chem.E. classification scheme, and the I.Chem.E. Keyword system should both be of use to members of this Society, either in unadapted form or by showing how an existing system may be improved. It is relatively simple to criticise the weakness of either system, much more difficult to suggest feasible improvements, and impossible without considerable experience to comment on the practical advantages.

I am neither a rheologist nor an information scientist, and my prejudices must be mainly personal. My intention is to start individuals thinking about both systems, in the hope that members' experience of using the "fluids" scheme may prompt them to make suggestions for its eventual improvement.

The B.S.R. scheme cannot be considered apart from its origins as a classification scheme to assist the investigations of a Working Party. One member of that Working Party is reported to have called it "a classification scheme and not an information retrieval system"; another is reported as saying that "although primarily a classification scheme it does act as a coarse mesh retrieval system." Papers mentioned in *Rheology Abstracts* are to be classified according to the scheme, a procedure which will be useless unless the code numbers permit the abstract to be retrieved: hence, by definition we have an information retrieval system. I certainly do not agree that the mesh

is coarse; in the I.Chem.E. system the whole of rheology is classified under a few keywords, whilst here we get a complete system devoted to rheology of the fluid state. Rheology of solids will form a later part of the scheme, and thus for all but full-time general rheologists (who, I am told, form a minority of members of the Society) the mesh would seem a very close one.

The feature of the scheme that I view with greatest concern is that the use of direct coding prevents the system being open-ended. There are 69 descriptors, each of which is to be represented by a single notch on an 80 hole edge-notched card. Indeed it was reported that the number of descriptors was reduced so that an 80 hole card could be used. This procedure seems not only wasteful but allows no room either for future expansion or personal additions. Did the Working Party consider the problems of incorporating the fluids system with the proposed scheme for solids? It seems ludicrous that a material just above its melting point must be dealt with in a completely separate filing system from the same material at a temperature 1°C lower. Most individual users will have a special interest in just one small part of the field, and so will wish to add extra retrievable terms to subdivide their particular interest, and probably add symbols for particular materials. How do they set about this task when the standard edge-notch card has no punch holes available? It is in planning the mechanics of the system that the Working Party might have benefited from the inclusion of an information scientist. At the very least the scheme might have been submitted for comment to such a specialist before being publicly released.

Edge-notch cards would seem a logical choice for any file system having no more than a few thousand entries. The cards do not need to be pre-sorted or arranged in any particular order, all the sorting being carried out by the needles during the retrieval process. Once retrieved cards can be replaced anywhere in the index without in any way affecting the efficiency of the system. It is normal, however, with all but the very smallest systems, to use an indirect coding scheme: the keywords or descriptors are rearranged into numbered blocks, and different combinations of a few notches are used to represent the descriptors in each block. In this way hundreds of descriptors could be accommodated on the selected type of card. Many descriptions of indirect coding methods have been published, and the use of one of these systems would have considerably extended the scope of the fluids classification. It is stated that the remaining 11 holes on the card are used for giving the reference to the original paper. But how is this to be accomplished? Surely here an indirect codification must be used; and if used for the final 11 holes why not for the first 69?

The choice of descriptors will naturally vary somewhat with the individuals who have selected them; we are assured that those finally chosen have been subjected to critical scrutiny both for ease of indexing and ease of retrieval, using already published *Rheology Abstracts* as the test material. Most of the terms selected look right,

and one ought not to cavil at the choice without doing one's own homework. I am surprised, nevertheless, that six out of 69 terms cover language of document. What is the purpose of this? Surely not to warn rheologists away from papers written in tongues other than their own, since the searcher is less likely to know of, and thus be more in need of knowing, work carried out in other countries. Language is not necessarily a good guide to country of origin. One recent polymer journal contained articles in English written by workers from France, Germany, Holland and Japan, and other issues of the same journal have contained papers from Eastern Europe written in French and German. Conversely, all papers in *Cahiers du G.F.R.* are in French, regardless of country or origin; and the Editor tells me that an equivalent ruling for English holds true for *B.S.R. Bulletin*.

The "keyword" and "fluid" systems both clearly indicate their origins. They appear, respectively, as the products of a professional committee and a group of enthusiastic amateurs. In a parallel manner one has a feeling that the keyword system might be too professional for individual use, whilst the B.S.R. scheme is unlikely to find much direct application in its complete form in a medium-sized commercial organisation. This conclusion is not surprising, since the two systems were formulated independently (despite I.Chem.E. co-operation in each), and have essentially different objectives. The I.Chem.E. hope that individuals will use the keyword system, but there must be few individuals who can afford the initial outlay of £75-£300 for hardware, plus tuition fees.

The keyword system operates smoothly and efficiently, declaring itself the brainchild of a panel of experts from a range of disciplines, and demonstrating the advantages of having undergone a large amount of practical debugging before release. One might query, therefore, why for trial and demonstration purposes a store of only 1500 items was used to cover the whole of chemical engineering. A store of this size could probably be coped with using a simpler retrieval system. The answer is that the 1500 indexed documents made use of 85% of the keywords, which ought to be sufficient for working trials of the system. Nevertheless I am left with the nagging feeling that the limit might have been set by the time needed to compile a larger index. A statement by the compilers of both "fluid" and "keyword" systems of the average time taken to classify a batch of, say, 20 assorted documents would have been useful for the individual considering changing all his references to a new system.

The I.Chem.E. Committee might have considered whether the keyword system should incorporate means of retrieving a document when the author and date of publication are known (perhaps they did, and then rejected the idea). One can always guess appropriate keywords, but it might have been preferable to incorporate a coding scheme for author's name. The system is, of course, open-ended, and allows the user to incorporate any additional retrieval aids he may require.

Any individual adopting the keyword system is faced with two problems (both of which become less serious if the system is run by a commercial organisation under skilled supervision): first he has to punch the keyword cards in the correct place; then he has to maintain the keyword cards in alphabetic, and the document cards in numerical, order.

The first task must be rather tedious using a very small hand punch, and is also subject to errors—which can be corrected by covering up the hole, at the expense of still more time. Each individual will thus need to carry out a personal cost-effectiveness study to determine whether he needs to spend another £75 on the recommended co-ordinate punch.

The prospect of having two sets of cards, both of which must be kept in order, is the drawback that really frightens me—although this is probably because I am too untidy to make a good chemical engineer. How to find a mis-sorted card in a hurry in a system which is not reported as having sub-divisions is a problem that the inventors of the system must have considered. [This last statement is made rather warily, because I have not seen full details of the system, and know only what the Editor has shown or told me. Presumably you do not get all the information until you have paid out £10 for a word list.] Possibly there is some sort of ordering mark on each card. By complicating the system somewhat it might have been made possible to retrieve any required keyword or document card by carrying out a preliminary sorting; although putting edge-notches on each card would probably be going too far.

The keyword card is comprehensive, and therefore needs to be large (15" × 8"). For the sake of rigidity when handling the card must be fairly thick, and so for individual use the bulk of the system could be a problem—although it is a problem which will fit into a single desk drawer. The document cards used at the demonstration were also massive-looking affairs, but presumably smaller cards would not be unsuitable.

It is understandable that the I.Chem.E., having developed a very good retrieval system, do not want the reputation of that system damaged through misuse. But the £10 or £25 fee for word list plus tuition seems rather steep for the individual, particularly one whose main concern is not chemical engineering, but whose work has affinities with it. A rheologist might be able to improve a retrieval system of his own by making use of some of the I.Chem.E. keywords, without going through the complications of changing his system to one involving co-ordinate indexing. Under such circumstances, recalling the quotation from Babbage given earlier, it would have been a generous gesture on the part of the I.Chem.E. to have made copies of the thesaurus available to individuals at cost price, without imposing the need for instruction in the full system. The B.S.R. is a much smaller and less wealthy body, and yet it is making its classification scheme freely available for anybody who wishes to use it.

All through this article I have been carping about the drawbacks associated with each of the systems considered, mainly because I was asked to criticise, but partly because the good features are largely self-evident, and have already been described. Both B.S.R. and I.Chem.E. have clearly done very good jobs, and can honestly claim to have constructed sturdy refuges against the fall-out from the information explosion.

E. TAL

[The opinions expressed above are not necessarily those of the British Society of Rheology or the Editor of the Bulletin. I am informed that the Working Party did consider an indirect coding system, but decided that the complexities were too great.—Editor.]

## THE RETRIEVAL OF CHEMICAL STRUCTURE INFORMATION

W. J. Wiswesser, Fort Detrick, Maryland, U.S.A.  
and

A. J. Barnard, Jr., J. T. Baker Chemical Co.,  
Phillipsburg, New Jersey, U.S.A.

The efficient management of chemical information has become compelling now that up to 150,000 new compounds are added each year to the incompletely mastered backlog of 3.5 to 4 million reported chemicals. Ever since organic chemistry began its accelerated growth with Butlerov's emphasis on *structure* in 1861, the kinds of atoms and their arrangements constitute major parameters of any body of organised chemical information. Today diverse interests are facing the information retrieval problem in chemistry by the use of tools based on structure. Developments have become so ramified and thereby confusing, that a historic review of some "hits and errors" is appropriate.

### LINE NOTATIONS

#### The Original Information Retrieval Tools of the Chemist

In one sense there are only two fundamental types of structural "IR" tools for chemistry: first atoms and functional groups, which may be designated by symbols; and, second, rules for ordering these groups or their symbols.

Our basic set of atomic symbols was advanced largely by J. J. Berzelius in 1813. He showed remarkable perception as an "IR" pioneer when he argued that, "the chemical signs ought to be letters, for the greater facility of writing, and not to disfigure a printed book." He developed most of the atomic symbols in use today and from the beginning provided a simple means for dividing elements into a small number of frequently mentioned *nonmetals* (inappropriately called metalloids by Berzelius) and a large number of infrequently cited metals: "In the class that I call metalloids, I shall employ the initial letter only, even where the letter is common to the metalloid and to some metal."

Today we can sharpen machine-based information retrieval by perceptive standardisation of this dichotomy with one-letter symbols for the common nonmetals (B, C, E for bromine, F, G for chlorine, H, I, N, O, P, and S) and *two-letter symbols for all other elements*, infrequently encountered in large listings. Expanded, doubly-suggestive symbols are adopted for potassium (KA), uranium (UR), vanadium (VA), tungsten (WO), and yttrium (YT). The internationally approved two-letter symbols can be maintained for the remaining elements. The unused letters of the English alphabet then can serve as discriminating signals for frequently cited *combinations* of atoms.

The primitive rule for the ordering of the elements in empirical formulas ("C and H followed by . . .") was initiated by R. F. Marchand (1839) and F. Wohler (1840), and suffered unhappy promotion by Charles Gerhardt (and by others), even though Gerhardt himself acknowledged that the H-count might better be cited last because of its uncertainty (in analysis *or* count).

A promising alternative to empirical formulas was well under way by this time (1845). Important *compound* atomic groups had been assigned special two-letter symbols almost since two-letter symbols for atomic radicals had been advanced by Berzelius. For example, the cyanogen radical was denoted as Cy soon after its discovery by Gay-Lussac in 1815. By 1845, symbols for many compound radicals were better standardised than the English spelling of their names: e.g., Ac for acetylene, Ad for amidogen (that is,  $\text{NH}_2$ ), Am for ammonium, Ci for cinnamyle, and Fo for formyle. Thus ammonia was properly symbolised in 1845 as AdH and cyano derivatives were indicated by CyBr, CyCl, CyH, and CyI.

Only a few of the early symbols for compound groups have survived, along with a few subsequent additions. However, Frankland's proposals of 1866<sup>1</sup> to use symbols for all *common* radicals for teaching and representing reaction sequences are still vital for the present day. Rarely used early symbols fell into disuse because they were structurally unrevealing and statistically unimportant: more effort was required in remembering the meaning of a symbol than was saved in writing it. The selection of each chemical-group symbol requires statistical as well as historic justification.

Our appreciation of the ordering of atoms and compound groups in organic molecules, for the purposes of discourse and retrieval, stems from the delineating intuition of the chemists of the period 1860–1865 that the atoms in organic compounds assumed fairly fixed arrangements. They sketched a chain structure with the carbon skeleton upright and then scanned the items from left to right and top to bottom to provide a linear "rational" formula that could be readily set in type. The sketch for lactic acid would yield the rational formula shown on the right.



Note that the success of this linear presentation lies in the rightward and downward polarised orientations.

### The Wiswesser Line Notation

With the mushrooming of chemical literature in the second quarter of the twentieth century, dissatisfaction was widespread with systematic chemical nomenclature as a structural "IR" tool. With the increasing availability of Hollerith punched card processing equipment, some workers foresaw the need for a simple, linear "Teletype" notation that could represent a chemical structure unambiguously.

The most explored and utilised of such line-formula notations was introduced by W. J. Wiswesser in 1950 and details were published as a monograph in 1954<sup>2</sup>; a second edition has appeared recently.<sup>3</sup> Much of the popularity of the Wiswesser Line Notation (WLN) may stem from the inventor's adherence to the two lessons from the history of early structural "IR" tools, namely: (1) the use of standardised atomic and atomic group symbols, with the shortest reserved for the most frequent elements and groups and (2) the standardisation of the traditional "line-formula" citation sequence. He also was determined to denote all structures using the limited group of symbols available on the keyboard of teletype and tabulating equipment; as a consequence, WLN's are transmittable by all modern channels of high-speed communication.

Many of the elements of WLN's can be appreciated from Table I.

Here some 250 compounds are presented in increasing order of notational complexity. The assignment of the symbols to the elements and structure fragments is delineated at the top of the Table. The reader should be able to write conventional two-dimensional structural formulas and line formulas by simple substitution of the indicated structural elements for the WLN symbols. More than a century ago, Thomas Graham in his colloid studies, first suggested that the viscosity of pure liquids might change in a regular manner within a similar series of compounds. The structures listed in Table I represent about 90% of the liquids early studied for rheological correlations!

In the subsequent development of physical chemistry, a distinction was properly drawn between additive properties (which depend on the number and kind of atoms in the molecule) and constitutive properties (which depend on the particular arrangement of atoms within the molecule). Most physico-chemical properties are in part both additive and constitutive. Consequently structural relations have been given attention with such properties as molar volume, parachor, critical constants, and dipole moments. Nearly all of the correlations recognise contributions from distinct monatomic entities, such as halogen atoms and relatively simple functional groups. Thus in 1938, when M. Souders, Jr., reported a correlation between molar volume and a log-log plot of viscosity, his "atomic and structural constants" acknowledged sharp differences between C and CH<sub>3</sub>, O and OH, N and NH, COO<sup>-</sup> and COOH, aldo and keto, tertiary and quaternary carbon, and double bonds and ring closures. For each of these pairs, differences are exhibited in the corresponding WLN descriptions.

The compactness of the WLN's of the compounds in Table I should be noted. For 5000 organic compounds offered to laboratory markets, 85% require 13 or fewer marks for the WLN's and 96%, 22 or fewer columns.<sup>4</sup> Even a file 10 times larger, of the U.S. Chemical-Biological Coordination Center (CBCC), shows a notation size of 13 or less for 63% and a size of 22 or less for 84%.<sup>5</sup>

TABLE I: 182 STRUCTURES CITED IN EARLY RHEOLOGICAL STUDIES  
(Arranged in increasing order of notational complexity)

| Code: —1              | U1               | —2                             | U2                             | E  | G  | M  | Q   | R                                                                                 | U | V     | W              | X      | Y                                                                                       | Z               | Ø    | O      |
|-----------------------|------------------|--------------------------------|--------------------------------|----|----|----|-----|-----------------------------------------------------------------------------------|---|-------|----------------|--------|-----------------------------------------------------------------------------------------|-----------------|------|--------|
| Key: —CH <sub>3</sub> | :CH <sub>3</sub> | —C <sub>2</sub> H <sub>5</sub> | :C <sub>2</sub> H <sub>5</sub> | Br | Cl | NH | OH  |  | : | >CO   | O <sub>2</sub> | >C<    |  (H) | NH <sub>2</sub> | zero | oxygen |
| R                     | G4               | 10H                            | 2Y2                            |    |    |    | QV1 | —HG—                                                                              |   | EYEE  |                | 11U5R  |                                                                                         | 1UY1U1          |      |        |
| 1R                    | GR               | 11H                            | 3O1                            |    |    |    | QV2 | 1R B†                                                                             |   | G2U1  |                | 1N1&R  |                                                                                         | 1Y2OV1          |      |        |
| 2R                    | GX               | 12H                            | 3O2                            |    |    |    | QV3 | 1R C                                                                              |   | GR B  |                | 1VOV1  |                                                                                         | 2OVVO2          |      |        |
| 2X*                   | GY               | 12R                            | 3O3                            |    |    |    | QV4 | 1R D                                                                              |   | GR C  |                | 1YVO1  |                                                                                         | GYGYGG          |      |        |
| 2Y*                   | I1               | 13H                            | 3V1                            |    |    |    | QV5 | 1VO1                                                                              |   | GR D  |                | 2N2&R  |                                                                                         | Q1Y7&6          |      |        |
| 3Y                    | I2               | 14H                            | 5O5                            |    |    |    | QVY | 1Y&Y                                                                              |   | GYGG  |                | 2OVO2  |                                                                                         | Q1YQ1Q          |      |        |
| 4Y                    | I3               | 15H                            | E1E                            |    |    |    | QY2 | 2O1Y                                                                              |   | I2U1  |                | GXGGG  |                                                                                         | QVR BQ          |      |        |
| 5H                    | 1R               | 16H                            | E1Y                            |    |    |    | SCS | 2OV1                                                                              |   | Q2U1  |                | GYG1G  |                                                                                         | ZR BQ1          |      |        |
| 6H                    | 1Y               | 17H                            | E2E                            |    |    |    | VH1 | 2OVR                                                                              |   | QR C  |                | QR BG  |                                                                                         | ZR CO1          |      |        |
| 7H                    | Q1               | 18H                            | G1G                            |    |    |    | VH2 | 2VO1                                                                              |   | VHO1  |                | QR CG  |                                                                                         | ZR DO1          |      |        |
| 8H                    | Q2               | 18R                            | G1Y                            |    |    |    | VHQ | 2VO2                                                                              |   | VHO2  |                | QR DG  |                                                                                         | L6TJ A‡         |      |        |
| 9H                    | Q3               | 1MR                            | G2G                            |    |    |    | VHR | 3OV1                                                                              |   | VHO3  |                | QY2&2  |                                                                                         |                 |      |        |
| E2                    | Q4               | 1OR                            | I1Y                            |    |    |    | WN1 | 3VO1                                                                              |   | VHO4  |                | R1O1R  |                                                                                         | 1UY&YU1         |      |        |
| E3                    | Q5               | 1V1                            | NC1                            |    |    |    | WNQ | 3VO2                                                                              |   | WSQQ  |                | WNR B  |                                                                                         | Q1Y2&12         |      |        |
| E4                    | Q7               | 1VR                            | NCR                            |    |    |    | WNR | 4OV1                                                                              |   | Z2U1  |                | WNR C  |                                                                                         | QX2&2&2         |      |        |
| EE                    | Q8               | 26R                            | OCO                            |    |    |    | Z1R | 5OV1                                                                              |   | ZR B  |                | ZR CE  |                                                                                         | L66J B6         |      |        |
| ER                    | QR               | 2M2                            | Q16                            |    |    |    | Z1Y | 7U5R                                                                              |   | ZR C  |                | L6UTJ‡ |                                                                                         | L6TJ AQ         |      |        |
| EY                    | QY               | 2MR                            | Q1R                            |    |    |    | Z2Y | E2U1                                                                              |   | ZR D  |                | L6VTJ  |                                                                                         | T66 CNJ         |      |        |
| FR                    | Z1               | 2O2                            | Q1Y                            |    |    |    | ZVH | ER B                                                                              |   | L5TJ‡ |                |        |                                                                                         |                 |      |        |
| G1                    | ZR               | 2V1                            | Q2Q                            |    |    |    |     | ER C                                                                              |   | L6TJ  |                |        |                                                                                         |                 |      |        |
| G3                    | ZZ               | 2V2                            | Q2Y                            |    |    |    |     | ER D                                                                              |   | L7TJ  |                |        |                                                                                         |                 |      |        |
|                       |                  |                                |                                |    |    |    |     |                                                                                   |   | L8TJ  |                |        |                                                                                         |                 |      |        |
|                       |                  |                                |                                |    |    |    |     |                                                                                   |   | T6NJ  |                |        |                                                                                         |                 |      |        |

\* The—X is a tertiary butyl terminal and—Y is an isopropyl terminal; their methyl branches (like units subscripts in empirical formulas) are understood by definition.

† Lone letters prefixed by a blank space locate ring positions for *methyl* groups. (That is, the groups are understood to be *methyl* unless otherwise specified.)

‡ Ring number citations are enclosed in carbocyclic L . . . J or heterocyclic T . . . J marks.

Equally important for information retrieval, these symbol sequences are self-orienting to yield a singularly correct indexing and a registering order. This uniqueness circumvents the need for analogs to the host of synonyms that plague chemical name indexes. The early Hollerith punched card files of WLN's were managed by placing the notations in alphanumerical order and securing a line-printer listing<sup>3, 4</sup>. The resulting indexes served as significant "IR" tools since terminal symbols for functional groups take precedence over those for carbon chains or benzene rings and symbols for all other ring structures take precedence over these.

Using WLN's, a computer today can check whether a structure is identical to one previously entered. For a duplicate, information can be read in to update that already stored and for a new structure the computer can assign a registry number and put it and associated information into addressable storage. In September 1967, a census revealed that 14 installations in the United States and Great Britain have combined WLN holdings of almost 800,000 structures.

Line notations stored as part of the computer records for structures can be searched for the presence of particular features. In addition, each symbol in the line notation can be listed in its part of a comprehensive alphanumeric order to provide a *permuted* index, showing the full notation at each entry as well as associated data.<sup>7</sup> For 4982 monomeric laboratory-use organic chemicals, the symbol-permutation program yielded 20,171 lines, corresponding to 4·1 permutations per compound.<sup>8</sup>

Notation systems have been suggested that afford unambiguous identification of a compound, but are non-unique in that for a given compound more than a single notation can be written. With adequate programming, such a notation system may attain usefulness in the generation of a unique connectivity record for computer-based retrieval systems. However, a unique, unambiguous notation, such as the WLN, will continue to be of broad interest.

For an appreciation of chemical notation systems and their use as of 1964-1965, the reports of the National Academy of Sciences—National Research Council should be studied.<sup>9, 10</sup>

## CONNECTIVITY RECORDS

The nineteenth century line formulas incorporated two types of structural information: the atomic groups and their connections. For some purposes of computer-based retrieval it is desirable to have the atomic groups and connectivity information separately available but yet forming a unity described as a connectivity table or record. Atom-to-atom connectivity tables have been investigated, but most single atoms in a molecule have no descriptive value for search purposes and an atom-to-atom connectivity matrix is a bulky record made up of descriptions of atoms and bonds. Hyde and associates<sup>11</sup> have shown the advantage of using a notation to derive an extremely

compact connectivity record. The approach can be appreciated by brief description. The delineation of the alkane



indicates that a tertiary CH group holds a primary  $\text{CH}_3$  side group. This structure can be generalised to



where A . . . G are any atomic groups. This representation implies that both E and F are connected to D. A connectivity "list" here reduces to recitation of the open-chain *branching transfer* in the string of the atomic groups, which are assumed to be numbered, as listed, from 1 through 7:

Atomic Groups:  
A B C D E F G

Branching Transfer:  
6-4

This expression implies that the "expected" 6-5 connection is transferred to 6-4. Where a ring closure is involved, an additional ring link must be recited. In place of the atomic groups, symbols may be introduced drawn largely from WLN's. Barnard and Wiswesser<sup>12</sup> have tried to reduce an appreciation of both WLN's and connectivity records, generated after Hyde and associates,<sup>11</sup> to the inspection of the few selected examples given as Table II, produced with permission from the original paper.<sup>12</sup> In a feasibility study,<sup>11</sup> it was concluded that such connectivity records average about 40 characters per compound when computer generated from a test collection of WLN's averaging 20 characters.

A connectivity record can on one hand be applied directly within the computer to answer search questions and on the other serve to generate special outputs, including fragmentation codes and classification numbers.<sup>13</sup>

## VISUAL STRUCTURAL DISPLAY

Chemists often would like a visually recognisable representation of a chemical structure as part of the answer to a search question. The on-line *computer generation* of such displays is an area for active study and several approaches have been reported. A computer output can be secured, corresponding to the previous input of punched tape or magnetic tape, resulting from the keyboarding of chemical structures on specially modified electric typewriters. A second possibility is the provision on-call of structures stored on microfilm.

A structural display has been secured from conventional punched cards and line printer characters.<sup>14</sup> The "dotplots" appear principally of interest where many similar structural outlines occur in the study of isomer possibilities and the equivalence of fused ring presentations.

Of greatest long-term interest is the "synthesis" of the visual structure display from more compact information stored in the computer, notably from a line notation or connectivity record. The production of visually pleasing chemical structure displays on an unmodified line printer has recently been demonstrated; the computer processing starts from stored connectivity table records, which in turn are computer-generated from a WLN input.<sup>15</sup>

TABLE II: EARLY VARIATIONS OF "RATIONAL FORMULAS" (1861-1867),  
WITH EQUIVALENT LINE NOTATIONS AND ATOM-CONNECTIVITY TABLES

| "Rational Formulas"<br>(1861-1867)                                                                  | Wiswesser<br>Line<br>Notation | Atom-Connectivity Table       |                        |               |
|-----------------------------------------------------------------------------------------------------|-------------------------------|-------------------------------|------------------------|---------------|
|                                                                                                     |                               | Atomic<br>Groups <sub>1</sub> | Transfers<br>Branching | Ring<br>Links |
| C <sub>8</sub> H <sub>10</sub> O, C <sub>8</sub> H <sub>8</sub> O, O, C <sub>8</sub> H <sub>8</sub> | 2OV2O2                        | ALOVLLOLA                     | —                      | —             |
| CH <sub>3</sub> CN·CO·Br                                                                            | NCIVE                         | NCLVE                         | —                      | —             |
| ClCH <sub>2</sub> ·CO <sub>2</sub> H                                                                | QVIG                          | QVLG                          | —                      | —             |
| H <sub>2</sub> N—CH <sub>2</sub> —CH <sub>2</sub> —CO <sub>2</sub> H                                | Z2VQ                          | ZLLVQ                         | —                      | —             |
| CHCl <sub>3</sub> ·CCl <sub>3</sub>                                                                 | GYGXGGG                       | GYGXGGG                       | 04-02, 06-04, 07-04    | —             |
| CO·OH—C(OH) <sub>2</sub> —CO·OH                                                                     | QVXQQVQ                       | QVXQQVQ                       | 05-03, 06-03           | —             |
| CH <sub>3</sub> ·CHI·COOH                                                                           | QVYI                          | QVYIA                         | 05-03                  | —             |
| CH <sub>3</sub> —CH·OH—CO·OH                                                                        | QYVQ                          | QYAVQ                         | 04-02                  | —             |
| C <sub>6</sub> H <sub>5</sub> —CH <sub>2</sub> —Br                                                  | EIR                           | ELTDDDDDD                     | —                      | 03-08         |
| C <sub>6</sub> H <sub>5</sub> ·CCl <sub>3</sub> H                                                   | GYGR                          | GYGTDDDDDD                    | 04-02                  | 04-09         |
| C <sub>6</sub> HBrHBrHNO <sub>2</sub>                                                               | WNR CE EE                     | WNTD'TDTDEE                   | 09-05, 10-07           | 03-08         |
| C <sub>6</sub> H <sub>5</sub> ·SO <sub>2</sub> ·OH                                                  | WSQR                          | WSQTDDDDDD                    | 04-02                  | 04-09         |

<sub>1</sub> Branching transfers are high-to-low connection *transfers* of the respective atomic groups in cited order. Ring closures are low-to-high connection *links*. The atomic group symbols are defined as follows:

Group: A D E G L Q T V W X Y Z  
Key: CH<sub>3</sub> :CH— Br Cl CH<sub>2</sub> OH :C— —CO— (:O)<sub>2</sub> C CH NH<sub>2</sub>

In the computer record, the atom-connectivity lists can be compacted as these examples: GYGXGGG-040206040704 and WNTD'TDTDEE-09051007-0308.

The display of chemical structures on the face of a cathode-ray tube was attained as early as 1959 by Opler.<sup>6</sup> This approach will become of increasing interest as cathode-ray tube output becomes common at general-purpose computer installations.

### MODIFIED CHEMICAL NAMES

Chemical names have undergone progressive systemisation, but offer only a feeble basis for the computer-based retrieval of chemical structures. Names, however, are familiar "tags" for the chemist and their inclusion as the part of an answer to a retrieval question is often desirable. Unfortunately chemical names present typographical complexities and consequently special practices<sup>10</sup> must be adopted for the transcription of names to the limited typography of conventional electronic-processing equipment.

In addition to the transcription of a chemical name, it must be kept within reasonable length and best within a fixed number of characters so that simple, compact, uniform formats are possible for computer input and output. Consequently long names must be contracted. *Contraction* of frequently cited functional groups, prefixes, and suffixes has been introduced, with the aim of achieving a compact result that can be understood by the chemist after only brief exposure to the practices adopted.<sup>12</sup> The approach is to use punctuation marks (or separating spaces) as systematic modifiers for the root names and to allow different meaning where the root symbol appears in different places. For example, the hundred-year-old symbol, BU, for the normal butyl group, denotes BUTYL in an initial position (followed by any other term), and BUTANE in a terminal position (not followed by any other term). The *-ate* suffix is expressed by the ampersand so BU& in a terminal position means BUTYRATE; the same symbol followed by any other term denotes BUTYRYL. The contractions BU&H, BU&A, BU&M, and BU&CL denote most succinctly BUTYRALDEHYDE, BUTYRIC ACID, BUTYRAMIDE, and BUTYRYL CHLORIDE, respectively. The *-ane* suffix, where it is needed in a non-terminal position, is indicated by an asterisk; thus, 1,4-butanediol is denoted by 1-4-BU\* DIOL. A check list has been organised of the most frequently used regularised contractions utilised in these studies.<sup>14</sup>

### EMPIRICAL FORMULAS

Empirical formula indexes have long been used for the retrieval of compounds. The power of an empirical formula index can be augmented by special treatment. For example, Skolnik<sup>17</sup> has used permuted empirical formulas bringing each heteroatom to the front of a formula. *Chemical Abstracts* is considering a heteroatom-in-context formula index, based on this principle.

Empirical formulas can be applied in the machine management of chemical information. In one mode of punching such formulas into Hollerith cards,<sup>4, 8</sup> fixed fields are employed for carbon and hydrogen

to  $C_{999}H_{999}$ , and the remaining elemental symbols are ordered alphabetically with their multiplier and separated from each other by single spaces. This format is economic in its space requirements. Listing of empirical formulas in alphanumeric order after an initial sort of the collection according to the 10 values of the A-digit of the BATCH number (see below) has proven a useful tool<sup>4, 5</sup>; this atomic-class sorting reduces drastically the number of empirical formulas to be examined in any search. The input of empirical formulas to a computer allows the computer generation of the formula weights.<sup>18</sup>

In Dow Chemical studies, the routine computer generation of the empirical formula of a compound from its WLN has proved feasible.<sup>19</sup> Indeed, the generated formula can be compared with the previously stored one, thereby providing a check. Garfield<sup>20</sup> has reported an algorithm for the calculation of the empirical formula from the conventional systematic name for an organic compound. Comparison of the resulting formula with that provided by an abstractor also permits a limited check.

## FRAGMENTATION CODES AND NUMBERS

Organic compounds have long been thought of in terms of structural fragments. For storage and retrieval purposes these fragments can be assigned unique addresses, which constitute a fragmentation code. For example, by locating all compounds having the addresses for both hydroxy compounds and carboxylic acids, all hydroxy-carboxylic acids are retrieved. Some degree of meaning can be worked into the addresses of a fragmentation code and thereby a system of fragmentation numbers can be evolved.

Many fragmentation codes have been described and have long been used in the hand, edge-notched card and Hollerith punched-card management of organic chemical information.<sup>9, 10</sup> Assignments by a chemist can be entered into a computer-based system and be "called" in answering retrieval questions. Additionally, listings by the structure fragments can be secured.

Of current interest is the computer assignment of compounds to the classes of a fragmentation code from other entered or generated information, notably a line notation or connectivity record.<sup>11, 12, 13</sup> Bowman and associates<sup>19</sup> generate computer-assigned fragmentation numbers from WLN's. Hyde and Thomson<sup>11, 12</sup> generate an open-ended fragmentation index from WLN input; within each structural class, the line notations for the fragments are listed in permuted-index arrangement.

## CLASSIFICATION NUMBERS

"Class" codes had long been used with collections of knowledge. Library classification schemes offer familiar examples. Empirical formulas written according to established rules may be viewed as classification numbers. Their power as retrieval tools can be enhanced by permutation.<sup>17, 21</sup> However, *digital* classification numbers avoid the

difficulties of mixed alphabetic-numeric orderings and offer possibilities for securing a balanced statistical distribution for the ten values of each digit.

W. J. Wiswesser in 1953<sup>22</sup> described a general structural-atomic classification number for organic compounds, to which he gave the mnemonic designation BATCH Number. The assignment of values for the B and A digits of the BATCH Number was revised in 1964; the scheme for the assignment of BATCH Numbers was detailed and salient uses for this tool were discussed.<sup>4</sup>

The implications of the BATCH Number can be appreciated from a brief description of its five digits. If the first digit identified the kinds (and pluralities) of rings in the organic structures, using zero through nine, this digit alone can subdivide a collection into 10 Basic or ring divisions. If the second digit identifies the kinds of elements in the structure, using zero for the simplest C, H, or C, H, O structures and nine for the rarest fraction containing metallic elements, this digit by itself can subdivide the collection into 10 Atomic subdivisions. If the carbon and hydrogen atom counts are reduced in a statistically efficient manner to single digits, recorded in the fourth and fifth digits of the BATCH Number, these CH-digital specifications can be combined with the BA-combination of a given structure or closely related group of isomers to increase retrieval selectivity ten thousandfold. Finally a fifth digit, received as the third digit of the BATCH Number, can provide additional discriminating power by denoting the number of atoms other than carbon or hydrogen, to a total of "ten or more"; this Total heteroatom digit is less selective than the others because of the prominence of compounds having three to five heteroatoms.

The BATCH Number has been demonstrated to be a valuable tool in the organisation of structural-atomic indexes, termed BATCH Directories<sup>23</sup> that allow many retrieval requests to be answered by simple examination, if the computer-oriented name and other relevant information are included.

Freeman<sup>24</sup> has described for dyestuffs a classification number having elements of the BATCH Number into which greater functional group discrimination has been worked.

The BATCH Number and other easily mastered, well-devised classification numbers are conservative measures that can be adopted in the hand-management of small collections<sup>25</sup> and be continued when machine-management becomes appropriate. They can serve as effective "screens" even where more sophisticated methods such as the WLN are available as retrieval tools.

## INORGANIC COMPOUNDS

The information retrieval tools discussed earlier in this article are principally applied to organic compounds. The computer-based retrieval of inorganic structures is of less applied interest and has been given less attention. For some purposes, subject indexing with parts

of the inorganic chemical names is useful, and for other purposes empirical formulas are preferred. But more sophisticated "IR" tools are needed.

Wiswesser and Smith have established rules for writing computer-oriented equivalents to conventional inorganic formulas; hierachial rules ensure that for a given structure only a single formula is written.<sup>7</sup> The formulas can be placed in alphanumeric order by machine sorting and be listed with associated information, including their computer-oriented names.<sup>24</sup> The listing of such computer-oriented inorganic formulas for over 1500 simple and complex *anions* with their systematic and common names has provided an insight into the inconsistencies and inelegancies in the assignment of some names.<sup>2</sup>

A classification number has been devised for inorganic compounds, elements, and alloys, namely the ABET Number.<sup>25</sup> This elemental classification number is distinguished by a prefixed asterisk or hyphen "sign"; consequently either the BATCH Number or ABET Number can be stored in a single 5-character field. In brief, the A, B, and E digits refer to the periodic subgroups in which fall the first three elements of the computer-oriented formula; the T digit value corresponds to the total non-hydrogen atoms present in the formula, omitting molecules of solvation and with a zero being recorded for 10 or more such atoms. The ABET Number with computer-oriented chemical names is of value in the development of so-called ABET Directories<sup>26</sup> that have usefulness paralleling that of the BATCH Directories described above.

### REMARKS

The aim of this paper has been to delineate some important retrieval tools for chemical structure information. It has not been the intent to show these tools in the context of a total data system. The design of a system is a most active area of effort and is influenced by the system size, the nature of the data stored, the expected depth and range of search questions, and the extent and nature of the computer installation available.

The machine-management of data can be undertaken by even small laboratory groups. Initially much can be accomplished with the simplest techniques for Hollerith punched cards. With the adoption of conservative measures, such as classification numbers, computer-oriented chemical names, and an efficient line notation, many search questions can be answered from directory printouts from a line printer. As the system grows, the measures adopted are applicable to computer-based operations.

### LITERATURE CITED

- <sup>1</sup> Frankland, E., "Contributions to the Notation of Organic and Inorganic Compounds," *J. Chem. Soc.* 19, 372-395 (1866).
- <sup>2</sup> Wiswesser, W. J., "A Line-Formula Chemical Notation," Thomas Y. Crowell Co., New York, 1954; "A Pictorially Direct Notation," offset, 1950.

- <sup>2</sup> Smith, E. G., "The Wiswesser Line-Formula Chemical Notation," McGraw Hill Book Co., New York, 1968.
- <sup>3</sup> Barnard, A. J., Jr., C. T. Kleppinger and W. J. Wiswesser, "Retrieval of Organic Structures from Small-to-Medium Sized Collections," *J. Chem. Doc.* 6, 41-48 (1966).
- <sup>4</sup> Wiswesser, W. J., unpublished findings.
- <sup>5</sup> Opler, A., and N. Baird, "Display of Chemical Structural Formulas as a Digital Computer Output," *Am. Doc.*, 10, 59-63 (1959).
- <sup>6</sup> Sorter, P. F., Granito, C. E., Gilmer, J. C., Gelberg, A., Metcalf, E. A., *J. Chem. Doc.* 4, 56-60 (1964); Granito, C. E., Gelberg, A., Schultz, J. E., Gibson, G. W., Metcalf, E. A., *ibid.*, 5, 52-55 (1965); C. E., Schultz, J. E., Gibson, G. W., Gelberg, A., Williams, R. J., Metcalf, E. A., *ibid.* 5, 229-233 (1965).
- <sup>7</sup> Barnard, A. J., Jr., W. C. Broad, C. T. Kleppinger, and W. J. Wiswesser, "Some Techniques for the Machine Management of Small Chemical Data Systems," Proceedings of the Wiswesser Line Notation Meeting of the Army Chemical Information and Data Systems Program, 6th-7th October, 1966, J. P. Mitchell, Editor, Edgewood Arsenal Special Publication, EASP 400-408, 1968, pp. 85-101.
- <sup>8</sup> NAS-NRC, "Survey of Chemical Notation Systems," Publication 1150, National Academy of Sciences-National Research Council, Washington, D.C., 1964, 467 pp.
- <sup>9</sup> NAS-NRC, "Survey of European Non-Conventional Chemical Notation Systems," Publication 1278, National Academy of Sciences-National Research Council, Washington, D.C., 1965, 78 pp.
- <sup>10</sup> Hyde, E., F. W. Matthews, L. H. Thomson and W. J. Wiswesser, "Conversion of Wiswesser Notation to a Connectivity Matrix for Organic Compounds," *J. Chem. Doc.*, 7, 200-204 (1967); Hyde, E., in Proceedings of the Wiswesser Line Notation Meeting of the Army Chemical Information and Data Systems Program 6th-7th October, 1966, J. P. Mitchell, Editor, Edgewood Arsenal Special Publication, EASP 400-408, 1968, pp. 57-67.
- <sup>11</sup> Barnard, A. J., Jr., and W. J. Wiswesser, "Computer-Serviced Management of Chemical Structure Information," *Lab Management*, 5, No. 10, 34-36, 38, 40-44 (October 1967).
- <sup>12</sup> Thomson, L. H., E. Hyde and F. W. Matthews, "Organic Search and Display Using a Connectivity Matrix Derived from Wiswesser Notation," *J. Chem. Doc.*, 7, 204-209 (1967).
- <sup>13</sup> Wiswesser, W. J., "The Dot-Plot Computer Program," paper, Division of Chemical Literature, 152nd meeting, Am. Chem. Soc., New York, September 1966; Proceedings of the Wiswesser Line Notation Meeting of the Army Chemical Information and Data Systems Program, 6th-7th October, 1966, J. P. Mitchell, Editor, Edgewood Arsenal Special Publication, EASP 400-408, 1968, pp. 103-116.
- <sup>14</sup> Barnard, A. J., Jr., C. T. Kleppinger, W. J. Wiswesser, "Computer-Oriented Chemical Names," *J. Chem. Doc.*, 6, 48-57 (1966).
- <sup>15</sup> Technical Information Services, J. T. Baker Chemical Co., "Checklist of Transcriptions and Contractions for Computer-Oriented Chemical Names," Phillipsburg, N.J., July 1966, 11 pp.
- <sup>16</sup> Skolnik, H., and A. Clow, "A Notation System for Indexing Pesticides," *J. Chem. Doc.*, 4, 22-227 (1964).
- <sup>17</sup> Barnard, A. J., Jr., unpublished Fortran IV program.
- <sup>18</sup> Bowman, C. M., F. A. Landee, and M. H. Reslock, "A Chemically Oriented Information Storage and Retrieval System," *J. Chem. Doc.*, 7, 43-47 (1967); Bowman, C. M., F. A. Landee, Betsy P. Smith, and M. H. Reslock, "Proceedings of the Wiswesser Line Notation Meeting of the Army Chemical Information and Data Systems Program, 6th-7th October, 1966, J. P. Mitchell, Editor, Edgewood Arsenal Special Publication, EASP 400-408, 1968, pp. 49-56.

- <sup>20</sup> Garfield, E., "An Algorithm for Translating Chemical Names to Molecular Formulas," *J. Chem. Doc.*, 2, 177-179 (1962).
- <sup>21</sup> Garfield, E., "Generic Searching by Use of Rotated Formula Indexes," *ibid.*, 3, 97-103 (1963).
- <sup>22</sup> Wiswesser, W. J., "Literature Sources of Mammalian Toxicity Data, With Special Emphasis on Tabulating Machinery Applications," *Advances in Chemistry Series No. 16*, American Chemical Society, Washington, D.C., 1956.
- <sup>23</sup> J. T. Baker Chemical Company, Phillipsburg, N.J., "J. T. Baker BATCH Directory," September, 1965; "J. T. Baker Special BATCH Directories for Halogen Compounds, Hydroxy Compounds and Cyclic Nitrogen Compounds," January, 1966; "J. T. Baker Special BATCH Directory for Polymer-Use Chemicals," September 1966.
- <sup>24</sup> Freeman, J. F., "A Novel Organization Code for Organic Structures Based on Functional Groups," *J. Chem. Doc.*, 6, 184-187 (1966).
- <sup>25</sup> Barnard, A. J., Jr., and W. J. Wiswesser, "Use of the BATCH Number with Hand-Manipulated Files," *ibid.*, 6, 188-189 (1966).
- <sup>26</sup> Barnard, A. J., Jr., W. J. Wiswesser, and W. C. Broad, unpublished work.
- <sup>27</sup> J. T. Baker Chemical Company, Phillipsburg, N. J. "J. T. Baker ABET Directory for Inorganic Compounds," April, 1967.

The Proceedings of the WLN Meeting, cited in (8), (11), (14), and (19) have now been published as AD-665, 397, Clearing House for Scientific and Technical Information, Springfield, Virginia 22151, U.S.A.

(Solely for reasons of limited space it has proved necessary to print the above paper in a form briefer than that originally received. I am grateful to the authors for permitting publication of the truncated version, and I have available for loan a copy of the full paper, together with copies of several of the references cited. Readers of this article would find many points of interest in Dr. M. F. Lynch's paper "Storage and retrieval of information on chemical structure by computer," *Endeavour* (1968), 27, 68-73. Dr. Lynch compares WLN with other forms of notation.—Editor.)

## N.L.L. TO THE RESCUE

Earlier this year I was fortunate enough to participate in a seminar on the use of the scientific literature held at The National Lending Library for Science and Technology, Boston Spa, Yorkshire. The seminar was one of a series held for university staff with the aim that the participants would themselves organise similar courses for undergraduate and post-graduate students. It is impossible to condense a complete course into a single short article, therefore I shall concentrate on just two topics, together with a resumé of the work of N.L.L. If my personal receiving and transmitting system shows evidence of malfunction I ask the staff of N.L.L. to forgive my errors, and remember that no retrieval system gives 100% recall or 100% precision.

### What and Where is N.L.L.?

Scientific periodicals have been flourishing for rather more than 300 years, with the total number of titles published being doubled about every 15 years since 1750. Not all these titles are still being published, but even so there are well over 30,000 periodicals at present covering the world's scientific literature—and the number can be expected to increase rapidly as the developing countries are able to enter the field. No university or industrial library can afford to subscribe to *all* the journals that *may* be relevant, and the N.L.L. was established (or rather grew out of existing organisations) so that scientists could obtain, wherever possible, a return-of-post loan service for periodicals in any language, for Russian scientific books, and, more recently, for most post-graduate level English language technical books. In addition to these better known functions N.L.L. holds large numbers of (mainly U.S. and British) technical reports (many of these only as microfilms or microfiches), keeps and catalogues translations into English (mainly from Russian), and is responsible for the cover-to-cover translation of several Russian periodicals.

To carry out its functions with the maximum speed N.L.L. is not organised as a conventional library. Except for translations there is no normal catalogue, periodicals being shelved in alphabetic order in one of three separate stores, with the 1000 most used journals being in A store closest to the centre of activities. Incoming requests are sorted; an assistant is given a stack of requests to search for; and as each volume is found it is placed on the conveyor system which links all parts of the library with the processing centre.

N.L.L. is housed in part of a former ordnance factory, on a rural site, and there is sufficient acreage available for it to extend to several times its existing area. By the time all the space is filled periodicals, as we know them, will probably have been replaced by less wasteful means of communication. One's first impression of N.L.L. must be that the library revolution is a hundred years behind the industrial revolution. There is primitive mechanisation in the form of slow-

moving, clanking, conveyors, but all the sorting and all the searching is carried out by hand, and there are always dozens of girls scurrying about amongst the miles of shelving. In a region of plentiful, relatively inexpensive, woman-power the present system achieves a high cost-effectiveness, but within a few years one feels that automation must enter and ultimately triumph. [In fairness it must be added that first impressions are often misleading, and there is certainly nothing out-dated about the methods of the photocopy department.]

N.L.L. subscribes to 28,000 publications, and yet keeps its collections in order and deals with about 3,000 requests per day with a total staff (including maintenance and industrial) of only 200. The reason for this is soon explained: 50% of the demand is for journals less than three years old, and 50% of requests are for 3% of titles. The most requested journals are those which it was thought every large library already took, and not surprisingly, the bulk of requests arrive immediately after publication or immediately after publication of the relevant abstracts journal.

Individuals cannot, in general, borrow from N.L.L., but must channel requests through their own library. There is, however, a reading room at Boston Spa open to any member of the public during normal working hours, Monday to Friday. This facility is extremely useful for anyone wishing to undertake a literature search, since the reading room has on its open shelves all the important abstracts periodicals back to 1960; and requests for day-loan of specific journals are handled within a few minutes.

For those readers who would like to know how their requests are processed N.L.L. has produced a 20-minute, 16mm. colour film describing its activities, which may be borrowed, free of charge, on request.

### The Use of Abstracts

The Royal Society was able to produce a catalogue, in only a few handsome volumes, to the complete world scientific literature of the nineteenth century. Today several hundred abstracts journals are needed, or at least are published, to give complete coverage; and a single current year of *Chemical Abstracts* is itself larger than the complete Royal Society catalogue. One's problem therefore is not only to find one's way about the literature, but to begin a step further back by knowing how to make efficient use of the available abstracting services.

We are all familiar with *Rheology Abstracts*, which is small and slow, but useful because it cuts across the normal boundaries between the sciences. For this reason it is not perhaps typical of abstracts publications, which range in size from *Chemical Abstracts* (dealing with 250,000 abstracts a year from 11,000 publications) to small occasional pamphlets such as *Cheese Abstracts* (produced at Sandwich).

Abstracts periodicals are essentially of two distinct types: Indicative, giving only full titles and subject headings; and Informative, giving a summary of each paper. They are used for two purposes: Current Awareness, where rapidity of production is important, dealing with essentials only, in a narrow field; and Retrospective Searching, where one requires wide coverage and an informative abstract, preferably with a cumulative index. The production of abstracts journals is now being speeded up by the use of computers, most frequently for producing indexes—unfortunately the possession of a computer occasionally leads a publisher to compile additional, unnecessary and extremely complicated indexes.

Current awareness periodicals are produced rapidly after initial publication, often using proofs of the contents page of the original journal as their raw material. They are produced cheaply, and are meant to be thrown away later when full abstracts become available. Since titles alone are given the searcher may miss some relevant papers.

*Chemical Titles*, published by the Chemical Abstracts organisation, covers 700 journals with an average delay of only two weeks. It is worth mentioning as an example of a periodical containing a KWIC index (an O.K. term in information retrieval circles), and is available on computer tape so that requested keywords may be readily retrieved. [KWIC is an acronym for Key Word In Context, which means in the case above that certain of the most meaningful words in the title (termed *Keywords*) are underlined or printed in bold type and subsequently indexed. The searcher selects a combination of keywords, and is then directed to all papers containing that combination. The system is simple and rapid to use, but relies on a title giving a description of a paper's contents that is both accurate and complete.]

Another type of current awareness publication, represented by *Current Contents (Physics, Chemistry, and Life Sciences)*, reproduces simply a photocopy of the contents pages of the journals covered, together with addresses of authors. No translations are given for foreign language titles. These publications are compiled with computer assistance, and are offshoots of *Science Citation Index (q.v.)*.

The largest informative abstracts periodical is, of course, *Chemical Abstracts*, which covers about 11,000 titles. Its very size is sometimes frighteningly off-putting, but the same organisation produces other smaller, more specialised journals, such as *Polymer Science and Technology*. *Chemical Abstracts* covers an enormously wide field, arranged under 80 subject headings, and is complete with a battery of indexes, which at present occupy seven volumes every six months. In addition there are five year collected indexes, published with a time-lag of several years; those for 1962–66 will occupy 28 large volumes. The use of computers has reduced the delay between paper and abstract from more than a year to three or four months, and in addition has given much more comprehensive indexing.

*Science Abstracts* is a much smaller affair; *Physics Abstracts* yielding about 40,000 abstracts a year from 1,000 journals; it is important to remember that applied physics is not covered. The only general tool in the engineering field is *Engineering Index*—which despite its title contains informative, sometimes critical, abstracts. This publication has a bias towards English language papers, and searching is not always easy, since one must look under the subject index, and hope that the correct headings have been selected. The Electronics and Plastics sections, now produced with computer aid, do, however, have different subject headings and much improved indexing.

*Biological Abstracts* lists 200,000 abstracts per year from 7000 journals. Each issue has several indexes, one index, BASIC, being a KWIC index. One must not forget *Bioresearch Index*, with its KWIC subject index. At first sight this might look like a current awareness periodical, but it covers some material omitted from *Biological Abstracts*.

The medical field is covered by *Excerpta Medica* and *Index Medicus*, both of which cover about 3,000 titles. The former, giving informative abstracts, appears in many separate monthly sections—sometimes it may be difficult to determine which is the section you really need. *Index Medicus* does not give informative abstracts. It has a complicated subject index (MESH), with an average of three subject headings per article. One particularly important feature of this journal is that it forms the basis for MEDLARS (Medical Literature Analysis and Retrieval System), the first fully computerised information retrieval system.

Incoming literature is searched by human indexers, who attach about 10 subject headings, selected from an official vocabulary (*thesaurus*), to each paper, three of these headings being accorded priority. The raw data are then put on magnetic tapes of two types: (i) GRACE tape, with only the three major subject headings, contains typesetting instructions to produce *Index Medicus*; (ii) the Compressed Citation File, which contains the bare bones of all data in coded form, without typesetting instructions. Duplicate copies of the latter tapes are sent from U.S.A. to several world-wide centres, including Newcastle University for U.K.

The tapes, which go back to mid-1963, contain more subject headings than *Index Medicus*. They are particularly useful for performing very exhaustive searches, or for tracing the intersection of two distinct ideas (e.g. the concurrence of two diseases)—both cases with a large amount of "eyeballing," where user fatigue can lead to inefficient searching.

At present the British end of MEDLARS is organised, on an experimental basis, as a free service by N.L.L., to whom all enquiries should be directed. About 30 searches per week are now carried out, all the week's searches being performed in a single batch process. N.L.L. act as an intermediary between user and computer because

of the complexities involved: the official vocabulary changes from year to year, and care must be taken not to ask a silly question that could generate tens of thousands of answers.

Computer tapes also form the basis of a printed retrieval publication of unorthodox type. The producers of *Science Citation Index* record not just the title of all published papers in about 1900 journals, but list in alphabetic order the references cited at the end of each paper. A key index then gives details of all authors citing a particular paper, and there is a host of other computer controlled indexes—some of which are more useful than others. The great advantage of the method is that it enables the rapid follow-up of references cited by relevant papers—a technique which an ASLIB survey showed the majority of scientists to prefer.

Although providing a useful, but expensive (about £500 a year), facility *Science Citation Index* does not replace standard abstracts journals. Its coverage is very patchy (but improving), and no steps are taken to remove irrelevant references (e.g. medical men seem fond of quoting Shakespeare). The reader must ask the right questions, and even so a relatively large amount of non-significant material is retrieved.

The above survey has been very selective, and review publications have not been mentioned. For those who require a full list of publications N.L.L. will provide free copies of "A KWIC Index to the English Language Abstracting and Indexing Publications Currently Being Received by N.L.L.," and "KWIC Index to Some of the Review Publications in the English Language Held at the N.L.L."

### **Personal Record Keeping**

Having traced a reference using an abstracts periodical the searcher then normally wishes to transfer the information to a personal record system. Most conventionally one has a book, or card file, arranged in author or subject order, each entry containing details of author, title of article and where published, keywords and cross-references. Sooner or later such a system will grow to an unmanageable size, particularly if the amount of cross-referencing is increased. An improved system would, ideally, have: speedy retrieval; lack of special order; all the cross-references on a single card; no distinct categories; and would be portable.

There are two ways of achieving these objectives: optical coincidence cards; and edge-notched cards. In the former system cards are printed with a numbered grid, each square of which can refer to a document. Each card represents one subject heading, and holes are punched wherever the document number is relevant. Documents (in a separate ordered file) can be retrieved by superposing several subject cards and noting which holes are common to each. The drawback to this system is that both subject and document cards are required, the former type being fairly expensive.

For most personal files edge-notched cards are to be preferred. These cards are bordered with one or more rows of circular holes, each hole representing some coded piece of information. Whenever a particular code number is relevant that hole is converted into an open notch. Retrieval is carried out by inserting a long needle through the pack of cards, which is then lifted and shaken, allowing notched cards to fall. A system using edge-notched cards is compact, allowing 10,000 cards to be stored in a single box, and inexpensive, since only one set of cards is required.

Retrieval is through concepts coded by numbers or letters (e.g. using first three letters of catchword). If the numbers are selected at random then a dictionary is needed to complete the system. Coding can be direct, with one notch to each number or letter, or indirect. The former method is extremely extravagant in its use of notches and consequently is rarely found except in very small systems.

The simplest indirect coding for numbers is to use successive holes to represent 7, 4, 2, 1 and parity check. With two needles any number 0-9 can then be retrieved. Similarly 0-99 will need 10 holes and 0-999 15 holes. One letter per hole is also wasteful, and so it is common to divide the alphabet into groups of nine letters and code as for numbers. If groups of corner holes, or multiple rows of holes, are used then even more compact coding is possible.

One often wishes to find a reference by a known author. This can be done by coding each card for the first three consonants of the author's name. Retrieval by those letters above is often sufficient without the need for secondary sorting.

Almost certainly one's coding system will need to be open-ended, so that new concepts can be added as the system grows. It may be convenient to represent each concept by two or more random numbers, each of which has to be retrieved in turn to recover the required documents. If there are more than six concepts per card it will probably be necessary to create new terms which describe certain combinations of primary terms, and sorting may need to proceed by stages.

Those of us who at present have only small personal files will probably wish to defer the transfer to edge-notch cards until our system has become unmanageably large, by which time transfer of all the records will take so long as to be practically impossible. It was suggested that a much better procedure was to use edge-notched cards for the smallest systems, but at first not to cut any notches, and use the cards in any way desired. Later conversion is then a relatively simple process.

D. W. HADLEY.

## CLASSIFICATION, INFORMATION RETRIEVAL AND RHEOLOGY

by Albert Yanez, licencié-ès Lettres, Head of the Automatic Data Processing Branch, CEDOCAR. (Centre de Documentation de l'Armement) Paris, France.

### Summary

Information retrieval must evolve to adapt itself to the new demands placed upon it by the scientific and technical "revolution." Electronic machines allow a new classification to be substituted for the traditional ones, namely scientific language itself—that is to say, the formulation of concepts using descriptors which, co-ordinated by logical operators, allow a more rapid and certain retrieval.

Today, there is no longer any real problem of classification, but only one of the constitution "thesauri" and the creation of descriptors.

These descriptors can be classified, or presented, in various ways which we have tried to illustrate. The essential thing is that the languages of information retrieval should give the image of a universal language.

### The literature explosion

The scientific and technical revolution which began at the end of the second world war has given rise to an "explosion" in published work, which has necessitated a fundamental revision of the problems of transfer of information and of communication between libraries and their readers.

These libraries are becoming centres of information processing, and a new methodology is developing, linked to the possibilities offered by the computer, according to an irreversible process.

The proliferation of documents, the diversification and interdependence of sciences and techniques, the shortening of the useful life of a document, and new demands made for a rapid and effective transfer of information, make the methods of the librarian of the last century useless.

Just as the small money lenders merged to form large banks, and still continue to do so, the small libraries are moving towards the creation of multidisciplinary systems; which are not any the less in touch with the user despite their being bigger and more powerful.

### Classifications

The field which has irrevocably changed is that of classification.

The art of classification has gradually become a science in its own right, called "taxonomy," whose golden era was at the end of the 19th century.

Just as a search for symmetry has appeared in the fine arts—in architecture, for instance—beautiful tree-like structures were constructed for each science and art which, from the trunk to the outermost tip, were supposed to represent and contain the whole science.

That which had some sense when it was just a matter of the vegetable or animal world, where everything is completely fixed (whatever the evolutionary theoreticians might say about it) becomes impracticable in a science in an exploding state. A universal classification has even been sought for, just as the alchemists used to seek the philosopher's stone—as if such a classification could be the universal panacea. I have even taken part myself in meetings of "well-known classifiers" where we tried to determine if the function created the organ or if the organ preceded the function, just as people would try to find out which came first—the chicken or the egg.

Don't let us throw stones at the workers in the field of classification. It is evident that for a long time, a good classification has been the only means of going from "heading" by "sub-heading" up to the subject under investigation, a little as Ariadne used her thread—but today the maze is longer than the thread and the thread is fraying. . . .

This is so true that such and such a learned society, not knowing any longer just where the boundaries of its activities lie, still hopes to find the necessary and sufficient classification, which will define its boundary.

The trouble is that all classification supposes some choice with an objective in view, within the limits of a particular context. If I were a naturalist, I would say that the dog is a carnivore, and in doing this I would ignore the other characteristics of a dog to concentrate on the nature of its teeth, and I would conclude, perhaps wrongly, that its canine teeth guarantee that it eats nothing but meat.

If I were a physiologist or research worker in a laboratory, I might class the dog as one of my test instruments.

If I were a policeman, I would count it among my weapons.

In other words, I choose for the dog a characteristic which serves my purposes, and then make a universal truth of it.

I am pleased to note that Sherwood Anderson, who was in part the spiritual father of Faulkner, considered that there are some wonderful truths in the world because they contain shades of meaning, but that these truths become either lies or falsehoods when a man appears on the scene and wants to make them his own truths. To desire at any price specialised classifications in a universe in a state of flux, badly known and badly delineated, is to desire the mutilation of truth at any price.

#### **Information retrieval**

"Alright," you say, "but what is the solution? We know quite

well that our classifications are always imperfect and to be reshaped—you propose to suppress them entirely—a negative and quite useless proposition—well, our problem of information retrieval does exist and gets worse every day.”

Certainly not. I answer that with the fortunate advent of optical or electronic machines which allow us to do what previously was impossible, it is the whole concept of retrieval which must evolve, and it is time for us to relegate our classifications to their proper place.

In effect, we have at our disposal a single and accurate universal classification—that is, the natural language itself, or more exactly concepts expressed in natural language.<sup>1</sup> When a concept is born with such a force that it must needs be expressed, language creates a new word or expression to classify it.

For instance, the study of the deformation and flow of materials, subject to stress, became the object of systematic research and was called “rheology.” This does not mean to say, however, that everything which is a deformation, flow, material or stress belongs only to rheology, and ceases to belong to other fields which have need of these concepts.

Today it is *easy* to co-ordinate descriptors, to associate them by logical expressions and thence to give a precise answer to a precise question, in the absence of any classification.

The document represented in Fig. 1 has been indexed, that is to say, tagged by a list of the most significant concepts it contains. This document will be retrieved with certainty when the question posed contains some of these descriptors.

The problem has become displaced. Instead of a problem of classification we are faced with a problem of language, terminology, identification of concepts, and rapid access to these concepts, either to index these documents or retrieve them.

Certain classifications or “word arrangements” may be found useful, albeit of secondary importance to the system, to gain access more rapidly and with greater certainty to the descriptors.

We can have a simple alphabetical classification (Fig. 2), or an alphabetical classification within a more general field (Fig. 3), or an alphabetical classification based on the words present in the descriptors (Fig. 4) or “hierarchical” classifications (Fig. 5) or synoptic tables of context (Fig. 6). It should be noted that all these displays are easily realised and updated by a computer, and can be modified without the information retrieval system suffering, because it doesn't depend on it.

It has been proved that libraries or information processing centres, having decided to automate their services and having been persuaded that they ought to cover a particular scientific field, have aspired to compile a specific thesaurus. It is also proved that every thesaurus has a rapid tendency to develop along unsuspected lines. In place

# PHYSICAL AND RHEOLOGICAL PROPERTIES OF NITROSO RUBBERS.

Descriptors: (\*Heat-resistant materials, Synthetic rubber), (\*Synthetic rubber, Physical properties), Rheology, Thermal expansion, Phase studies, Transition temperature, Chemical analysis, Solvents, Polymers, Spectrum (Infrared), Nuclear magnetic resonance, Viscosity, X-ray diffraction analysis, Differential thermal analysis, Thermogravimetric analysis, Fluorine compounds

Several nitroso gums were partially characterized. Linear expansion coefficients of a typical nitroso gum were determined to be 0.000076 from -72°C up to the glass transition with the glass transition

FIGURE 1.

—AUA  
Voir COMPOSES RHENIUM  
«RHENIUM»

## RHEOLOGIE

- Notions diverses de mécanique  
Rat. MECANIQUE FLUIDES  
Voir VISCOSITE  
ECOULEMENT FLUIDE  
ELASTICITE  
PLASTICITE  
«RHEOLOGY»

## Rhéostats

Emp. RESISTANCES VARIABLES

## Rhin+

FIGURE 2.

|         |                      |
|---------|----------------------|
| 2015021 | RHEOLOGIE            |
| 2015023 | SHOCK WAVES          |
| 2015023 | ONDES CHOC           |
| 2015024 | SKIDDING             |
| 2015024 | DERAPAGE             |
| 2015025 | SPIN                 |
| 2015025 | VRILLE               |
| 2015026 | STATICS              |
| 2015026 | STATIQUE             |
| 2015027 | TORQUE               |
| 2015027 | COUPLE               |
| 2015028 | VELOCITY             |
| 2015028 | VITESSE              |
| 2015029 | AEROTHERMODYNAMICS   |
| 2015029 | AEROTHERMODYNAMIQUE  |
| 2015030 | ANGULAR ACCELERATION |

FIGURE 3.

|                 |                        |
|-----------------|------------------------|
|                 | REVOLUTIONS PER MINUTE |
| CORPS           | REVOLUTION             |
| NOMBRE          | REYNOLDS               |
| INCOMPATIBILITE | RH                     |
| ALLIAGES        | RHENIUM                |
| COMPOSES        | RHENIUM                |
|                 | RHENIUM                |
|                 | RHEOLOGIE              |
|                 | RHIZOPODES             |
| ILE             | RHODES                 |
| ALLIAGES        | RHODIUM                |
| COMPOSES        | RHODIUM                |
|                 | RHODIUM                |

FIGURE 4.

## MECHANICS

CONTINUUM MECHANICS  
FLUID MECHANICS  
DYNAMICS  
AERODYNAMICS  
HYDRODYNAMICS  
HYPERSONICS  
SUPERAERODYNAMICS  
SUPERSONICS  
TRANSONICS  
STATICS  
AEROSTATICS  
HYDROSTATICS  
SOIL MECHANICS

## PROPERTIES

DEFORMATION  
CREEP  
PLASTIC FLOW

FIGURE 5.

FLUID DYNAMIC PROPERTIES  
ABLATION  
AERODYNAMIC HEATING  
AEROELASTICITY  
AEROTHERMOELASTICITY  
AREA SUCTON  
BOUNDARY LAYER  
LAMINAR BOUNDARY LAYER  
TURBULENT BOUNDARY LAYER  
BOUNDARY LAYER CONTROL  
BOUNDARY LAYER TRANSITION  
BUBBLES  
CAVITATION  
EXPLOSION BUBBLES  
BUOYANCY  
DOWNWASH  
DRAG  
DRIFT  
FLUTTER  
GROUND EFFECT  
GUST LOADS  
HYDROSTATIC PRESSURE  
JET ACOUSTIC OSCILLATIONS  
LIFT  
LIQUID JET  
LOAD

STRUCTURAL PROPERTIES  
MICROSTRUCTURE  
CRYSTAL STRUCTURE  
BRILLOUIN ZONES  
LATTICES  
COLOR CENTERS  
SYMMETRY (CRYSTALLOGRAPHY)  
ISOTROPISM  
ANISOTROPISM

MECHANICAL PROPERTIES  
DEFORMATION  
CREEP  
PLASTIC FLOW  
STRESSES  
SHEAR STRESSES  
THERMAL STRESSES  
BRITTLENESS  
HYDROGEN EMBRITTLEMENT  
FAILURE (MECHANICAL)  
CRAZING  
FATIGUE (MECHANICAL)  
FRACTURE (MECHANICAL)

FIGURE 6.

of a precise scientific language there have been developed, due to the "ivory tower" type of mentality, "specialised" languages and differing dictionaries, which together have led to a state of considerable disorder.

The U.S. Department of Defense has reacted promptly to this situation. Just let us note that it was necessary to start the LEX project (because lex = lexicon = law) and to use the authority of COSATI (Council for Scientific and Technical Information) to bring back the lost sheep to a wiser frame of mind, and the LEX project (final report)<sup>2</sup> carries this recommendation on its title page.

"Think like a wise man, but communicate in the language of the people."

In conclusion, and also to avoid certain errors to which we might be liable, let us endeavour to create a universal scientific language which will enable us to do certainly more and better than was hitherto possible with the help of the classification at our command.

<sup>1</sup> "Outline of a general theory of index terminology and indexing methods." By Frederick Jonker, AFOSR 1911/AD 272820.

<sup>2</sup> "The making of test thesaurus of engineering and scientific terms," final Report, project LEX, Nov. 1967, AD 661001.

*(Translated in collaboration with Dr. K. M. Beazley.)*

## A Commercial Miscellany

B.S.R. Bulletin does not accept advertisements, is not sponsored by any commercial organisation, does not recommend particular branded goods and has no desire to become a rival to "Which." Nevertheless, since details of three completely different commercial products, having in common only the aim of lightening the task of information retrieval for the working scientist (at the same time as increasing their manufacturer's profits), have recently come to the Editor's notice it seems appropriate to mention them in the present context.

Materials Data Ltd., of Farnham, Surrey, announced in 1967 an information system for thermoplastics. The complex behaviour of these materials, together with the lack of physical properties data based on common test methods, produce an almost insuperable obstacle to a really satisfactory system, particularly when simplicity is a vital criterion. The materials themselves therefore form the basis of the system, which will consist of about 2,000 data sheets of three types: (a) raw materials; (b) semi-finished forms; (c) continuation sheets, one for each manufacturer, giving properties of each grade of material produced. Card selection is based primarily on processing, finishing and chemical properties. Queries and answers regarding the suitability of a material under particular conditions are printed on the relevant sheet, and hence the user need not refer back to another part of the file. Because each card can be considered independently the compilers have decided that symbols need not be consistent throughout the complete system, a procedure that could lead to misunderstanding on the part of a casual user.

With some (admittedly perhaps a lot of) additional thought and effort special edge-notched cards might have been used in such a way as to form the basis of a semi-automatic and much more comprehensive retrieval system.

The system is really a loose-leaf notebook of large size, and can be considered the simplest form of retrieval tool. Many of its limitations arise, however, through limitations in testing methods, and for those plastics technologists who can persuade their companies to pay the annual rental of £105 the system will be a great improvement over the assorted (often incomplete) collection of manufacturers' leaflets which is sometimes all that is readily available.

A step up in complexity of presentation and in ease of retrieval is achieved by the Datamatic Index, announced recently by Pergamon Press. Each Datamatic card contains comprehensive information from the manufacturers' technical literature on a particular product, and is edge-notched to provide a comprehensive cross-referenced index. The first volume, containing 400 pre-notched cards each 11" x 8", contains all manufacturers' published data on rubber polymers and latices available in Britain, together with other information from U.S.A. Updating is achieved by the publisher issuing new cards, which can then be added to the existing file in any order.

Amongst Datamatic Indexes now in preparation are systems to cover thermosetting and thermoplastic polymers.

Each double-sided card is edged with 160 holes arranged in a single row, and direct coding (one hole per concept) is used. A sample card gives the name of the material and its manufacturer, together with a concise description of e.g. moulding and handling methods. A wide range of chemical and physical (including mechanical and electrical) properties are summarised at several temperatures both in tabular and graphical form. The edge holes are divided into several major sections, including, amongst others: type of polymer; special features; processing details; compounding studies; applications; and supplier. Six holes are left free for personal coding.

The sheer size of the cards probably makes handling more cumbersome than advertising photographs suggest, and this system also suffers from the limitation of non-standardisation in presentation of data between different manufacturers. We cannot ask the index to sort out all materials combining, for example, a particular combination of tensile modulus, fatigue life and abrasion resistance, but if what we require is a list of rubber polymers suitable for flooring materials, whose resistance to oil and solvents is known, then Datamatic would appear to make life so much easier. The cost of Datamatic Index Volume 1, Rubber Polymers, is £31 10s. for 400 cards, case and sorting needle, with an additional £10 10s. per year for the twice yearly up-dating service.

Finding out information on a particular material is difficult enough, but coping with information on information is a problem of a different order of magnitude. How can you be certain that you do not miss any important papers in your particular field of study? The answer is that you can't, but the ASCA service provided by the Institute for Scientific Information (the producers of *Science Citation Index*) is an attempt to see that you miss as little as possible.

ASCA utilises the computer tapes produced for *Science Citation Index* for extracting current items related to the particular interests of the individual subscriber. These tapes contain titles, authors' names and addresses, and all references cited for all papers in just under 2,000 periodicals. There is considerable variation in coverage of the literature from field to field—as must be expected from an input for all the sciences only 20% that of "Chemical Abstracts"—and one is totally dependent on titles adequately indicating the complete coverage of a paper. ASCA is therefore subject to the advantages and disadvantages of its parent, with the addition that sensible questions must be asked before a sensible selection of answers can be produced.

A subscriber submits a "profile" of the type of search that is required. This may have several forms: word questions, in which each title is retrieved which contains a certain word or combination of words; cited reference questions, which indicate those papers citing

a particular earlier work; all papers citing the work of a particular author; or all papers published by a given author or organisation. In addition certain negative logical combinations may be used to exclude unwanted material.

All titles containing the selected keyword combinations will be sent to the subscriber, since there is no prior subject classification of entries. Thus the precision (signal to noise ratio) of the system is not high—although it is very simple to determine which might be the relevant papers from a list of about 20 titles.

Each week the subscriber receives the results of the latest search, providing for each "hit" the authors, title, journal, volume, issue and page, together with a reminder of the questions that led to the notification. The cost of the service is typically about £40 per year, and whether this is good value depends on an individual's particular circumstances. Does the tape input cover your field satisfactorily, and are most titles in that field sufficiently informative? Do you receive regular company-produced abstracts bulletins, or are you dependent on personal searches in a library with very little basic material? Above all, do you have the time to spend searching all that might be of interest? Time; that word inseparable from all discussions on information retrieval. Less time needed for literature searching to give more time for new work—combined with a greater certainty that the work really is new and not a repetition of un-retrieved information. Despite their obvious limitations the three schemes discussed above all in their various ways contribute towards the required objective.

## GUIDES FOR THE PREPARATION OF SCIENTIFIC PAPERS AND ABSTRACTS

The Guides for the Preparation of Scientific Papers for Publication and for the Preparation of Authors' Abstracts which are printed below are based on earlier guides which had been circulated widely by Unesco. They were approved at a conference held recently under the joint sponsorship of Unesco and the Publications Commission of the International Union of Pure and Applied Physics which was attended by Editors of Primary Journals in Physics from many countries. It was agreed at the conference that the Guides should be made widely known and that where possible editors of journals would reproduce them in their own countries.

### GUIDE FOR THE PREPARATION OF SCIENTIFIC PAPERS FOR PUBLICATION

#### Preamble

The essential aim of a scientific paper is to communicate ideas and information clearly, concisely and honestly. Authors and editors must keep this aim paramount. Most of the specific recommendations in this guide are directed towards this goal.

A code of good practice among scientific journals is needed to facilitate the exchange of information between scientists in all countries, especially because of the increase in the volume of scientific publications.

#### General recommendations

1. Every manuscript submitted for publication in a scientific or technical journal should be accompanied by an informative abstract.

This abstract should be drawn up in accordance with the recommendations of the *Guide for the Preparation of Authors' Abstracts for Publication*.

2. The title should be concise but as informative as possible.

Titles of articles should be sufficiently descriptive and informative concerning their contents to be of practical use in title lists, in indexing and in coding for information storage and retrieval; abbreviations and specialist jargon should be avoided.

3. A manuscript submitted for publication will normally be in one of the following three categories:

(a) Original scientific paper, describing new research, techniques or apparatus.

A text is regarded as belonging to the category "original scientific paper" when it constitutes a significant extension of knowledge or understanding and it is written in such a way that a qualified research worker is able, on the basis of

the information given, (i) to reproduce the experiment and secure the results described with equal accuracy or within the limits of experimental error specified by the author, or (ii) to repeat the authors' observations, calculations or theoretical derivations and judge his findings.

(b) Provisional communication or preliminary note.

A text is regarded as a "provisional communication or preliminary note" when it contains one or more novel items of scientific information, but is insufficiently detailed to allow readers to check the said information in the ways described above. Another type of short note, generally in letter form, gives brief comments on work already published.

(c) Subject review article.

A subject review article is a survey of one particular subject, in which information already published is assembled, analysed and discussed. The scope of the article will depend on the journal for which it is intended.

It is the duty of the author of a review article to endeavour to give credit to all published work which has advanced the subject, or which would have advanced it had it not been overlooked.

4. Manuscripts should not be submitted for publication if they have already been published or accepted for publication elsewhere. They should not be submitted for simultaneous consideration by more than one journal.

5. The introduction should explain the aim of the paper.

If a historical or critical review of existing knowledge is included in a research paper, it should be confined to the immediate subject of the paper. More comprehensive reviews should be published separately as such.

6. Authors should make clear what in the paper represents their contribution and what represents the work of others. They should be very careful to specify the limitations of their work: the source of error and probable errors in their data, and the range of validity of their conclusions. They should not make over-optimistic claims for the precision of the work, the generality of their conclusions, or the applicability of their results. They should avoid *ad hominem* arguments in criticising other related work. Criticisms should be directed only at the scientific aspects of the related work.

7. Most journals have their own *Instructions to Authors* giving particular details of presentation to be followed. These instructions should be followed carefully. They normally cover general layout, form of bibliographical reference, symbols and abbreviations, etc.

Study of the format and style of similar papers on related subjects in the journal to which the paper is to be submitted may be a valuable guide to proper form and content.

Bibliographical references should be complete. They should be presented in the style required by the journal, e.g. initials of authors, page number, publishers and town for book references, etc. They should specify all authors (not just the first one *et al.*) unless the number of authors is excessive.

8. Material should be presented as concisely as possible in simple straightforward language, avoiding words not in common use. Words appropriate to any particular subject should only be used if they have been fully accepted by other workers in the field. If it is essential to introduce new words these should be carefully defined.

Etymologically unsound nomenclature, ambiguities, technical jargon and colloquial language should be avoided.

9. Authors should try to make their papers reasonably self-contained by including adequate explanation and identification of symbols used.

Authors should not use excessive non-standard abbreviations, especially of infrequently used words, since this makes papers very hard to read. Authors should check the final draft of their manuscript very carefully to eliminate typographical errors, omission of symbols and grammatical errors.

Careless preparation of manuscripts wastes time of editors, reviewers and (if published) readers, and tends to create bias against the scientific validity of the work being reported.

Authors who write papers in a language in which they are not completely fluent should make every effort, by consultation with colleagues who are fluent in the language, to present the work in acceptable form. It is unfair to expect either editors or referees to undertake the task of rewriting the paper to correct inadequacies of language.

10. Should industrial or national security considerations lead to significant restriction of the amount of scientific information which can be published in the article, the text should be presented as belonging to the category 3(b) a preliminary communication, and not to 3(a) an original scientific paper. Any restriction should be clearly indicated in the paper.

11. Explicit bibliographical reference should be made to previous work published on the same subject where this is necessary to show how the new material advances knowledge of the subject. Reference should in general only be made to fully published material or to material which has been accepted for publication.

References should be brought up to date to the time of submission of the manuscript. This is especially important in the case of papers submitted for publication a long time after completion of the work being reported.

Work discussed in the text may be referred to by naming all

the authors, by specifying the first author *et al.* or by referring to the institutional group if the work of a group is involved. Work should not be identified by specifying the country of origin.

12. Reference to private communications and documents with limited circulation (i.e. not available to the general scientific public) should be avoided whenever possible. Documents of a "classified" nature (company or government restriction) should never be referred to.

There is no question of prohibiting allusions to conversations or private communication, but it is certainly not justified to make a statement or advance a theory based solely on a mere conversation which has not been checked.

### **Recommendations to editors**

13. Editors are asked to ensure that the instructions for authors compiled for journals under their control do not contravene the basic principles set out in this document.

14. In accepting material for publication the editor, besides satisfying himself of the general good standard and suitability of the paper for publication, should also pay particular attention to the abstract to ensure that it is sufficiently informative, and to the title to ensure that it is concise but adequate.

15. In journals carrying articles of more than one category the editor should request the author to state in which category his work should be classified, e.g. 3(a) original scientific paper, 3(b) provisional communication or preliminary note, or 3(c) subject review article.

When the category of a paper is not otherwise made clear in the journal, the editor should ensure that the category is indicated in the abstract.

This is valuable both to readers and to editors of abstracting journals.

16. In order to ensure that authors' abstracts may be freely republished anywhere and at any time, it should be clearly stated within the journal that reproduction of authors' abstracts is authorised.

17. Efforts are being made to evolve a consistent series of symbols and abbreviations and a uniform pattern for bibliographical references. As these are determined, editors are being asked to co-operate by using them in the journals under their control.

## **GUIDE FOR THE PREPARATION OF AUTHORS' ABSTRACTS FOR PUBLICATION**

### **Preamble**

With the increasing amount of publication of scientific material it becomes more and more important that the basic content of all papers,

long or short, published in the scientific journals should be easily available. This can be achieved by ensuring that titles are adequate and by inserting at the beginning of every article an author's abstract of the paper.

The following notes set out the main points to be considered by authors in preparing such titles and abstracts.

### General

1. Titles of articles should be sufficiently descriptive and informative concerning their contents to be of practical use in title lists and in coding for information storage and retrieval.

2. Every article should be preceded by an abstract (in most journals this includes Letters to the Editor).

The purpose of an abstract is threefold:

- (i) to help workers in the subject of the article to decide whether the contents of the paper are such that they desire to read it in full;
- (ii) to give readers for whom the paper is of "fringe" interest as much information as possible, so that it is unnecessary for them to read the whole paper;
- (iii) to expedite the work of the abstracting journals by making it possible for them to reproduce immediately the author's abstract; this will be of great assistance in the general improvement of information services in the scientific field.

In writing abstracts authors should bear in mind that these may be the only parts of the papers that are read.

### Content

3. The abstract should contain a brief but *informative* summary of the contents and conclusions of the paper and should refer to any new information which it contains. The abstract should not contain information or claims not contained in the body of the paper. The abstract should not contain inessential details.

4. The abstract should be self-contained but may refer to the title.

5. It is valuable to indicate the treatment of various aspects of the subject by such words as "brief," "exhaustive," "theoretical," "experimental," etc. The abstract should also mention the category to which the paper belongs (cf. paragraph 3 of *Guides for Preparation of Scientific Papers for Publication*) when it deviates from the standard content of the journal.

6. New information should include observed facts, conclusions of an experiment or argument, and essential points of a new method of treatment or of newly designed apparatus, etc.

When feasible, it is preferable to give specific numerical results rather than merely to say what was measured.

7. Reference should be made to new material (compounds, etc.) and new numerical data, such as physical constants. Attention should be drawn to these even though they may be incidental to the main purpose of the paper.

Otherwise valuable information may be hidden.

8. When an abstract includes experimental results there should also be some indication of the method used. Reference to new methods should include their basic principle, the range of operation and the degree of accuracy of results.

Care should be taken to avoid distortions and misinformation. Statements of conclusions and inferences should be accompanied by an indication of their range of validity.

Comparison with earlier results (when this is of major importance to the paper) should be as specific as possible.

### **Presentation**

9. The abstract should be written to form completely connected sentences and not as a list of headings. Standard terms should be used rather than proprietary names.

In order that the abstract can be taken over unaltered by an abstracting journal, it should not be written in the first person.

10. It should be assumed that the reader has a general knowledge of the subject, and abstracts should be intelligible without reference to the full paper.

The abstract should not contain non-standard abbreviations, symbols or terminology unless there is space in the abstract itself to identify them. It should not make specific references (by number) to a section, equation, figure or table of the paper.

11. Specific references and citations should in general not be included in abstracts.

If such reference is necessary citations to scientific journals should be made in conformity with the standard practice of the journal for which the paper is written. (The Unesco International Conference on Science Abstracting, 1949, recommend the standard proposed by the International Organisation for Standardization (ISO), Technical Committee 46, names of journals being abbreviated as in the *World List of Scientific Periodicals*. Note: a new ISO document is in preparation.)

12. The abstract should be as concise as possible while fulfilling the above requirements. It should in general not exceed 200 to 250 words, and will often be much shorter.

It may then, for example, when printed be cut out and mounted on a 7.5cm x 12.5cm card.

The International Conference on Science Abstracting commended the practice of certain journals in which all the abstracts appearing in a single issue are printed together either inside the cover or with advertisements on the back in such a way that they can be cut out and mounted on index cards for reference without mutilating the pages of the journal itself. For this purpose the abstract should not be more than about 10cm wide.

13. The International Conference on Science Abstracting recommended that abstracts be published in at least one of the more widely used languages, no matter what the original language of the paper, in order to facilitate its international usefulness.

*Any comments on the above text may be sent to the Department of Natural Sciences, Unesco, Place de Fontenoy, Paris VII\*.*

#### **List of ISO recommendations concerning scientific publications**

- ISO/R4 "International code for the abbreviation of titles of periodicals."
- ISO/R8 "Layout of periodicals."
- ISO/R9 "International system for the transliteration of Cyrillic characters."
- ISO/R18 "Short contents list of periodicals or other documents."
- ISO/R30 "Bibliographical strip."
- ISO/R77 "Bibliographical references. Essential elements."
- ISO/R214 "Abstracts and synopses."
- ISO/R215 "Presentation of contributions to periodicals."
- ISO/R233 "International system for the transliteration of Arabic characters."

Project ISO No. 315 "Transliteration of Greek into Latin characters."

Project ISO No. 379 "Transliteration of Hebrew."

The above documents are available either from the Headquarters of the ISO (International Organization for Standardization), Geneva, Switzerland

or from

The National Standards Organization of any country.

(The Editor wishes to acknowledge the assistance of the Institute of Physics and the Physical Society, from whose Bulletin the above text is reprinted.)

Fletcher & Co.,  
63 Osborne Road  
Farnborough.