ADVISORY REPORT 14

ADVISORY REPORT 14

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE, 92 NEUILLY SUR SEINE, FRANCE

Research and Development in the Handling of Scientific and Technical Information in the United Kingdom

by

A. H. Holloway

★

1968

NORTH ATLANTIC TREATY ORGANIZATION

44

AGARD Advisory Report 14

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

RESEARCH AND DEVELOPMENT IN THE HANDLING OF SCIENTIFIC AND TECHNICAL INFORMATION IN THE UNITED KINGDOM

by

A.H.Holloway

Naval Scientific and Technical Information Centre

1968

SUMMARY

This survey contains notes about a number of developments in the field which are taking place in the United Kingdom. The list is known to be incomplete because at the time of writing adequate information was not available about some of the work which was known to be going on, because some of the projects were thought to be of insufficient interest and because some important developments have started since the document was compiled. So far as is known, no more complete list is available, but it is hoped that these notes will be kept up to date and extended and that similar lists will be compiled in other countries, since knowledge of work being done elsewhere is of vital importance to all developments.

RESUME

La présente étude est constituée par un recueil de Notes concernant un certain nombre des évolutions en cours dans ce domaine au Royaume-Uni. Il ne s'agit, on le sait, que d'un répertoire imparfait, et ce, pour les raisons suivants: au moment de la rédaction de ces Notes on ne disposait pas d'informations suffisantes sur certains des travaux que l'on savait être en cours; quelques projets ne semblaient pas présenter suffisamment d'intérêt pour être signalés; certains projets importants n'ont été lancés qu'après la date d'établissement du présent document. Autant qu'on le sache, il n'existe pas de liste plus compréhensive, mais on espère pouvoir tenir à jour et compléter les présentes Notes, ainsi que de voir l'établissement de répertoires analogues dans d'autres pays, car la connaissance des travaux entrepris par ceux-ci est d'une importance particulière pour l'évolution de tous projets nouveaux.

025.5:659.2:681.3.01

CONTENTS

RESEARCH AND DEVELOPMENT IN THE HANDLING OF SCIENTIFIC AND TECHNICAL INFORMATION IN THE UNITED KINGDOM

INTRODUCTION

Most workers in the field are aware of a number of investigations and developments which are taking place or have recently been completed in different parts of the world, but they will be painfully aware that they cannot hope to have heard of all of them that would be of interest, even in their own countries, and there appears to have been no systematic attempt to compile a list of such projects apart from the selective lists prepared by some grant-dispensing bodies.

The notes which follow represent an attempt to compile a list of projects in one country which attracted the attention of the compiler and about which information is available. It is being distributed because such a list, with all its defects, is thought to be of general interest to workers in the field and in other countries and so lead towards the accumulation of a comprehensive collection within the NATO community.

This collection was produced by a combination of personal knowledge, responses to enquiries, information extracted from publications and data about commercial undertakings. It is known to be incomplete, for circumstances enforced a choice being made more or less at random; in some instances expected information was not available in time and in some there was simple ignorance on the part of the compiler.

An appendix contains more detailed information about a scheme which has been prepared, principally by Canadian Industries Limited, for the conversion of structural organic chemical formulae to a connectivity matrix suitable for mechanised searching. Mr E. Hyde, Dr F. W. Matthews and Miss L. H. Thompson have kindly provided a detailed description of this scheme and because of its international flavour and of the fact that it is unsuitable for presentation as a short summary it has been thought appropriate to present their paper in full as an example of the several attacks which are known to have been made on this difficult problem.

THE ASLIB CRANFIELD RESEARCH PROJECTS

This project owes its origin to a discussion at the Annual Conference of the Aslib Aeronautical Group in 1955 at which it was agreed that tests ought to be instituted to compare the efficiencies of different subject indexing systems. A grant was made by the (US) National Science Foundation and a collection of 18,000 documents relevant to high speed aerodynamics and subjects related to it was indexed by the Universal Decimal Classification, a faceted classification system devised for the purpose, the Uniterm system and by an alphabetical subject catalogue; times ranging from 1 to 16 minutes were allowed for indexing each document by the staff of varying degrees of library and technical expertise, and the results were tested by questions based upon specific documents (and usually upon their titles), the assessment being upon the ability of the different indexes to retrieve the documents upon which the questions were based.

The results showed surprisingly little difference between the indexers, the systems or the times and the project workers then analysed the causes of failure to retrieve the documents that were the subjects of the questions. In these early tests there was little sophistication, but expertise has grown, and in addition to the analysis of the results obtained at Cranfield, the project has conducted an examination of the index of metallurgical literature at the Western Reserve University, which is prepared by part-time abstractors with appropriate subject knowledge, using role indications, who assigned an average of about 30 terms to each document, and retrieval was by computer. This was compared with a uniterm type index using facets in which an average of about seven concepts were assigned to each document. A comparison was made between 114 searches made by both methods among 950 documents common to both systems in response to questions set by metallurgists from the documents in the systems.

The comparison shows that the Cranfield searches achieved a 90% success rate against 82% at W.R.U., but the W.R.U. searches also retrieved more than twice as many other documents. These documents were examined for their relevance to the questions and sorted into three grades – those equally relevant with the source document, those of some relevance and the irrelevant ones. An examination of the whose collection gave an indication of how many documents it contained which were relevant to each question, and this leads to two ratios, recall (the ratio of relevant documents retrieved to those contained in the collection) and relevance, later referred to as pertinence, (the ratio of relevant documents retrieved to all those retrieved). When these are plotted together for trials conducted under the same conditions a curve is obtained which indicates that in the Cranfield tests the product of the two ratios is approximately constant. In these tests it was found that whereas the W.R.U. recall ratio was similar to that at Cranfield, the Cranfield relevance ratio was about twice that at W.R.U.

The reasons for this disparity were examined by the project staff but the relation between recall and relevance has attracted a great deal of attention in a number of countries. C.W.Cleverdon, the Director of the project is convinced that the form of the recall/relevance curve is fixed and that attempts to improve either factor by variations in search programmes in the same conditions only result in a movement along the curve with a corresponding deterioration in the other factor, and Cleverdon goes further and maintains that it is impossible to achieve really high relevance at the same time as really high recall, though of course unwise changes in the conditions could result in deterioration in both, although some studies elsewhere have reported recall and relevance well over 80% at the same time.

In furtherance of the enquiry on the relation between recall and relevance some 1400 documents on high-speed aerodynamics were examined in great detail in order to examine the effect of such devices as synonym control, word-form control, hierarchical linkage and coordination on the results obtained to some 350 questions set by the authors of the papers from whose references the documents had been selected. This is indicative of the development in the Cranfield activity from the early attempt at an overall comparison between the performance of indexing systems towards the study of the methodology of such

tests, and this second project there is in fact a consideration of the factors determining the performance of indexing systems in which comparisons were made between 29 index languages, leading to the conclusion that in the conditions of the study the best results were obtained from a group of languages using single terms, intermediate results from a group based on the Engineers Joint Council Thesaurus and the worst from a group based on concepts. It is concluded that the only two factors likely to have much effect on performance are the level of exhaustivity of indexing and the level of specificity of terms in the index language.

Cleverdon concludes that for any operational situation, it appears that there is an optimum level of exhaustivity of indexing and an optimum level of specificity in the search terms, that it is unlikely that the environment of the test is unique, and consequently that the best results are therefore generally to be expected by the coordination of single terms in the natural language of the documents and further that there appear to be strong doubts as to whether the improvement in operational performance obtained with indexers as against using key terms from titles or abstracts is economically justified.

ASLIB RESEARCH DEPARTMENT

During the last five years the Aslib Research Department has conducted a number of enquiries and produced a number of reports which have attracted a good deal of attention and reached important conclusions or drawn attention to major anomalies in the use of the scientific and technical literature. These reports include one on the barriers presented to English-speaking scientists by foreign language literature, and methods of surmounting or lowering them, which received considerable attention in the Press in view of the growing realisation of the importance of making the greatest possible use of the published literature.

An investigation of the literature searching done by research scientists in connection with current projects yielded, for the first time, valid evidence on the incidence of unwitting duplication and avoidable waste in research due to failure to find literature in time. The investigation was carried out in two stages. At the first stage, 800 scientists were asked about the literature searching they had done. At the second stage, seven months later, the same scientists were asked if they had since found information which they wished they had found earlier, and to identify such finds. The response at both stages was excellent, over 80%. One outstanding result was that 22% reported making late discoveries (many of them more than one) which either revealed unintentional duplication of research, would, if previously known, have caused them to plan their whole research differently, did, in practice, cause an alteration in the plan of research or would, if previously known, have saved time, money or research work. A factual report was prepared ("Literature Searching by Research Scientists", Aslib, 7s.6d.) and papers discussing the significance of the findings were written for separate publication.

A pilot study of each "act of library use" in 25 selected technical libraries was conducted on a test day. This was aimed primarily at identifying groups of users with differing patterns of demand, and thus providing a classification which could be used to predict group demands. The results yielded a classification showing, among other things, that nature of employment, e.g., industrial or academic, is the most powerful factor in influencing demand, but that, in addition, the patterns of demand by scientists, by engineers and technologists and by technical administrators, differ in important respects.

Studies in depth of the information needs of five scientists were made by daily tape recordings, supplemented by interviews, and the subjects were very cooperative. Unfortunately, only one of their employers agreed to the publication of the diary, but it is hoped to prepare a general review of the lessons to be learned from them about the sources, nature, use and flow of information and ideas.

A series of tests were made on abstracting journals to find the percentage of the total literature in a number of fields covered by each of several English language abstracting organisations, the amount of duplication, and the numbers of items found through various entry words in the subject indexes.

Three new projects were started during 1966. The first is to identify the rank UK journals carrying original scientific work. This should provide data on which decisions for rationalising the publication of original British science could be based, and which could provide working guide lines for journal selection in libraries. The journals are being ranked by an index number that will be derived from (a) the number of original papers published in each journal during 1964, (b) journal use as indicated by such figures as demand on the National Lending Library, and (c) the number of times that recent papers in each journal were cited during 1965. The last figures are derived from a specially prepared print-out of data from the Science Citation Index. The second is a sequel to the pilot study on the usage and users of technical libraries, published in 1964, and consists of a larger survey in which over 100 libraries are taking part. The study is assessing the use made of technical libraries and information services, and its results should have practical application to the planning and provision of such services. The third is an attempt to assess the use in the UK of the literature of social sciences relative to that of the natural sciences and technologies. The resources that need to be devoted to literature consultation and loan services should be related to the demand for these services. The existence of the N.L.L. makes it possible to estimate the absolute demand in science and technology. The project aims at determining the relative demand for literature in this field and in social science. Demand will be measured by references cited in a 10% sample of 1965 British publications.

Other new projects which have been approved and are in progress or preparation are a systems study of library operations, the development of techniques of thesaurus construction, a survey of the forms and uses of bibliographic records in British libraries (in collaboration with the British National Bibliography), an investigation of the techniques and costs of publishing a bibliography by computer typesetting (which will use the Aslib Index to Theses as a test-bed), an assessment of the use in the UK of the literature of the social sciences relative to that of the natural sciences and technologies (which will use citation as a measure but will use data from the National Lending Library for the natural sciences), a study of the national availability of natural science and technology literature relative to the distribution of potention users, and a study of the factors affecting the rate of diffusion of information in science and technology.

The attention of the Research Department had been increasingly drawn towards mechanisation, and as the staff has increased several members have been recruited with expertise in the applications of computers. A major report on mechanised systems prepared by Dr H. Coblans of the research department was published in 1966. Entitled *The use of mechanised methods in documentation work: a report on problems, achievements and potentialities with special reference to the situation in the United Kingdom*, the report presents the results of intensive studies and assessments of mechanised systems made by the author during the previous eighteen months. Even in a fast-moving field, it should provide for some years the basic knowledge of systems in action, their potentialities and limitations, which anyone contemplating mechanisation will need.

The work of the Department, in addition to the research investigations, includes a number of consultancies which arise more or less directly from the expertise gained by its members. These include a major one for the Institution of Electrical Engineers in assisting its plans for the mechanisation of Physics Abstracts and its associated services

by the use of computer typesetting; there has also been a detailed study of the library system at the Atomic Energy Research Establishment and an analysis of potential index mechanisation.

The work of the Department was originally found by the Office of Scientific and Technical Information, although the overheads were supplied by Aslib, and O.S.T.I. now makes a basic grant and gives contracts for a value equivalent to grants which are made by Aslib members. The cost of the Aslib Cranfield projects has been entirely funded by grants from the (US) National Science Foundation.

CENTRAL ELECTRICITY GENERATING BOARD

Nearly all bulletins, lists and S.D.I. notifications issued by the Information Services Section and the catalogue cards relating to the items in these publications are produced by typing on and reproducing from tape typewriters. Specifications for the tape typewriter machines and the programs for the tape typewriter operations were devised within the Information Services' section; work continues to seek new and improved uses for tape typewriters. Each tape-typewriter operation has been subjected to detailed study and cost analysis: figures have been produced comparing earlier production methods and those using the tape typewriters. Economies have been demonstrated.

A thesaurus of terms, based on the Engineering Joint Council's Thesaurus of Engineering Terms, covering the subjects of interest to the Generating Board has been compiled for use in information storage and retrieval. Manual filing of cards in a 'non feature' card index has been written up as a code of practice for staff use.

At the London headquarters of the C.E.G.B. - preliminary and exploratory work has been done on an integrated programme of library and information systems development. A proposal for using computer techniques for handling periodicals literature in the Board's Central Library is being considered and if this is approved some further thought will be given to the economic justification for mechanising more information work.

INSTITUTION OF CHEMICAL ENGINEERS

The Institution of Chemical Engineers is one of the groups which has been struck by the extent to which money is being wasted in the duplication of research and development and has set up a project to test the practical retrieval value of a coordinate indexing system using optical coincidence cards in comparison with a Uniterm system and the system used by the British Patent Office.

A descriptor list has been compiled comprising some 650 indexing terms and 1150 lead-in terms covering the field of chemical engineering. A collection of 1500 documents has been analysed and indexed and the comparison is being made on the basis of a number of questions

devised by independent collaborators. The results are to be assessed on the bases of relevance, recall and the occurrence of 'false drops'.

It is intended when this exercise has been completed to publish the descriptor list and a manual for its use for the benefit of the chemical industry, and the Institution is acting in an advisory capacity to the Council of Engineering Institutions on matters of information retrieval.

THE CHEMICAL SOCIETY

The Society's research unit in information dissemination and retrieval at the University of Nottingham is to undertake an experiment with the aid of a grant from O.S.T.I. in which five hundred Ph.D. students in their last year who receive grants from the Science Research Council will, at fortnightly intervals, be receiving selections from the literature relevant to their work which have been selected by a computer.

Six liaison officers will interview the students and draw up a profile for each of them based upon the references which they have already found to be useful. These profiles will be fed into the computer to guide its selections and will be adjusted from time to time on the basis of the students' reactions to the material they have been receiving as established at later interviews.

The objects of the experiment are to assess the usefulness of the system and to give the students a good introduction to the use of mechanised information services which should be valuable to them in their later careers. The system, which is to be based upon bibliographical references only, will also probably be modified in the light of the experience gained in order to meet the true needs of the users and to increase its acceptability.

CITY UNIVERSITY

Researches in progress at the City University are:

1. Information retrieval by relational indexing and new methods of general concept organization (classification). Some 1200 documents (abstracts) have been indexed and 180 questions are being processed, with checks by the original enquirer and an independent subject expert, and with various statistical methods being developed by us.

2. Psychological investigation of how people make logical jumps in asking questions, i.e., when they pre-classify or condense their real requirements we wish to be able to follow this by equivalent condensations in the indexed material. This will be carried out with a variety of technical staff and students, and results will be compared with indexed material.

3. A small pilot survey has been carried out on the attitudes of small (electroplating) firms to their needs for scientific and technical information. The work is now being extended on a larger scale, the first investigation being with light engineering firms.

INFORMATION SERVICE IN PHYSICS, ELECTROTECHNOLOGY AND CONTROL (INSPEC)

The Institution of Electrical Engineers, the publishers of Science Abstracts, have a project for setting up a comprehensive information service in the fields of Physics, Electrotechnology and Control, with the support of O.S.T.I., which is to establish a system by which abstracts and bibliographical and indexing data can be committed to a magnetic store from which reference publications and indexes can be produced by the use of a computer and photo-typesetting equipment, and references to information identified by specific criteria (e.g. subject, author, source) can be disseminated or retrieved. The fundamental feature of the proposed system is that all the data referring to each item of the literature required to provide the various services will be selected by a once-for-all intellectual effort and committed to store by a single keyboard operation. All the services, including the printing of the periodicals and indexes, the provision of bibliographies, lists of references, etc., will then be produced by machine operations.

INSPEC produces only secondary publications, but its objects demand a very considerable information research programme which includes an investigation of the Selective Dissemination of Information, the evaluation of index languages, user studies, an investigation of the optimum subject average and of the type of material to be used, the relations with primary journals and with other information services and a comparison of document representations for relevance assessment and acceptability. All these are being studied from the point of view of the mechanisation of the INSPEC operations.

The SDI investigation, which is the continuation of the preliminary study undertaken by the National Electronics Research Council will provide, free of charge, a service to six hundred electronics research workers for eighteen months, during which they will receive weekly notifications of English language periodical articles in their individual fields. The notifications will be based upon interest profiles furnished by the users which will be compared with subject indexing of the articles in the periodicals in terms of a language based upon the Engineers' Joint Council's Thesaurus of Engineering Terms, without the use of voles, links or weighting.

The users will provide feedback by assessing the degree to which the notifications they receive coincide with his interests, from which a precision index will be calculated, and the assessments will also be used to adjust the profiles during the run of the investigation. A recall index will also be obtained by sending each user an occasional printout of the complete week's accessions on which he will mark those which he considers to be relevant.

The project will also be the medium of other investigations, such as the comparative acceptability of different types as layouts of the notifications and a survey of the changes which the project itself makes to the information-gathering habits of the users, by means of a questionnaire issued at the start of the project and repeated a year later. Valuable information will also be obtained on the value, usefulness and acceptability of the service and on the rate of change of the profiles.

It is hoped that only one indexing operation and, essentially, one index language will be required for the complete range of INSPEC services eventually established. Thus the language chosen must be suitable for the printed indexes to the abstracts journals, and searches of the machine file, as well as S.D.I.

Initially it is proposed to evaluate the various possible languages on the basis of their use for retrieval in the machine file. When this evaluation has been completed the relationship of the index languages to the printed index, their use in the S.D.I. system, and their relationship to other aspects of INSPEC will be considered.

In all, six index languages will be investigated. To avoid confusion perhaps it should be explained that "index language" is being used here to include not only added index terms, but also the words of the title or abstract when used for retrieval.

The six languages to be investigated are:

1. Terms in the title of the paper, report, etc.
2. Terms in the abstracts, i.e. similar to (1) but of greater exhaustivity.
3. Science Abstracts subject headings with modifier line. This is the system used at present for the printed indexes to the abstracts journals. The subject heading is a controlled language whereas the words of the modifier line, a modification of the title where required to make it more informative, are uncontrolled.
4. Selected natural-language, single terms, i.e. single terms selected by indexers as indicating the subject content of the document, with complete freedom of choice in synonyms, etc.
5. Descriptors used in the S.D.I. Investigation.
6. A specially-developed, controlled, faceted language, which it is hoped will be available from our US associates.

User studies will be undertaken on reactions to the publications Current Papers in Electrotechnology.

Of the many other aspects which will be studied, including the formats preferred and the sectionalising or amalgamation of the Current Papers and Science Abstracts publications, one of the most interesting investigations will be the present use of Science Abstracts for retrospective searching. It is hoped to obtain the cooperation of a number of librarians and information officers in a variety of organisations to provide a record of the queries for which they sought answers in Science Abstracts. Such a collection of typical queries (preferably in the language of the questioner) will be invaluable for consideration in the development of the printed indexes and in setting up the retrospective searching facilities of the machine file.

ELECTRONIC MATERIALS INFORMATION CENTRE

This Centre, which was started by the Royal Radar Establishment at Malvern in October 1966, sets out to provide for British workers in electronics a service similar to that provided by the Oak Ridge National Laboratory of the US Atomic Energy Commission. It will

make information available under the headings of references in the literature to specified topics or combinations of topics, availability of research specimens and special materials and location of specialised knowledge of, and facilities for, crystal-growing, purification of materials, analysis etc.

The input to the Centre consists partly of copies of relevant articles from the literature and partly of data sheets contributed by interested participants. The same data sheets are used for formulating enquiries. At first the input will be confined to the data sheets, which will be retrieved by specialised card equipment; but as the Centre expands computer-indexing will be included. Over part of the subject field Oak Ridge and Malvern have complementary information and each will refer appropriate questions to the other.

This is one of the ways in which the Royal Radar Establishment is giving a new service to industry, and other approaches are being developed there in an Industrial Applications Unit which provides a channel of contact and offers a consultative service as well as devices which are suitable for commercial development; an example of these is the use of touch wires in the face of a cathode ray tube which is the display unit of a computer. Contacts through these wires serve as a means of interaction between the computer and its user.

ENGINEERING SCIENCES DATA UNIT

The Technical Committees of the Royal Aeronautical Society, the Institution of Mechanical Engineers and the Structures and Materials Panel of AGARD/NATO working with The Engineering Sciences Data Unit have produced a wide range of authoritative data in the form of data sheets, memoranda and handbooks. There are already many hundreds of different Items available. Most of these are data sheets in the long established Aeronautical Series. These are being supplemented by new Items in the Aeronautical Series and by Items in the recently commissioned Mechanical Engineering Series. Many of these Items have a potential application far beyond that originally intended. All of them are on public sale.

It is of great importance that a potential user of this system of data should be able to locate any information pertinent to his work, for there is little point to the provision of these working aids unless they can be readily located and obtained. For this purpose an Index has been produced.

No existing thesaurus or classification system was found to possess a sufficiently realistic series of terms or sufficiently precise sub-division to represent adequately the material existing in the various Engineering Sciences Data Series. Since the Index has been prepared to assist, primarily, the designer or other worker in industry or research, the terminology chosen reflects those headings which, in the experience of the staff of E.S.D.U., are those under which data are most likely to be sought by such a user. In brief, this is intended as an "engineers index" rather than as a "documentalists index" and there is no suggestion that the entries used have any universal application.

INSTANT LIBRARIES

A number of commercial organisations have come into being with the object of supplying the literature which is constantly used by technical officers. Many small firms and some larger ones have no libraries and it is common to find that engineers and designers keep their own collections of data sheets and trade catalogues, often of some antiquity. If the collections are kept up to date the technical officers have to devote a considerable amount of valuable time to this activity and it has been appreciated that there is a market for a service which does just this. For a moderate fee the undertakers supply the material and equipment needed to set up a small library, essentially of trade literature, and visit the subscribers at regular intervals to replace old documents by up-to-date ones and supply new ones. As most of the material is supplied gratis the charge is essentially one for service and indeed as the producers of the documents are interested in having their material efficiently used and are saved the expense of individual distribution they may be prepared to pay the undertakers a fee in addition to that paid by the recipients.

A few details are given of three such undertakings in the field of engineering, but there are others, such as architecture, in which the system is being applied.

Technical Indexes Limited

This firm located at Ascot is offering libraries of firms literature and the principle is to replace the engineers' "sedimentary deposit" of firms catalogues, etc., by a properly indexed and organised library. Firms pay for their literature to be included and T.I. staff visit each library once a month to bring the collection up to date.

The Electronics Engineering Index costs 50 guineas a year and over 300 sets have been installed. The Components/Materials Index is 40 guineas and 250 are in operation.

The E.E. Index is also available on 16 mm microfilm. It is used in cassettes and with a 3 MMs reader printer costs £650 p.a.

The firm is now considering microfiche.

Materials Data Limited

The firm offers data sheets on materials. These are prepared by specialists from all available information and are printed in standard form on 6" x 8¼" sheets. Peek-a-boo cards are provided for retrieval on specification, properties, etc.

The Non-ferrous Metals system is £40 p.a., Iron and Steel £65 p.a. and that on Thermo-Plastics, expected to be issued mid 1967, about £105 p.a.

This material can also be obtained on 16 mm microfilm using the 3 Ms reader printer for viewing and the print out of cards.

Engineering Index

A very interesting scheme at a somewhat lower level is being operated by Engineering Index. Subscribers are provided with a library of trade catalogues which is updated at monthly intervals and with indexes to them which are updated at longer intervals by the organisers; a very moderate fee is charged for this service and there are about 400 subscribers. A development is being planned by which the library would be in the form of microfiches and a reader would be provided as part of the service: the space needed by the microfiches and reader would be considerably less than that occupied by the catalogues, and the material would be more readily accessible. The scheme is welcomed by users since the information is available at one point instead of being scattered and is kept up to date at a cost which is less than that represented by more haphazard methods, and by the firms producing the catalogues since one centralised supply replaces a large number of uncoordinated

requests and there is greater assurance that the catalogues go to recipients who will really be using them.

THE LIBRARY ASSOCIATION

The Library Association has three projects of interest one of which is a survey of the major indexing and abstracting services for library science and documentation by H.A. Whatley, which is published as a separate report of some 78 pages. The author, who is editor of one of the services surveyed (Library Science Abstracts) deals with sixteen services in Czechoslovakia, France, East and West Germany, Hungary, Italy, the Netherlands, Poland, Sweden, the United Kingdom, the United States and the USSR and gives general information about their content, coverage and preparation, bringing out their variation between strict librarianship and information handling and between abstracting and indexing services. There follows an analysis to provide a basis for selection between the services from a range of points of view – subscription rates, coverage by country, language and content of sources and material, method of compilation, style of entries and presentation, classification of the contents and indexing, the timeliness of appearance, the services provided and the use made of them. These sections are of course factual, but there is also a discussion of the quality of the abstracts and an assessment of the services, and the conclusions include some eighteen detailed recommendations including the author's proposed list of subject headings for an abstracting service in the field. The preface is dated March 1965 and it is unfortunate that at least one of the services reviewed is now defunct.

The second development is the production of the British Technology Index, which is complementary to the H.W. Wilson Applied Science and Technology Index with a minimum overlap. About 30% of its content is believed to be covered by no other abstracting service although its coverage is restricted to English language periodicals, and it combines a prompt current awareness service with good indexing by subject headings in a wide field of applied science and engineering. In order to maintain the high standard of currency the present method of production by a combination of Varityper setting, Fotolisting and offset lithography with a good deal of manual work is being reviewed with the idea of using computer processing for sorting operations, the comparison of input headings with standardised file headings, operations of term manipulation, the production of an authority file and the introduction of computer typesetting. The changes are not expected to reduce costs, but to enable more to be done with the present resources and without the struggle which is at present necessary to meet the publication programme.

The third project is a long-term investigation into a new general faceted scheme of classification which is being carried out under the guidance of the Classification Research Group. This is in its early stages and is expected to absorb more than ten man years of work, but a preliminary investigation under NATO funds is nearing completion; in this the theory of integrative levels as a basis for a classification system has been explored and studies made of such areas as geology, mining and sculpture to test how the ideas would work out in practice. A number of problems have been examined such as the distinction between physical and chemical entities, particularly in fields whose content consists of concrete entities. In several fields there are difficulties in establishing a satisfactory sequence of levels and in some of them no acceptable solution is in sight and a concept which appears at one level in one field may well require a markedly different level in another. Although some patterns may be seen in a number of areas their application is by no means universal and the search for universality may be a delusion.

THE UK MEDLARS SERVICE

An information retrieval service using tapes prepared by the US National Library of Medicine has been in operation in the UK since May 1966, and about 1,000 individual literature searches were performed during the first year. Experiments are also in hand on "current awareness" searches, and an economic assessment of the system is being made.

During the experimental period, the service is free, but users are requested to provide "feedback" about the usefulness of each reference found.

Orientation courses for librarians and research workers have been held, to improve the critical step of communication of the users' requirements to the operators of the system.

MEDLARS must be seen in its context of conventional guides to the literature, and in particular "Index Medicus", which contains exactly the same references. MEDLARS searches are very specific sub-sets of the references in "Index Medicus"; other, broader sub-sets are published (e.g. "International Nursing Index", "Index of Rheumatology") and some more specialised sub-sets in narrow fields are available for distribution by MEDLARS centres.

The UK MEDLARS service is organised by the National Lending Library for Science and Technology, and the computer processing is carried out at the Computing Laboratory of the University of Newcastle upon Tyne. Enquiries should be addressed to the UK MEDLARS Service, National Lending Library for Science and Technology, Boston Spa, Yorkshire.

NATIONAL INSTITUTE OF MEDICAL RESEARCH

Project FAIR

This is a cooperative project for Fast Access Information Retrieval in the field of Biomedical Engineering by optical coincidence feature cards, and the published objects are to produce a formula for creating efficient indexes to feature card information retrieval systems, to explore the possibilities of members of a learned society helping in the setting-up of an information retrieval system for their use and to test the practicability of providing a whole library on the individual's desk. A collection of documents has been formed by gifts from the collaborators who are all active workers in the subject field and batches are sent out to the collaborators for analysis: they are asked to assign up to 15 descriptors to each document, each descriptor being a word or phrase to represent a single concept and only one descriptor is to be assigned to any concept; the collaborators have a free choice of the words they use apart from a few general rules, but the descriptors are to be arranged in order of importance.

Each document is analysed by two of the collaborators and the selected descriptors are then recorded on 80 column tabulator cards, which will be fed in groups to a computer, which is programmed to select variable numbers of descriptors according to several criteria, and thus to enable a language for information retrieval to be built up which will in due course be formed into a thesaurus which the collaborators will be asked to use instead of having a free choice of descriptors.

The ultimate intention is to establish a number of satellite libraries or information centres, each of which would have a copy of the thesaurus, a set of optical coincidence

cards or a printed index, a set of microfiche microcopies of the documents and a reader-printer, so that any enquiry could be formulated in the terms of the thesaurus, the relevant documents selected, the microfiches read and any necessary copying done on the spot in the course of a few minutes.

The literature collection comprises over 2,100 reprints from a total of 225 periodicals and 180 collaborators have received up to five batches of reprints and returned them with their subject indexing. Over 850 reprints have been analysed twice, and from this information three different information retrieval languages of lists of descriptors used in indexing 500 papers each have been processed by the computer in order to assess the degree of coincidence between them in terms of the frequency of use of the descriptors. In the most often used hundred descriptors there was about 70% coincidence between any two of the three languages, and this decreased to about 55% coincidence between the top 500 descriptors. A language is to be made by combining two of the three, and this is to be issued to the collaborators for their use as well as being used to re-index the reprints which have already been circulated.

NORTH-WESTERN POLYTECHNIC SCHOOL OF LIBRARIANSHIP

Some of the investigations to be undertaken here are of interest.

An investigation into the problem of indexing and classification in the building and construction industries. This will be made jointly with the Brixton School of Building and in close cooperation with RIBA. It will most likely include the production of a classification and thesaurus for the industry. This project will engage two senior researchers at least, for two or three years.

An analysis of the bibliographical structure of a technological field (Computer technology and Paper technology are the guinea-pigs) - essentially a statistical analysis of who produces what kind of information in the field, with every useful parameter such as language, place of publication, publisher, 'level', etc. analysed.

A study of the New Anglo-American Cataloguing code has just been completed under funds provided jointly by the Library Association and the British National Bibliography. This study will be of particular importance to national bibliographies.

UK COLLABORATION IN PRODUCTION OF NUCLEAR SCIENCE ABSTRACTS

Nuclear Science Abstracts is a publication compiled by the United States Atomic Energy Commission's Division of Technical Information, Oak Ridge, Tennessee. It has been in existence since 1947 and has established itself as a comprehensive abstracting and indexing service for the international literature in nuclear science and technology. At present it publishes about 50,000 abstracts per year.

The U.S.A.E.C. has felt for some time past that other countries should play a part in the compilation of N.S.A. For this reason agreements have been set-up for collaboration by certain other countries - notably Canada, the Scandinavian countries, and the UK. The U.S.A.E.C. are looking to a time when N.S.A. will become a truly international project.

Mr J.Terry, of Harwell, visited Oak Ridge in the Autumn of 1966, to make detailed arrangements for UK participation.

UK participation commenced in January 1967, and involves (a) selection of matter appropriate to N.S.A. from the many UK journals, reports, etc., and (b) abstracting of this matter (or editing of author abstracts, where provided). Indexing will be provided later. Computer techniques are being kept very much in mind.

Selection and abstracting are being undertaken by several U.K.A.E.A. libraries and also by Science Abstracts, but the major part of this work, including the coordination of the UK contribution, is carried out at Harwell by Mr R.W.Clarke, who has been at Harwell for the past 18 years.

The U.K.A.E.A. and the Office of Scientific and Technical Information are collaborating in meeting the cost of the UK effort, which covers not only publication by the U.K.A.E.A. and S.R.C. staffs, but also the work of universities, firms, etc.

The UK contribution at present amounts to an average of about 75 abstracts per week, or about 7% of N.S.A.'s total abstracts.

A successful Conference, attended by Mr Terry and Mr Clarke, was recently held in Sweden with the Swedish Documentation specialists, together with representatives from D.T.I.E., the I.A.E.A., and Euratom, agreement being reached on methods of operation, and recommendations framed for future guidance.

DEPARTMENT OF EDUCATION AND SCIENCE

Office for Scientific and Technical Information

This Office, which is a substantially independent unit within the Department has the mission of being the vehicle for British Government support for the development of information services in science. It acts by making financial grants rather than by carrying out investigations and developments itself, and its policy is determined by an Advisory Committee under the chairmanship of Dr F.S.Dainton, the Vice-Chancellor of the University of Nottingham; in 1966 its expenditure amounted to £300,000.

Most of the activity of O.S.T.I. consists of letting contracts to outside bodies. Most often this means universities, but research associations, learned societies, and industry are also eligible for assistance. As part of the plan to collaborate with information services abroad, O.S.T.I. is supporting the trial in Britain of the MEDLARS information retrieval service devised by the United States National Library of Medicine, and the corresponding system in chemistry sponsored by 'Chemical Abstracts'. But the office has also helped with grants to the University of Sheffield for research on the automatic detection of structural similarities among chemical structures, and what is called a National Reprographic Centre for Documentation at the Hatfield College of Technology and many others. The terms of reference for O.S.T.I. have allowed it to provide a grant for the support of 'Physics Abstracts', at present published by the Institution of Electrical Engineers. The intention is that the grant should enable the abstracting journal to investigate new techniques of compilation and dissemination. 'Physics Abstracts' is also supported on a continuing basis by the United States Government by means of a grant through the American Institute of Physics.

The interests of O.S.T.I. include the support of specialised information centres, such as those on electronic materials at the Royal Radar Establishment, on biodeterioration at the University of Aston, on intestinal absorption at the University of Sheffield, on high-temperature processes at the University of Leeds and on mass spectrometry at the Atomic Weapons Research Establishment. Information activities in the social sciences have been underdeveloped, and are now receiving support from O.S.T.I.; some preliminary studies have been completed and deeper studies in areas of importance are now being encouraged with the cooperation of the Social Science Research Council; the National Lending Library is also developing its collection of the literature of the subject.

Work on the automation of cataloguing and other library procedures and information activities is being supported at a number of centres; in addition to Physics Abstracts, there is work on library cataloguing at the University of Newcastle and mechanised selective dissemination of information in different fields is being supported under the Atomic Energy Authority and the Institution of Electrical Engineers.

Some of the projects which are being supported but which are not otherwise mentioned are to be found in the following l'st, and others of them will be described under the appropriate organisations.

LIST OF O.S.T.I. SUPPORTED PROJECTS (AS AT 28TH FEBRUARY 1967)

Institute and Investigator	Short description of investigation	Commencing date and duration
City University (J.E.L. Farradane)	The consolidation and testing of a new theory of indexing which used certain well-defined relationships between index terms.	1.10.63 (4 yrs 6 mths)
Newcastle University (Prof. E.S. Page)	MEDLARS - Medical literature analysis and retrieval system - an experimental service provided for medical research workers based on a computer file of indexed medical literature provided by the National Library of Medicine, USA.	4. 3.65 (3 years)
Institution of Electrical Engineers (Dr G.F. Gainsborough)	Development of an information service for physicists.	1. 5.65 (1 yr 9 mths)
University of Aston (Dr M.O.W. Eggins)	Biodeterioration specialised information centre.	1. 6.65 (3 years)
Cambridge University (Dr O. Kennard)	Provision of uniform X-ray single crystal structural data; this forms part of an international publication on X-ray data.	1.10.65 (3 years)
City University (J.E.L. Farradane)	A psychological study of the ways in which people sub-consciously compress information in their minds.	1. 4.66 (3 years)
Oxford University (C.F.H. Tapper)	Testing the use of mechanical searching of the complete text of legal reports to facilitate the retrieval of relevant reports.	1. 6.66 (2 years)
Chemical Society (located at Nottingham University) (J.R. Ruck Keene)	Research unit in chemical information dissemination and retrieval.	1. 8.66 (1 year)
Sheffield University (W.L. Saunders and Prof. D.L. Smyth)	Intestinal absorption - specialised information centre.	1. 9.66 (3 yrs 3 mths)
Library Association (E.J. Coates)	Feasibility study for computerisation of the British Technology Index.	1. 9.66 (1 yr 7 mths)
British Standards Institution (G. Weston)	Support of activities in documentation.	1. 9.66 (3 years)

Institute and Investigator	Short description of investigation	Commencing date and duration
Lancaster University (A.G.Mackenzie)	Systems analysis of the university library.	1. 9.66 (1 year)
Newcastle University (A.B.Littlewood)	Provision of data to enable many gas chromatography data results to be translated into absolute values.	15. 9.66 (3 years)
Sedgwick Museum, Cambridge (A.G.Brighton)	Feasibility study of computer-orientated systems for geological data handling.	1.10.66 (1 year)
Leeds University (Dr A.Williams)	High temperature processes - specialised information centre.	1.10.66 (3 years)
Durham University (Dr J.Hawgood)	Cost benefit study of the value of university libraries.	1.10.66 (1 year)
Sheffield University (W.L.Saunders)	Automatic detection of structural similarities between chemical structures.	1.11.66 (3 years)
Leeds University (Dr D.L.Baulch)	Kinetic data evaluation for high temperature processes.	(3 years)
London School of Economics and Political Science (D.A.Clarke)	Investigation of literature and information services available for the social sciences.	(1 year)
Newcastle University (M.B.Line)	Investigation into information requirements of the social sciences.	(3 years)
Sheffield University (W.L.Saunders)	Subject indexes and automatic retrieval of information.	1. 6.65 (3 years)
Edinburgh University (J.N.Wolfe)	Economics of libraries.	1. 1.67 (1 year)
Hatfield College of Technology (G.H.Wright)	National reprographic centre for documentation.	1.12.66 (3 years)

NATIONAL PHYSICAL LABORATORY

Information Processing and Language Processing Group

The new emphasis in this group is to be on the development of computer-based fact retrieval techniques and on the computer processing of natural language items within fact retrieval systems. Document-retrieval studies are continuing, but are in a final evaluative phase. The project on Russian-English machine translation has been terminated. A further new project is on computer transcription of a shorthand machine's output.

Fact retrieval. Studies in this area will focus on the organisation of the computer storage of specific facts in an information system (as distinct from document descriptions; see below), so as to allow efficient and flexible retrieval, from large data bases, of both explicit and implicit items. Further to develop efficient means of handling items expressed in natural language within such systems. Operational interaction of user and machine is also of interest.

Initial studies will relate to the requirements of an information system for crime-detection, sponsored by the Home Office.

Document retrieval. The work here is just beginning the final evaluative phase wherein the effectiveness of computer derived descriptor sets for indexing documents and of various computer retrieval strategies using these sets on the collection of 11,500 documents from which they were derived is to be subjectively measured by a panel of evaluators.

Palantype shorthand transcription. A Palantype shorthand machine, such as is currently widely used in recording verbal proceedings, has been modified to provide direct computer input. A Palantype code-English computer dictionary is being compiled and will be used by a transcription programme to convert Palantype input into standard English on an output typewriter. This system would have clear application in fixed situations such as the House of Commons and the Law Courts, and later stages may prove its suitability as a fast typing service.

Machine translation. This project has been terminated, but the results of an evaluation experiment on the quality of translations produced by the (now-deceased) ACE computer are available, as are many examples of whole-article translations. A comprehensive report on the project will shortly be issued.

NATIONAL REPROGRAPHIC CENTRE FOR DOCUMENTATION

A National Reprographic Centre for Documentation has been set up at the Hatfield College of Technology by an O.S.T.I. grant with Mr G.H.Wright, the County Technical Librarian, as Director. The Centre is particularly concerned with photographic methods of reducing original documents to microforms (such as microfilm and microfiche) for storage, handling, retrieval and enlargement and will act as an important source of unbiased and informed advice on the application of reprographic techniques.

The functions of the Centre are:-

1. To act as a national clearing house for information on microrecording and associated reprographic techniques. All relevant published information will be collected, evaluated and abstracted and will be disseminated to interested users. The service

will go out in microform and will itself be used to assess the technical and economic parameters of, and users' response to, various microform systems.

2. To test and evaluate equipment on the British market and to coordinate this evaluation with similar work being undertaken by sponsored organisations in the USA and the Netherlands.

3. To examine and evaluate users' needs and encourage research and development to meet them.

4. To identify specific areas where further research and development are necessary and to sponsor and, where appropriate, carry out such work.

The activities of the Centre are carried out by a small team with appropriate library, design and photographic experience and is guided by an advisory committee which includes representatives of Aslib, the Library Association, the Institute of Reprographic Technology, the Microfilm Association of Great Britain and O.S.T.I.

The services available on subscription include a periodical bulletin, courses and symposia, evaluation reports from the Centre and other sources and an information service published on microfiche as well as an enquiry service.

The first evaluation report is of a microfilm reader. It includes a specification, details of the construction, operating information for 35 mm and 16 mm microfilm, microfiche and aperture cards, information on maintenance and a detailed evaluation including legibility tests to ISO Recommendation No. 648. There are recommendations for improvements to later models.

5. To assess specific areas where further research and development is necessary and to sponsor this in appropriate specialist organisations.

ROAD RESEARCH LABORATORY

A partially computer-based information storage and retrieval system is being established to provide rapid retrieval of information and a current awareness service.

Input to the system includes abstracts of research reports and published articles, selected by members of the International Road Research Documentation (IRRD) scheme, and summaries of current road research projects in the United Kingdom and many foreign countries.

The IRRD scheme was established in 1965, under the auspices of the Organisation for Economic Cooperation and Development, for the exchange of information on road research. The scheme is based on selecting and disseminating information in the form of abstracts. The Laboratory's Technical Information and Library is one of the three Coordinating Centres for this work. The other two Centres are Laboratoire Central des Ponts et Chaussees in Paris and the Forschungsgesellschaft für das Strassenwesen in Cologne, supported by the Bundesanstalt für das Strassenwesen, Cologne. The other member countries are Austria, Belgium, Canada, Denmark, Netherlands, Norway, Portugal, Spain and Sweden. Each IRRD member is responsible for analysing and indexing its own literature; material from non-member countries is shared. Information is therefore analysed and indexed once only.

Abstracts are prepared in one of the three official languages, French, English, and German, and are indexed using keywords selected from a trilingual thesaurus of terms in the field of road and road traffic research and related subjects. The abstracts are prepared on standardised forms, abstracts of current research projects on Project Sheets and abstracts of published articles on Information Sheets, and are sent to the Coordinating Centre dealing with the language in which the abstract has been written. The Centre allocates an IRRD number, processes the information and distributes the sheets to all member countries.

The trilingual thesaurus from which the index terms are selected contains some 2,500 coded terms. Some 53 basic ideas or subject areas have been adopted to help the indexer select appropriate keywords. These areas embrace the whole field covered by the thesaurus and each constitutes the central point of a diagram, with arrows linking the keywords corresponding to related ideas. External links connect associated keywords appearing on different diagrams. The diagrams have coordinates to permit easy codification.

The input to the RRL Technical Information Service is being stored on the Laboratory's Pegasus II computer. Each IRRD sheet number followed by the code numbers of the keywords assigned to the document are recorded; in the case of published articles, some bibliographic details are also included. Computer programmes have been written and tested and the material exchanged in the IRRD scheme since its inception in 1965 is being put into the computer.

SCIENTIFIC DOCUMENTATION CENTRE LTD - Dr P.S.Davison

The Scientific Documentation Centre at Dunfermline is chiefly known for its activities relating to the operation of Current Awareness/S.D.I. Services. The Current Awareness Services cover a wide range of subjects in the scientific, technical and medical fields including such items as adhesives, cybernetics, spectroscopy, deuterium, entomology, luminescence, management, pattern recognition, computers, tantalum, tissue culture and water desalination: a retrospective searching facility is available for many of them, and a searching service is also available on published indexes for most subjects.

The Scientific Documentation Centre is also known for its services on spectra, which are based on its very large collection of ultraviolet, visible, infrared, microwave, nuclear magnetic resonance, electron spin resonance, Raman, optical rotary dispersion and mass spectra. The published collections of spectra are supplemented by numerous spectra from laboratories and a loan and copy service is provided to subscribers, who can be supplied with spectra for individual substances or for groups of substances with specific relationships. Extensive indexes are maintained and the service copes both with current awareness on an S.D.I. basis and with retrospective searches.

Current Research projects are mostly those directly concerned with the operation of current awareness and spectra services. A very simple and apparently effective and cheap means of dissemination of scientific information has been developed. This gives prompt notification of research publications and operates on an S.D.I. basis. The methods used are manual. Based on this, the Centre is at the moment planning a user requirement survey to obtain further validated information on the types of information scientists wish, and the sources they are at present using. To obtain information about the needs for S.D.I. Services, the Centre has studied, and is studying, the distribution of scientific information on a series of topics in the literature. The national sources of research publications

have also been studied with a view to assessing the contribution made by different countries. The Centre is involved at present in two projects to compare its own current awareness services with other comparable publications. It is active in the preparation of research bibliographies on a number of topics of contemporary interest, and as a matter of routine, records sources of its information for these. One of these bibliographies is on costs in information retrieval and the Centre has made some costings itself. The Centre carries out a substantial number of searches for spectra of various kinds and systematically records the sources in which these have ultimately been found. As a major part of these spectra are ultimately found in unpublished laboratory sources, it is hoped this will give a means of assessing the number of different spectra actually available in laboratories throughout the world. A research project which will soon be published provides a simple device for assisting in the accurate measurement of literature spectra whose scales and base grids are often inadequate. The Centre has special problems relating to the retrieval of the very large amount of spectrographic data which it holds. Indexes of a novel kind are planned for this. A number of practical experimental projects have been carried out to obtain information needed for the immediate operation of services.

The field of operation is being expanded and a number of investigations are planned or being conducted on aspects of the indexing, storage and retrieval of scientific and technical information, including the costing of indexing, reproduction and dissemination, the comparison of some existing services and techniques of photoelectric data handling.

SHEFFIELD UNIVERSITY POSTGRADUATE SCHOOL OF LIBRARIANSHIP

This School is at present operating three research projects with financial support from the Office of Scientific and Technical Information. The students at the School also conduct special studies as part of the requirements for their Diplomas, both in the field of pure librarianship and of scientific information studies.

The first project concerns the automatic generation of subject indexes. A technique has been devised, and is at present being programmed, which enables a set of title-like phrases which describe the contents of documents to be manipulated into the form of an articulated subject index, i.e., one that closely resembles the index to *Chemical Abstracts*. The title-like phrases, or notations of content, will be derived by indexers, and will comprise nouns or noun phrases which can act as subject headings in the index. The computer will then transform these into potential index entries by rearrangement of the constituent parts of the phrases. These will then be edited and those entries selected which lead to the most highly organised form of display. The method is based on a study of structure in the entries in *Chemical Abstracts* indexes; this showed that a simple method for turning the entries into title-like phrases could be devised. From this, the logic underlying the transformation from title-like phrase into index entry was deduced. The advantages of the method are that indexers' efficiency is increased, and that an easily used index, with, on average, more access points, can be produced with a minimum of further human intervention. Studies on retrieval using subject index data are also in progress, and have given greater insight into certain steps in the indexing process.

The second project is concerned with the automatic detection of structural similarities among chemical structures and seeks to extend the range of manipulations possible on chemical structures stored in computers' memories. Although techniques for searching files of chemical structures for identify or partial identity are already well-established, there is as yet no established means of finding similarities in terms of the maximum overlap of

a pair of structures. A method has been devised, based on the generation of fragments of each structure, starting with the individual atoms of each, and by concatenation, fragments of increasing size. Each fragment generated comprises full information on the constituent atoms and the bonds which connect them, and at each step in the process the fragments formed from one structure are compared with those from the other; non-common items are discarded, and growth continues only from those fragments which are common to both. The procedure is continued until the structural 'highest common factor' has been determined. This system is used in a computer programme which is being written for the automatic determination of similarities among pairs of acyclic chemical structures, and it can be used for a number of related applications, such as the analysis of the structural changes which take place in the course of a chemical reaction and thus it would permit information on chemical reactions to be analysed, stored and mechanically searched from a wider variety of viewpoints than is at present possible.

The third project is on science information education in the United Kingdom. The study will first, ascertain the types of work being carried out in libraries and information services and the background of persons engaged in this work, and from this build up a picture of the various grades of staff, and types of knowledge required, to operate present services, second, determine what additional staff would be needed to improve and expand these services in the national interest, third, survey existing education and training facilities and fourth, determine the requirements of education and training for all levels in terms of content, duration, standard and general character of various courses.

Interested persons and institutions are invited to make observations and recommendations, and in addition it is intended to collect other evidence by visits to as many organisations, institutions and persons, as is feasible in the time available. It is also intended to make full use of advanced experience abroad.

"SHELL" RESEARCH LIMITED

Woodstock Agricultural Research Centre

Work here comprises machine methods for handling (a) a large and growing file (> 50,000) of compounds (mostly organic) and test data thereon and (b) literature - both Company and open literature. Some of the more important aspects are outlined below.

(a) *Compound and test data files*

Compounds are coded (IUPAC* ciphering) and the coded structures held on magnetic tape for sub-structure searching by computer.

In another approach, the structures are entered into the computer via a typewriter, designed by Shell Development Company, Emeryville Research Centre, California, USA. This machine enables a two-dimensional structure to be typed, component by component, with the simultaneous production of a tape. The tapes are used to (a) mechanically reproduce the structures via the typewriter or (b) as input to the computer.

Methods have also been worked out for whole compound matching using sorted cipher files, for selecting IUPAC fragments for entry onto feature cards and for file sub-division. The practicability of generating feature card systems via the computer for many search purposes

* International Union of Pure and Applied Chemistry.

has been demonstrated and methods of mechanising the punching of the feature cards are being studied.

Handling laboratory and glasshouse test data is relatively simple and is done by fixed field punching followed by simple mechanical card sorting or computer handling according to the nature of the search. Work is also in hand for dealing with the much more variable data arising from field work.

(b) *Literature handling*

This comprises KWIC indexing and similar work. Experiments have been started on Selective Dissemination of Information (SDI), based on Company records and on "Chemical Titles" and "Chemical-Biological Activities" (CBAC) tapes obtained from Chemical Abstracts Services, with a view to establishing a collection of personalised (i.e. user-orientated) reference files held on magnetic tape.

Woodstock Publications

1. Computer-based chemical information system.
H.F.Dammers, New Scientist, 11.8.66.

2. Mechanisation of a feature card system.
H.F.Dammers. Paper presented at Aslib Symposium on feature card indexing. (8.4.64, London).

3. Computer handling of literature information and research data in an industrial research establishment.
H.F.Dammers. Paper presented at the 36th International Congress on Industrial Chemistry (11th to 16th September 1966, Brussels).

"SHELL" RESEARCH LIMITED

Thornton Research Centre

A machine-sorted punched card index is being developed by Technical Information Division at the Thornton Research Centre of "Shell" Research Limited with the active cooperation of the Combustion Research Division to cover the interests of this fundamental research group which uses the most advanced techniques including electron spin resonance, various forms of spectroscopy, shock tubes and molecular beams to investigate combustion and ignition processes, flame noise, ionization reactions and the behaviour of free radicals. The index which was started twelve months ago now contains some 1500 cards and is based on codes of the reactants, reaction products, type of reaction, experimental techniques used, measured values, experimental conditions, calculated or theoretical values, mechanisms, intermolecular properties and quantum mechanics.

Preliminary tests have shown that the system is working well. Consideration will be given to a full description when more extensive tests have been made.

STANDARD TELECOMMUNICATIONS LABORATORIES LIMITED

Automated Information Dissemination System

A large-scale experiment has been carried out in order to evaluate the technical feasibility of an information dissemination system using a computer for both selective dissemination and retrospective retrieval.

The experiment was in the field of electronics and the material used was about 10,000 abstracts in Science Abstracts B over nine months, supplemented by a number of technical reports and Patents Abridgements. Some 336 engineers in the organisation took part and user profiles were compiled for them in terms of the thesaurus, concentrating on their main interests.

No published thesaurus was considered suitable for the project and it was decided to use one based on the natural language. The 6,000 descriptors were divided among 52 subjects into which the field was split and again into three classes – the basic keywords of the fields, those which define and qualify the basic keywords and those which further qualify the others, so that there were three weighting levels. Synonyms were given the same coding as each other.

The indexers analysed the material with freedom to select any descriptors they thought appropriate, but the computer programme linked terms into selected bound terms and also worked on stems, neglecting inflexions. The indexers saw all the material in the store and gave their profiles, which were used by the computer to make a selection for them, so that they were enabled to calculate recall ratios for themselves, and recall ratios were also calculated for all users in the smaller fields of patents and retrieval requests.

All the participants made assessments on the relevance of the documents which were sent to them either on selective dissemination or in response to retrieval requests both for publications and for patents, and these assessments were used for the calculation of precision ratios.

For current literature the indexers had recall percentages of 74 and precision percentages of 60, while the other users had precision figures of 58. With patents recall was 63 and precision 67, and when adjustments were made to some of the users' profiles significant improvements were made, in one case over 90 per cent recall being accompanied by over 80 per cent precision. Other assessments were made of the performance of different indexers, particularly according to their background, and of other factors.

THE MINISTRY OF TECHNOLOGY

One of the principal roles of the Ministry is to assist Industry to make the maximum possible effective use of available scientific and technical knowledge so that British design, development and production incorporate the most up-to-date technology. A very wide variety of means is deployed to bring the work of the Ministry's research establishments and of the Ministry-supported research associations to the attention of industry, ranging from publications of research papers and in technical journals to the organisation of exhibitions and seminars and the production of firms.

To assist this process, the Ministry has established a national network of nine Regional Offices and, working with these, a growing number of Industrial Liaison Centres sponsored by the Ministry at Colleges of Technology and at some technological universities. The main functions of the Regional Offices are to assist industry to make full and profitable use of the technical information, advisory and research resources - Government and otherwise - available to firms, and to provide the Ministry with information on industry's needs for programmes of research, development and technical support. The services of the Industrial Liaison Centres are directed mainly at the small-to-medium sized firms, which comprise the great majority of British manufacturing establishments and which are usually unable to support their own development units. The centres rely very much, as do the Regional Offices, on personal contact for achieving their aims. The educational, advisory and laboratory facilities of their Colleges, especially those relating to research, design and production matters, are a key element in the work of the Centres.

All these developments depend for their success upon a cooperation between Industry and the sources of information, and the Regional Offices and Industrial Liaison Centres depend essentially upon Industry asking them for advice and information. Some of the other projects, however, are examples of the selective dissemination of information, and one of these is Techlink.

Techlinks are essentially information sheets about particular ideas and small developments which have occurred incidentally to work in the establishments and laboratories of the Ministry, of other Government Departments or of Research Associations. A unit is being established which is scanning a wide field of unpublished and published technical information, selecting useful items and then producing the leaflets which give the essential information and which are sent to those who have declared their interest in the subject area. There are over 50 subject areas, which range from aerodynamics through plastics and rubber to food processing, and the Regional Offices act as contact points for those who wish to use the service. In addition to the staff scanning documents, the service is also receiving unpublished material from some laboratories, and although the sources at present are entirely Government controlled or supported, it is hoped that Industry will be providing material for Techlinks as well as using them.

Some work is being done which is quite unsuitable for publication but which may none the less throw up ideas which are capable of development for Techlink. The projects could not support development of these ideas for publication, and in some establishments a small staff is now being maintained by the Ministry specifically to work up such ideas to the stage at which they would be suitable for dissemination. Another project is Interlab, which is organised on a regional basis to encourage the more intensive and economical use of specialised research and development facilities in Industrial, Government and Academic establishments, which agree to provide cooperating organisations with advice on equipment and techniques, loans of instruments and other services.

Recent developments have been the introduction of a Production Engineering Advisory Service by the Ministry; the launching of a drive, "Approaching Automation", to encourage the wider and more rapid application of low-cost automatic control devices; and the offering of training facilities for industry through short course at the Ministry's Building Research Station and, at a higher level, through the new Institute of Machine Tool and Control Technology at the National Engineering Laboratory, East Kilbride. Yet another approach being developed by the Ministry is the sponsoring of Design Data Sheets to put key, critically evaluated, engineering data into the hands of designers in a readily usable form. In this programme, the Ministry is collaborating with the professional engineering institutions.

UNILEVER RESEARCH LABORATORY, PORT SUNLIGHT

At this laboratory a selective dissemination of information (S.D.I.) service which has a number of interesting features is operated for the benefit of the staff there. The first step is that a number of information scientists scan the 370 incoming periodicals and select the articles which are of general interest to a large part of the laboratory or of great interest to a specific group or to some individuals; in the latter case the common topic is indicated and any particular individuals who are concerned are named. There is a weekly bulletin of bibliographic information about the articles thought to be of interest to more than a few individuals, but with no abstracts, and the individuals who are thought to be specially concerned with the contents of any articles are separately notified of them. There are no abstracts and there is no circulation of periodicals since all the customers are on the site and can walk to the reading room in a very few minutes. The Information Bulletin has been produced for a considerable period and the S.D.I. service and other refinements have been added to increase its usefulness.

The information scientists who select articles for the system visit the laboratories regularly in order to keep up to date with the interests of the scientists, which are served by the main bulletin, by some eight supplements addressed to group interests, by individual notifications and by a keyword-in-context (KWIC) index. The information scientists mark search sheets for each periodical on which they indicate the pages on which interesting articles appear, the subject area for any supplements to the bulletin, the groups and individuals who are concerned, any expansion of the notes to make them more informative together with the keywords for K.W.I.C. indexing and any additional keywords. This information is then transferred to 80-column punch cards which are then fed to a document writing system and to a computer. The latter produces the K.W.I.C. index and organises the information used by the document writing system for producing the bulletin, supplements and the specialist notes which are sent to individuals; the steps in the computer after the correction of errors without holding up the computer are the compilation of all titles in a form suitable for the K.W.I.C. index, the collection and sorting of items for the bulletin and associated publications, which are automatically punched onto tabulator cards for the document writing system.

WARREN SPRING LABORATORY

On the basis of previous experience in the microfilm field and after consultation with users it has been decided to make available a large number (initially about 40 for 120 users) of cheap, portable, desk-projection readers designed primarily for use with fiche. It is anticipated that these will cost no more than £25 and they may be actually available for as little as £10.

To back up these portable readers the Library will have a reader-printer which it is hoped will use a dry-process for printing and a fiche copier is being installed to provide positive fiches from either negative or positive originals. It is intended as a general practice to provide fiches instead of loans and full size copies.

Negotiations are currently in hand between a microfiche producer and various publishers for fiche to be sold, with the hard copy, by the publisher's own organisations. Where this is achieved it is not expected that there will be any difficulty in obtaining back runs. In other cases it may be necessary to get fiche specially made.

Ultimately it is hoped to have all journals on fiche possibly to the exclusion of hard copies. As it becomes advantageous, other material e.g., pamphlets, reports etc. may be put on fiche.

Whilst the timing of this move has been largely controlled by space considerations, other advantages include economy on provision of copies and greater efficiency in handling and filing of standard units in place of the miscellaneous sized parts and volumes of normal hard copy.

Ergonomics Abstracts for some time now has had its cumulative index maintained on aperture cards but this is being used only within the Laboratory and has not been publicised. Further developments in Air Pollution Abstracts and internal Mineral Processing Abstracts are planned; Air Pollution Abstracts is pre-indexed under a fixed concept (numerical) list of headings and a Thesaurus for Mineral Processing Abstracts is now being compiled for the same purpose.

Appendix I

CONVERSION OF WISWESSER NOTATION TO A CONNECTIVITY MATRIX FOR ORGANIC COMPOUNDS

E. Hyde*, F. W. Matthews† and L. H. Thomson†

INTRODUCTION

Investigations have been carried out into methods of recording organic compounds for use in computer systems. The objective of these investigations was to establish a compound file which would be suitable both for the analyses of structure/property relationships, and also for use in generic classification for information retrieval purposes. The study investigated an atom-by-atom connectivity system based on mathematically derived matrices and showed this method to be too cumbersome for the proposed system. It also clearly demonstrated that in any method adopted the identity of chemically significant groups must be preserved. It was, therefore, decided to examine the Wiswesser notation of a molecule, which by avoiding the use of mathematical arrangement of symbols, preserved the integrity of molecular arrangement. A further point in favour of the notation was that it produced a compound record which was concise and, hence, an efficient computer language for input purposes. These investigations showed that a matrix maintaining the chemical identity of the molecular arrangement could be computer generated from the notation, and that the resulting compound record averaged 60 characters. The compacted matrix constitutes an unambiguous record of a compound, and is in a form suitable for search and correlation purposes.

INVESTIGATIONS

When the project was set up the two most promising candidates for compound description were thought to be

(a) Atom by atom connectivity

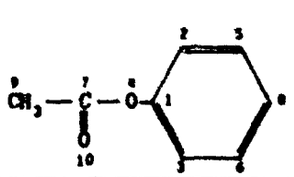
(b) Notations.

During the first six months of the project various existing systems were compared, and the system based on the Wiswesser notation was devised. It has now been developed to the point where it has been shown to comply with the objectives.

ATOM BY ATOM CONNECTIVITY

There are two problems associated with any atom-by-atom approach. Firstly, the vast majority of single atoms in any molecule have no descriptive value for search purposes, and secondly any atom-by-atom matrix comprising, as it does, not only a description of atoms but also that of bonds, is a bulky record. If the next step is a mathematically generated matrix in order to ensure a canonical ordering of the atoms, then all chemical significance is destroyed in the resulting element listing.

The following example will give a clear picture of the resulting disruption of the record of a simple molecule.



* Imperial Chemical Industries Limited.
† Canadian Industries Limited.

The canonical ordering of the atoms derived on a mathematical basis for ultimate magnetic tape storage is as follows:-

Thus the record states

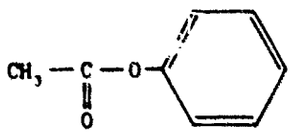
Atom No.	El	Bond	Connection
1	C	L	-
2	C	L	1
3	C	L	1
4	O	1	1
5	C	L	2
6	C	L	3
7	C	1	4
8	C	L	5
9	C	1	7
10	O	2	7

Ring closure 8-6.

NOTATIONS

There is ample evidence that a significant advantage of notation is that they provide an extremely cheap method of describing compounds in a computerisable form. At a Wiswesser seminar organised by the US Army users of the notation including Dow Chemical Company claimed high levels of both accuracy and input speed (J.Chem.Doc. 7, No.1, p.43).

Using the Wiswesser notation the example compound given previously would become



1 V O R

If this form is examined it becomes obvious that the notation has overcome a number of the problems created by an atom-by-atom system. It is canonical in the linear ordering of the notation symbols and this ordering has not destroyed the arrangement of the atoms in the molecule. It is concise because bonds and atoms have been compacted into one symbol and due to the linear arrangement there is no need to state connectivity. Finally, it has enriched certain elements to the point where their chemical significance and differences are clearly shown. The carbons in the example are described as 1 in the methyl group, V in the carbonyl and R in the ring atom. Thus scrutinising a molecule by computer becomes a much simpler task, the symbols acting as a fragment screen.

However, in achieving these linear representations of molecules the resulting cyphers are unintelligible except to those people skilled in their use.

CONNECTIVITY DERIVED FROM NOTATION

When retrieving data from an organic chemical file the questions are usually composed of part structures. They require the searcher to retrieve two or more atoms connected in a specific manner. Thus it is of prime importance to show the functional differences of elements and the way they are bonded to other elements as quickly and as effectively as possible. A notation contains the data, but in complex molecules not in an immediately accessible form. It was logical, therefore, to examine the possibility of computer generating a connectivity matrix from a notation, and in so doing preserve the advantages of notations outlined above.

"DOT PLOT" SYMBOLS

The Wiswesser notation does not spell out every single atom in a molecule, but instead points out the type, shows repetition and indicates change. It is therefore necessary to generate from the notation all excluded atoms, because these constitute nodes in any derived connectivity network. For example the notation for naphthalene is L66J from which is inferred that the compound is composed of two fully unsaturated carbon rings fused together. If it had been other than t' is the notation would have made suitable notes to this effect. Thus quinoline would be T66 BNJ. The T indicating a heterocyclic ring system and the BN indicating that the carbon in the B position has been replaced by a nitrogen atom.

If a connectivity network is to be composed then some symbols must be used which do not appear in the notation record.

Earlier W.J.Wiswesser had been working on an entirely different approach for describing ring systems. This system, "Dot Plot", comprised spelling out every node in the rings using the following symbols for ring carbon.

L	$-CH_2-$	
Y	$-\overset{\vert}{CH}-$	(as in the notation)
X	$-\overset{\vert}{\underset{\vert}{C}}-$	(as in the notation)
D	$-CH=$	
T	$-\overset{\vert}{C}=$	

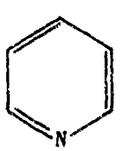
The above letters had been carefully chosen so that they would not interfere with existing symbols in the notation. It was obvious that these symbols could be used to expand the ring notation and provide the nodes essential for a connectivity network.

The problem remaining was therefore to examine the possibility of decyphering a standard notation and to generate the above symbols for the omitted portions of the ring record.

GENERATION OF "DOT PLOT" SYMBOLS FROM WISWESSER NOTATION

A programme has been written which builds a connectivity matrix using both Wiswesser notation and Wiswesser Dot Plot symbols. This programme is better understood by the consideration of actual examples.

Pyridine



TGNJ

The programme detects the number following the T symbol and allocates a linear record of that number of D symbols

D D D D D D
1 2 3 4 5 6

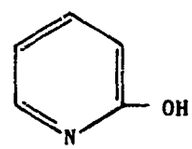
The next step is to read the N , which indicates a nitrogen, with no hydrogen attached, at the first position, and the programme overwrites the first D with an N

N D D D D D
1 2 3 4 5 6

Thus the matrix for pyridine would be

```
N * †
D  ††
D   ††          D = -CH=
D    ††
D     ††
D      ††       * indicates ring closure
D *     †
```

If this compound had contained a substituent e.g.



then the notation would be

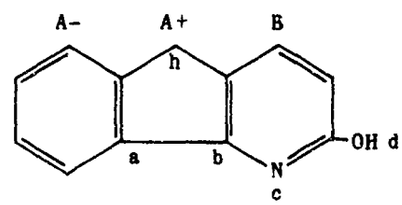
T6NJ BQ .

The programme reading the BQ adjusts the D at the second location, and the units of the matrix become

N T D D D D Q

where T = $-\overset{|}{C}=$

If fusions were involved as in the following compound



then the notation would be T B656 CN HHJ DQ.

The programme works character by character through the notation and commences by examining the first ring which is the one whose lowest character is B , and then the adjacent ring (A+1) and finally ring (A-). The record generated would be Rings

B	D D D D D D	Modified to	T N T D D T
A+	D D D D D		T T T L T
A-	D D D D D D		T T D D D D

CN and DQ would modify the appropriate characters as in the earlier examples and HH , showing that an additional H on the H atom, causes D to be replaced by L to indicate $-CH_2-$. The programme notes the overlapping symbols at the fused positions and modified a D to a T , i.e. changes these from -CH= to $-\overset{|}{C}=$.

The resulting connectivity matrix for this example is given on page 36. For magnetic tape store the matrix is recorded as follows.

Units TNTDDTTTTLTTTDDDDQ

Connection Transfers 18.3

Ring Block 1-6, 8-9, 7-11, 12-13, 12-17
size fusion size fusion size

(Connection transfers show the modifications to the matrix diagonals caused by substitution)

In arranging this record the following criteria has been taken into account.

1. Ring atoms should be clearly identifiable as being in the same ring.
2. Ring size should be capable of being readily deduced.
3. Record should indicate position and type of fusion.
4. The linking of substituents should be stated.
5. The linear arrangement of atoms should be clearly indicated so that atom by atom searching can be carried out as far as possible from the compacted record.
6. It should be possible to reconstruct the matrix efficiently when a search question demands an exhaustive searching of a structure.

APPLICATION OF THE CONNECTIVITY MATRIX

Fragment codes are a convenient way of describing a molecule in a file on which mathematical analysis is to take place. To use fragmentation codes for this purpose, however, the code must be specifically designed to reflect the topic under evaluation. Therefore one application of the connectivity matrix derived from the Wiswesser notation has been to generate fragmentation codes by algorithms.

Most computer systems in operation today give only a file reference number as the output to any search. A few systems carry a digital representation of the structure, which is available for display either on a computer line printer or a chemical typewriter. Obviously, a computer system which, as output, economically produces structure diagrams is preferable to one giving only file reference numbers. During investigations into various forms of output, consideration has been given to computer generating the structural formula from the search record.

PART I - A COMPUTER GENERATED OPEN ENDED FRAGMENT CODE

The object of this work has been to allow the computer to generate fragments, having been programmed to follow established guide lines: at the commencement of the operation, the fragments which will be generated are not specifically designated. As novel compounds are added to the file, the programme will generate new fragments as it meets a new situation, and hence the fragment code has the advantage of being open-ended.

The programme operates directly from the compacted matrix. Each fragment generated is composed of a string of Wiswesser symbols in canonical order and varies from 2 to 10 symbols in length, the majority being 4 symbols long. In general the programme reads from a ring or alkyl chain to a terminal group and picks up all symbols surrounding non-aliphatic branching units.

The programme, by direct examination of the compacted matrix located

(a) all branch units and catalogues these as

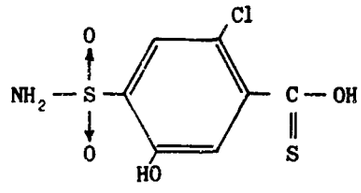
Group I - those which can act as starting points for fragments e.g. rings and alkyls.

Group II - those which are the centre of fragments e.g. Nitrogen in amine groups, sulphur in sulphonamides.

(b) all terminal groups e.g. hydroxyl, chloro.

(c) in addition, the programme is required to generate the longest path in the notation and the points on this path where branching occurs.

Consider the following compound.



which is represented by the compacted matrix record,

Units ZSWRQGYQS Connection transfers 32, 54, 64, 87

123456789

(a) The branch units in this molecule are at positions 2, 4 and 7. These are tagged so that unit 4 is recorded as a "Group I" unit and Unit 2 and 7 as "Group II" units.

(b) the terminal groups are at positions 3 5 6 8 9

(c) the longest path consists of units 1 2 4 7 9 and the side branches are 23, 45, 46, 78.

The programme reads from the beginning of the molecule and using this data develops the following unit combinations:

1 2 3 4
4 5
4 6
4 7 8 9 .

Note that during this operation it was not necessary to examine the Wiswesser units. The routine was performed entirely from the numeric data available in (a), (b) and (c).

As a final step, the four unit combinations listed above are converted into the following fragments expressed in Wiswesser units as:

Z S W R	- NH_2SO_2R	R = Phenyl
R Q	- ROH	
R G	- RCl	
R Y Q S	- RCSOH	

Every fragment is assigned a number, and a compound is registered by entering its serial number under each fragment contained in the molecule in an inverted file.

The fragments thus obtained may be listed using a KWIC programme (Appendix II). This brings together all fragments containing the Wiswesser symbols in common. An enquiry made of the file is examined against pertinent sections of the KWIC to establish under which fragments the search should be performed.

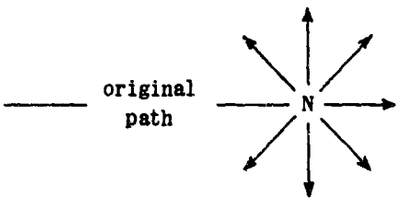
By altering the rules for deriving the stop units and reclassifying the definitions of Group I and Group II it is possible to generate different sets of fragments. Therefore for structure/property relationship, molecules can be fragmented specifically for the problem under examination.

PART II - GENERATION OF STRUCTURE DISPLAY FROM A WISWESSER CONNECTIVITY MATRIX

The object of this work was to establish the feasibility of using the compact record derived from the Wiswesser notation to generate an acceptable structure display for output on a lineprinter. An advantage of this approach is that one record serves the dual purpose of both search and display.

However, a programme for generating display must compete cost-wise with the alternative method of holding a separate tape record for display. At some point the computer generation of a structure will be more expensive than holding a separate record. In these cases it is the intention to create a separate display file.

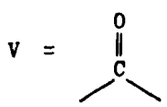
The programme is basically a free plotting routine which considers each single ring of a ring system or branching units as plotting and inspection points. At each such point the programme allows for seven possible changes of direction.

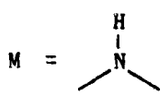


N = branching atom or centre of origin of ring

Rings

The first routine in the programme deals with rings and generates the linking bonds between the ring atoms. This routine reads the ring portion of the matrix, tags all atoms which are shown by their symbolic representation to be single bonded within the ring, e.g.

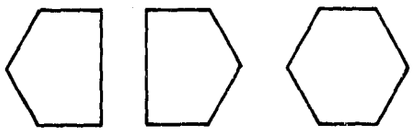




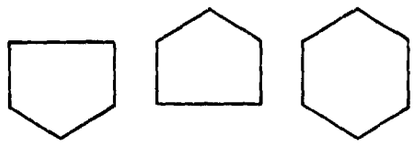
and then inserts alternating double bonds between the remaining atoms, commencing with a double bond. The final step in this routine is to mark the atom from which the point of origin is generated for each single ring within a ring system.

When commencing to plot a ring the programme next established the ring centre as its point of origin. By inspecting the direction from which the ring was approached the programme is able to select either a horizontal or a vertical form of 5 and 6 membered rings.

Horizontal

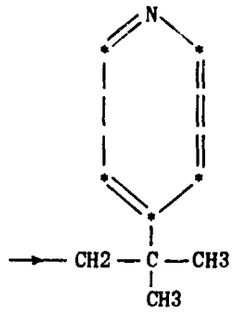


Vertical



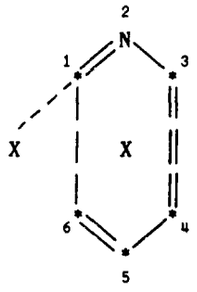
It will proceed to plot these using * for carbon atoms and will insert the given bonds

e.g. Pyridine



by noting the coordinates of the plotting position of the lowest ring fusion atom in the first ring the programme is enabled to develop the point of origin of the next ring

Quinoline



X = points of origin

Branched Units

Plotting from a branched unit uses the same programme routine as that used for plotting rings from the point of origin. However, the specific branching atom is read by the computer and the information used to select the preferred paths for that atom.

An additional routine keeps track of the area used up, and this information is inspected at each plotting point. If a particular path would lead to overwriting then this tracking route will modify the preferred paths.

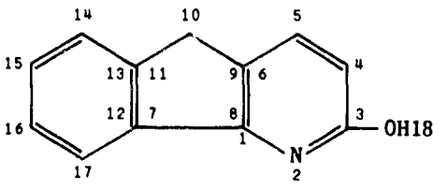
Line Printer Character Set

The programme has been designed for an IBM 1410 but the structures given in Appendix III were printed on the ICT 1004.

It is yet too early to state that the programme will meet the desired objective of economically generating at least 85% of the structures on file.

Connectivity Matrix from Wiswesser Notation

```
 1  T * †             †
 2  N  † †
 3  T    † †                      †
 4  D      † †
 5  D        † †
 6  T *       †       †
 7  T *         †           †
 8  T           † † †
 9  T             † † †
10  L               † †
11  T *               †     †
12  T *                 † †
13  T                   † † †
14  D                     † †
15  D                       † †
16  D                         † †
17  D *                       †
18  Q                           †
```



* - Ring Closure

† in a Vertical Column indicates connection

L $-CH_2-$

D $-CH=$

T $-\overset{|}{C}=$

COMPACTING THE MATRIX

The compacted matrix form for this molecule is

Units TNTDDTTTTLTTTDDDDQ

Connection Transfers 18.3 (substitution)

Ring Block 1-6, 8-9, 7-11, 12-13, 12-17

size	fusion	size	fusion	size

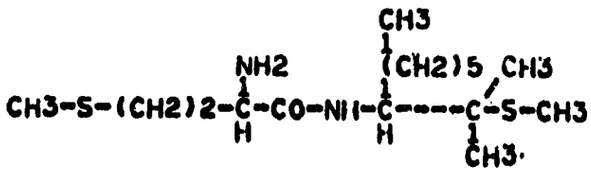
(Connection transfers show the modifications to the matrix diagonals caused by substitution)

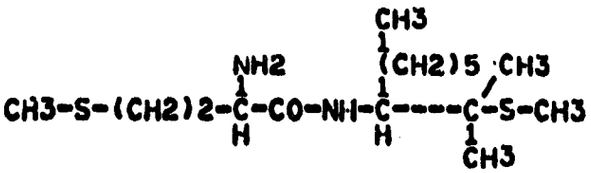
Appendix II

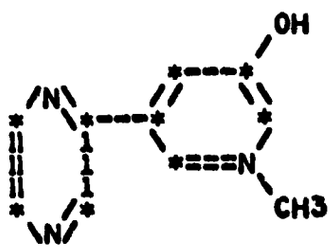
SECTION OF KWIC LISTING OF OPEN ENDED FRAGMENT CODE DERIVED FROM WISWESSER CONNECTIVITY MATRIX

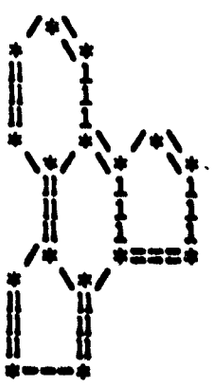
	Fragment #
*R O | M R	588
*R S W | M R	518
*A M Y M | M R	693
*A | M R	235
*A S W | M R	756
*R | M R	802
*A O | M R	278
*A | M V A	815
*A M V M | M V M A	615
*A M V | M V M A	271
*A | M V M A	159
*A | M V M M V M A	615
*A | M V M V M A	271
*A | M V N A R	493
*A | M V Q	837
*R | M V Q	791
*A | M V R	570
*R | M V R	135
*R | M V Z	456
*A | M V Z	059
*A M Y M | M Y M A	476
*A | M Y M M R	693
*A | M Y M M Y M A	476
*A | M Y M M Y Z M	032
*A | M Y S N R R	312
*R | M Y S S Y S N A A	680
*A M Y M | M Y Z M	032
*A | M Y Z M	259
*R S W | M Z	123
*R V | M Z	234
*R | M Z	468
*A C | N	061
*R C | N	066
*R | N A	246
*R Y S | N A A	891
*R M Y S S Y S | N A A	680
*A | N A A	499
*R O | N A R	812
*A M | N A R	494
*A M V | N A R	493
*R | N N R	578
*R | N Q A	801
*A | N Q A	834
*R N | N R	578
*R S W | N R R	911
*R | N R R	136
*R M | N R R	024
*A M Y S | N R R	312
*R | N W	913
*A | N W	810
*R V | N Z A	567
*A | O A	045
*R | O A	046
*R V | O A	026
*A M | O A	838

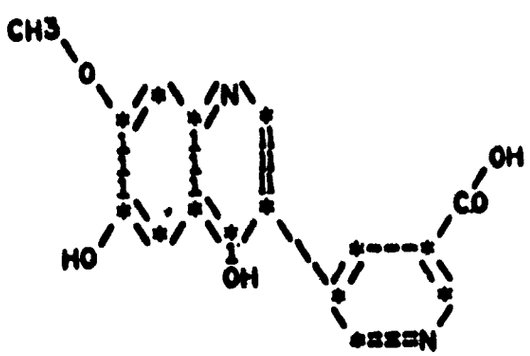
Appendix III











N. B.

These structures have been printed on the ICT 1004.

They have *not* been generated from the Wiswesser notation but are displayed in the way they would be generated by the programme under test.

Information regarding the availability of further copies of AGARD publications may be obtained from

The Scientific Publications Officer,
Advisory Group for Aerospace Research and Development,
7, rue Ancelle,
92 Neuilly-sur-Seine,
France.

Printed by Technical Editing and Reproduction Ltd
Harford House, 7-9 Charlotte St. London, W.1.