
INFORMATION MANAGEMENT RESEARCH IN EUROPE

Proceedings of the EURIM 5
Conference, Palais des Congrès,
Versailles, France, May 1982.

Edited by
Peter J. Taylor and Blaise Cronin

Aslib

in association with
en collaboration avec
in Zusammenarbeit mit

Association Nationale de la Recherche Technique
Bibliothèque Royale de Belgique
Consiglio Nazionale delle Ricerche
Deutsche Gesellschaft für Dokumentation
NOBIN
NORDFORSK

INFORMATION MANAGEMENT RESEARCH IN EUROPE

Proceedings of the EURIM 5
Conference, Palais des Congrès,
Versailles, France, May 1982.

Edited by
Peter J. Taylor and Blaise Cronin



Aslib

in association with
en collaboration avec
in Zusammenarbeit mit

Association Nationale de la Recherche Technique
Bibliothèque Royale de Belgique
Consiglio Nazionale delle Ricerche
Deutsche Gesellschaft für Dokumentation
NOBIN
NORDFORSK

Adise
Paula
S/O

157284
D.N.
D.N.
8/12/89

© Aslib 1983
3 Belgrave Square,
LONDON SW1X 8PL,
United Kingdom.

PC
029.07204
TAY

ISBN 0 85142 171 7

Contents

SESSION 1

| | | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------|----|
| <i>The management of information systems and libraries</i> | | 1 |
| Do-it-yourself manuals: research based aids for the information service manager | <i>T.D. Wilson</i> | 3 |
| Results and questions of research in information economics | <i>Werner Schwuchow</i> | 10 |
| A generalised profile of an internationally — oriented information management system for libraries and information centres, with MINISIS as a case study | <i>C.A. Godfrey</i> | 25 |
| Interfirm comparison: a management technique | <i>Maureen Grieves and M. Moffatt</i> | 30 |
| Summary of discussion | | 43 |

SESSION 2

| | | |
|----------------------------------------------------------------------------------------|--------------------------------------|----|
| <i>Retrieval techniques</i> | | 45 |
| Automatic indexing and classification for mechanised information retrieval | <i>T.M. Aitchison and P. Harding</i> | 47 |
| Une stratégie d'indexation automatique de documents écrits | <i>Richard Bouché</i> | 56 |
| Automatic indexing: a summary | <i>Gerard Salton</i> | 66 |
| Techniken für den Zugriff auf numerische Daten in alphanumerischen Referenz-Datenbasen | <i>Dieter König</i> | 78 |
| Summary of discussion | | 89 |

SESSION 3

| | | |
|---------------------------------------------------------------------------------------|------------------------------------------------------|-----|
| <i>Information transfer and delivery</i> | | 91 |
| New technology and R & D information: an international comparison of trends | <i>A.J. Meadows, A. Singleton and M. Van der Lem</i> | 93 |
| Planning the future: computer-based modelling and the dynamics of information systems | <i>B.C. Vickery and R.G. Heseltine</i> | 98 |
| Informed enterprising | <i>R.H. da Silva</i> | 106 |

| | | |
|---------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------|-----|
| Technical aspects of document delivery using broad band techniques and satellites | <i>Jørgen A. Richter</i> | 114 |
| Video discs in information technology: actual and possible applications | <i>Fredrik Backlund</i> | 121 |
| The neglected resource | <i>Margaret Slater</i> | 126 |
| Utilisateur, producteur, serveur: un ménage à trois pour le meilleur ou pour le pire? | <i>Bernard Marx</i> | 133 |
| The social context of information transfer: an analysis of facilitating and inhibiting factors | <i>Mark van de Vall and Nelleke Schoemaker</i> | 140 |
| Changes in occupation and profession in information work: the impact of new communication technologies on information work | <i>Thomas Seeger</i> | 152 |
| Summary of discussion | | 159 |
| SESSION 4 | | |
| <i>Multilingual and international systems</i> | | 161 |
| Multilingual information systems backed up by machine aided partial translation can provide the information needed for the 80's | <i>Peter P. Canisius</i> | 163 |
| Machine translation's big ten years: 1980-1990 | <i>Loll Rolling</i> | 169 |
| TITUS IV: système de traduction automatique et simultanée en quatre langues | <i>J.M. Ducrot</i> | 177 |
| An alternative strategy for steady growth towards high quality translation networks | <i>A.P.M. Witkam</i> | 196 |
| Summary of discussion | | 205 |
| CLOSING SESSION | | |
| Closing address | <i>J. Meyriat</i> | 209 |

Session 1

THE MANAGEMENT OF INFORMATION SYSTEMS AND LIBRARIES

'Do-it yourself' manuals: research-based aids for the information service manager

Dr. T.D. Wilson

*Department of Information Studies,
University of Sheffield*

INTRODUCTION

The means whereby the results of research in the information field may be put to use in practice is a subject of perennial interest to the funders of information research and to practitioners. Often the call is made for 'more effective dissemination' but this rarely seems to work (for reasons which I have attempted to explain elsewhere (Wilson, 1981)) In fact it seems that the ways in which the results of research may be disseminated in an effective way, to bring about change in the operation and management of information services, are limited in number:

- 1) it is possible to carry out the research itself in such a way as to effect change in the course of the project. This is generally called 'action research' and has been employed to considerable effect in the fields of education and organizational design. (For an information science application see Wilson and Streatfield, 1981);
- 2) training courses may be designed on the basis of the research which has been carried out, either for the staff of the organizations that have been involved in a project, or for the information community at large. For an application to members of staff of organizations which participated in a project see Streatfield, Wilson and Clark, 1981. A 'softer' and probably less effective version of this approach is to hold seminars or conferences to disseminate research results.
- 3) manuals, which embody the results of research, may be produced for the practical guidance of the managers of information services.
- 4) finally, the researcher may turn consultant and seek to use the knowledge gained through research for the benefit of organizations not involved in the original work.

The purpose of this paper is to report on the production of two manuals of the kind noted above, now published by the British Library Research and Development Department (Francis, Mullings and Wilson, 1980 and Mullings, Francis and Wilson, 1980).

BACKGROUND

Following a major investigation into information needs and information use in local authority social services departments (Wilson and Streatfield, 1980) a grant was received from the British Library R & D Department for further work towards the production of tools which would enable library/information service managers to carry out similar work themselves. Although the focus of the work was information service in local government, it was hoped that the manuals could be made sufficiently general to be of use to the managers of information services in any other field.

To test the research methods and data collection instruments produced for the manuals, field work was carried out in a number of local authorities. In the case of the information needs manual, this involved 52 interviews in 4 different departments, as different in character as a recreation department and a public health department, and a postal survey covering 62 respondents in six departments: architecture, planning, engineering, housing, finance and environmental health. The bulletin evaluation manual involved two existing bulletins, one general-purpose covering all fields of local government, the other a special-purpose bulletin covering social work only. Again, the work involved extensive data collection and interviewing of users.

THE INFORMATION NEEDS MANUAL

The information needs manual includes information on the background to the questionnaire and the interview schedule, copies of the two instruments, instructions on how to use them, notes on the analysis of the resulting data manually or by computer, and, finally an account of the use of a draft of the interview schedule in a 'live test' to discover the information needs of staff of the London Fire Brigade. Three appendices give information on the purpose of the questions and some examples of typical responses, and copies of the coding manuals for the questionnaire and the interview schedule.

The interview schedule, based on the earlier version used in Project INISS (Wilson, in press) contains 29 questions, many subdivided, under the following 8 headings:

Work interests; People as information sources;
Information-seeking and receiving; Information needs;
Information use; Document-based information sources;
Journals and information bulletins; and Background data.

The questionnaire, which is intended to be self-administered as in a postal survey, is similarly structured, with the following sections:

'Your work'; Information seeking and receiving;
Information needs; Information sources (subdivided
into 'People as information sources', 'Local collec-
tions of information', Journals, and Information
bulletins); and Background data.

In the case of both instruments the term 'information' is broadly defined; a definition is given for recipients of the questionnaire:

'INFORMATION - In the following questions, whenever we refer to information, we mean any knowledge that you already have about a subject; advice, facts and opinions that you receive from other people; and data contained in documents'.

A similar definition is printed in the interview schedule at the top of a response sheet for different kinds of information used by respondents. Curiously, in the earlier work under Project INISS, no difficulty had been experienced with the concept of 'need', but a number of respondents had asked the question, 'What do you mean by information?' When the definitions were added to the instruments the only comments received tended to be those seeking clarification that a particular type of information of relevance to the respondent was included in the definition.

The aim of the information needs manual was to offer as much flexibility to its user as possible: thus, both an interview schedule and a self-administered questionnaire are presented, and both are in modular form to allow the user to select those sections which will be most appropriate to the nature of the intended investigation. Furthermore, advice is given on the ways in which the instruments may be used: it is suggested, for example:

'... that one of the most valuable uses to which the interview schedule can be put is to use it informally. By this we mean that a newcomer to local government information service (or someone planning such services) should use the schedule as a basis for discussions with individuals he is expected to serve. The learning process that takes place through this contact with the potential user is much more valuable than all the data collected by other persons, no matter how well the results are presented'.

On the other hand, advice is also given on the formal use of the schedule with trained interviewers if they are available or with members of staff of an information service. In the latter case attention is drawn to the need for training and how to acquire it.

Flexibility is also offered in relation to the analysis of data: manual, edge-notched card, and computer analysis are discussed and, in relation to the latter, advice is given on the use of the SPSS statistical analysis package.

A considerable amount of information is also given on the wording of questions, including *why* a question was included, and an appendix gives sample answers, drawn from the pilot tests, to many. It is intended that this kind of information should enable the user of the manual to select questions, or whole sections, which are most appropriate to his/her needs for data.

THE BULLETIN EVALUATION MANUAL

As in the case of the information-needs manual, this manual arose out of earlier work under Project INISS. During that project a number of libraries in the East Midlands of England had begun to produce, cooperatively, a current-awareness bulletin directed at a specialized group of users, the *Social Work Information Bulletin*, and some evaluation of its contents and use had been undertaken. Other bulletins produced through cooperation, but more general in their intended audience, were being produced in other parts of the country and it was agreed with the British Library R&D Department that a manual would be produced which would allow any library producing such a bulletin to evaluate it.

Again, field tests of data-collection methods, questionnaires and interview schedules formed part of the process of developing the manual. Two modes of evaluation are covered by the manual: the analysis of demands and costs, and 'user-evaluation', that is, the collection of information directly from the users of the bulletin as to its suitability and usefulness in meeting their information needs.

As in the information-needs manual, questionnaires and interview schedules are presented and their use is discussed, but a special feature of the work on this manual was the preparation of computer programs for the analysis of journal input to the bulletins and demand data (SPSS being used for the analysis of interview and questionnaire data). The computer programs are fully documented in the manual and remain available for use in the Computer Services Department of the University of Sheffield.

Alan Gomersall, Research Librarian of the Greater London Council, who served on the advisory committee to the project and who wrote the Foreword, commented there:

'For those who may doubt the adaptability or even the practical value of a manual created within the research environment, I would assure them that its use during the summer of 1979 in my own organization generated considerable interest amongst our abstracting bulletin customers. We were given useful information on their appreciation or otherwise of *Urban Abstracts* and subsequently we have made changes in arrangement and content, recommended by customers who felt that such changes would be of benefit to many of their colleagues throughout the Council.'

RESPONSE FROM THE FIELD

As yet the response from the field has been limited; there are some indicators, however: firstly, the user needs manual was actually tested by an information service while it was still in the development phase; and, secondly, there have been reviews by people with experience in the field which provide response of another kind.

To the best of my knowledge, at the time of writing, only the information needs manual has been reviewed (perhaps the larger, bulletin evaluation manual is too daunting). One reviewer has noted that:

'The principle characteristic of the manual is its flexibility. It does not set out rigid rules for the investigation of information needs, in recognition of the fact that departments and authorities may vary widely. It is capable of considerable adaptation.'
(Grayson, 1982: 72)

She goes on to note the diverse ways in which the manual and the survey instruments may be used and recognizes most of the points made above in the section on this manual.

Another reviewer comments that the research team,

'... have provided an invaluable and time saving tool quite unlike the more usual end-products of academic research. Its principle strength for the hard-pressed practitioner is that it can be taken out of the file and actually put to use in most 'user survey' situations with a minimum of modification for local circumstances. The case study carried out by librarians at the Greater London Council...shows this clearly'.
(Kennington, 1982: 54)

The conclusions of the GLC researchers who tested an early draft of the questionnaire (using it as an interview schedule) point to the advantages to be gained from an exercise such as this:

'What the survey did achieve was to provide the impetus for an overview of the Brigade's information needs rather than merely focussing on the library.'

'... the questionnaire supplied a large amount of information that will be invaluable in further dealings with Brigade management on library matters; it also provided a vehicle for publicizing Research Library services.'

'Those who supported the idea of an improved facility viewed it mainly as a reference library, but the survey has brought out the fact that the Brigade also needs an information service whether or not this is realized or admitted. Even if that had been the only result of the survey the exercise would have been worthwhile.'

CONCLUSIONS

The aim of this Project was two-fold: to discover how far the findings of research in one area of local government (social services) were applicable to other areas, and to prepare manuals to guide the local government information officer in the conduct of his/her own research or evaluation. Different measures of success need to be applied to both of

these aims, clearly. The first aim was fulfilled entirely; it was found that, with very little modification, the research instruments used in social services departments were applicable to other areas and, although not reported here, the results, in terms of such variables as the kinds of information used, contact with outside agencies, contact within the department, the constraints of work upon information-seeking, etc., were compatible with those obtained in the earlier study.

Fulfilment of the second aim can be measured in different ways: the Project led to the successful production of the manuals and it was found possible to write these in such a way as to make them applicable not only within local government but also in other spheres of activity. Ultimate success, however, depends upon the manuals being used, and here we must be less sanguine about the possibilities. The problem, clearly, when one tries to give guidance on the performance of complex tasks such as needs investigations and evaluation of activities such as bulletin production, is that there is no way in which such guidance can be offered in a very brief document - detailed instruction is an absolute necessity. Nor can one hide the fact that both of these forms of investigation are going to be time-consuming for the information officer. Both of these factors are likely to be deterrents to the use of self-help manuals, and the motivation to change organizational behaviour must be strong if the task of applying the results of research presented in this form is to be undertaken. The dangers of this attitude are pointed out by both reviewers of the information needs manual:

'Local government information specialists have been far too willing in the past to assume that they know best both about the information that is needed and the ways in which it should be presented, and they have signally failed to provide any real justification for their products. The recent demise of several services is surely a sign that a more rigorous, and less patronizing, approach is necessary, and this manual goes a long way towards providing the means.' (Grayson, 1982: 72)

'Of course many of those involved at the practical level will protest that...they are far too busy providing services to have time to investigate needs. And in times when resources are hard to acquire...this can be, on the surface at least, a reasonably valid excuse for inaction in this area. It is, however, a myopic view since even those providing highly professional...services are certainly not fully aware of the *real* needs of their user groups and especially of how these are changing and evolving. Ignorance of these needs will, in the fullness of time, bring about an increasing risk of reductions in resources as the information service loses touch with the real issues faced by these users.' (Kennington, 1982: 53)

ACKNOWLEDGEMENTS

The work reported here would have been impossible without the funds provided by the British Library Research and Development Department whose support is gratefully acknowledged. It would also have been impossible without the very hard work of Gill Francis and Christine Mullings, respectively Research Officer and Principal Investigator.

REFERENCES

- FRANCIS, G.M., MULLINGS, C. and WILSON, T.D. (1980). *A manual for the evaluation of current awareness bulletins*. London: British Library R&D Department.
- GRAYSON, L. (1982). A manual for the investigation of local government information needs. (Review) *Journal of Librarianship*, 14, 72-73.
- KENNINGTON, D. (1982). A manual for the investigation of local government information needs. (Review) *Social Science Information Studies*, 2, 53-55.
- MULLINGS, C., FRANCIS, G.M. and WILSON, T.D. (1980). *A manual for the investigation of local government information needs*. London: British Library R&D Department.
- STREATFIELD, D.R., WILSON, T.D. and CLARK, D. (1981). A problem-based training approach to information for practitioners. *Libri*, 31, 243-257.
- WILSON, T.D. (1981). Letter to the editor. *Library and Information Research Group News*, 4, (15) 15-19.
- WILSON, T.D. (in press). Questionnaire design in the context of information research, in: *The research interview: uses and approaches*, edited by M. Brenner, J. Brown and D. Canter. London: Academic Press.
- WILSON, T.D. and STREATFIELD, D.R. (1980). *You can observe a lot... A study of information use in local authority social services departments conducted by Project INISS*. Sheffield: University of Sheffield.
- WILSON, T.D. and STREATFIELD, D.R. (1981) Action research and users' needs. Paper for the 4th International Research Forum on Information Science, Boras, Sweden.

Results and questions of research in information economics

Dr. Werner Schwuchow

*Sektion für Ökonomie (GID-Sfö),
Gesellschaft für Information und
Dokumentation mbH,
Heidelberg*

1. VORBEMERKUNGEN

Viele von Ihnen, die mit mir bereits an der EURIM 1-Konferenz im November 1973 in Paris teilgenommen haben, werden sich erinnern, daß damals bereits ganz überwiegend Themen diskutiert wurden wie: Methoden der ökonomischen Bewertung von Informationssystemen und Bibliotheken, Fragen statistischer Erhebungen im Bereich der wissenschaftlichen und technischen Information (wie es damals noch hieß), Methoden und Ergebnisse der Benutzerforschung, Managementtechniken für Bibliothekare, Kostenuntersuchungen und Kosten-Effektivitäts-Analysen in Informationssystemen, Kosten-Nutzen-Untersuchungen usw (1).

Obwohl also bereits auf der EURIM 1-Konferenz ökonomische Aspekte des Transfers von Fachinformation (wie wir heute sagen) im Vordergrund standen, sprach man damals noch nicht von einem eigenen Forschungsbereich 'Informationsökonomie'.

Dieser Teilbereich der Informationswissenschaften (im Englischen: '*Economics of Information*' oder '*Information Economics*') ist noch ganz jung. Ich würde sagen, daß er in seiner heutigen Gestalt eigentlich erst ein Kind der 70'er Jahre (vielleicht, sogar erst der späten 70'er Jahre) ist.

Die Gründe für die Entwicklung dieses Bereiches informations-wissenschaftlicher Forschung liegen sicherlich u.a. in:

- den Krisenerscheinungen in den Volkswirtschaften nahezu aller Industrieländer der Welt
- den damit verbundenen Einsparungsbemühungen in den öffentlichen Haushalten westlicher Industriestaaten und der stärkeren Rückbesinnung auf die Kräfte der Marktwirtschaft
- dem schnellen Wachstum des Marktes für Online - Informationsdienste
- den großen Entwicklungsmöglichkeiten des Informationsmarktes generell aufgrund der sich schnell entwickelnden neuen Informationstechnologien (Telekommunikation, Mikro- und Mini-computer, Videotext-Systems, Kabelfernsehen, Satellitenkommunikation, Bildplatte usw.)

A generalized profile of an internationally-oriented information management system for libraries and information centres, with MINISIS as a case study

Dr. C.A. Godfrey

*International Development Research Centre,
Ottawa*

SUMMARY

Any commercial or non-commercial organization interested in developing or marketing information-management systems internationally will begin by analyzing the generalized functions and features required by the user community, in the case at hand, libraries and information centres. Potential users can use the same profile as a tool in designing their own user specifications. Such a generalised profile is developed in this paper and is then used to evaluate MINISIS, an interactive computer system designed specifically for the above user community and in use in Europe since late 1979.

INTRODUCTION

Although we often have the illusion that the computerised information society has already arrived, as yet relatively few information management and retrieval systems have been developed for libraries and information centres that are widely accepted by them. Such software packages have generally been developed for mainframe computers to handle the needs of a specialised user group and have in most cases either provided processing for one or more library technical functions or retrieval from large data bases, but not both.

However, at least two major forces are currently creating pressure to change this situation. The first is well-known and is the decreasing relative costs of computer hardware associated with the advent of mini- and microcomputers. Computer processing power is reaching even small information centres through online computer services and/or through systems in-house. In relation to the latter, we can expect the appearance of new software packages for information management and retrieval in growing numbers and diversity.

The second and less widely recognized force is the creation and growth of international cooperative information networks such as AGRIS in the field of agriculture and DEVSIS in the field of social and economic development. Local, national and regional centres participating in these information networks require tools both to supply input to, and to process output from, the networks. These tools must satisfy a variety of international user needs.

Interfirm Comparison: a management technique

Maureen Grieves

*British Library Research and Development Department,
London*

M. Moffatt

*Centre for Interfirm Comparison,
London*

1. BACKGROUND

In 1977 the British Library Research and Development Department (BLRDD) received several approaches on the subject of comparisons in libraries. A review of the comparison work in industry and commerce led to the Centre for Interfirm Comparison (The Centre) which had been set up by the British Institute of Management and the British Productivity Council in 1959, following several years' research work funded by the (then) Board of Trade.

The Centre is not a commercial organisation; it is a non-profit national body whose aim is to help improve management standards in the United Kingdom and elsewhere, and was deliberately planned to be an independent neutral, expert body. Many trade, professional and other organisations have commissioned The Centre to establish comparisons, which have now been carried out in more than 100 separate fields of activity, including professional services, local authorities and charities, as well as industry.

Interfirm comparison is an activity designed to show management how the operations of their organisation compare with those of others; where there may be areas of strength and weakness; and what the most fruitful lines of action for improvement may be. Management can use interfirm comparisons both as diagnostic tools and as a means of target setting.

Having decided that the interfirm comparison technique looked interesting, the BLRDD commissioned a feasibility study in 1977. The feasibility report was discussed by a large group of librarians, who expressed interest in the technique and some scepticism as to its use in libraries, but recommended that further work should be undertaken on a phased basis. Between 1978 and 1981 the Centre was commissioned by BLRDD to: design a comparison for public libraries and carry out a field trial and pilot comparison; prepare a revised public library comparison; design a comparison for academic libraries and carry out a test comparison.

Session 2

RETRIEVAL TECHNIQUES

Automatic indexing and classification for mechanised information retrieval

T.M. Aitchison and P. Harding

*INSPEC,
Hitчин*

INTRODUCTION

The project on automatic indexing and classification for mechanised retrieval, which we shall briefly discuss today, is the latest in one of the lines of research and development in progress at INSPEC. Some of the work in this development was funded by INSPEC but most of it, like this project, was supported by grants from the Research and Development Department of British Library.

Since 1971 each item in the INSPEC Database has been indexed by three main indexing elements. First, there is the classification, which divides the database into a comparatively small number of sections (about 2,800): it is used for arranging the entries in the abstracts journals. Second, there are the thesaurus-based controlled subject headings which divide the database into a larger number of parts (currently about 5,200) and are used in the six-monthly and four-yearly indexes to the abstracts journals.

The third indexing element, free indexing, was introduced in 1971 as a result of an INSPEC study, which is known familiarly as the DEVIL (Direct Evaluation of Index Languages) study. This showed the advantage of free indexing in conjunction with a well-developed thesaurus for mechanised retrieval. Free-indexing terms, incidentally, may be defined as words or phrases chosen by the indexer to represent the subject of the document, from the title, abstract or text of the paper or from his own mind. Free indexing gives as fine a distinction in retrieval as the English-language vocabulary will allow.

To facilitate the production and maintenance of the thesaurus and to continue to show its relationship to the classification codes, our Vocabulary File Maintenance software was developed. This allows new thesaurus terms to be added and validated, and reciprocals to be generated automatically.

With the development of our integrated classification and indexing system, we were able to continue to the next stage of our overall plan, which is to make maximum use of the computer in the production of our database and the products derived from it. This next stage was, of course, concerned with an examination of automatic indexing and classification.

Une stratégie d'indexation automatique de documents écrits

Professor Richard Bouché

*Université Claude Bernard,
Lyon*

RÉSUMÉ:

Le système d'indexation automatique qui est présenté a été conçu par le groupe SYDO. Il utilise un modèle linguistique défini à partir de l'observation d'une collection homogène représentative des documents. Le système d'indexation consiste à découper les textes en unités dont on calcule la pertinence. Au départ la stratégie adoptée consistait à prendre le mot comme unité. Une réflexion de fond sur la nature linguistique des descripteurs montre qu'il est indispensable de procéder à une analyse automatique plus poussée de la surface linguistique des textes et que les unités à considérer sont des syntagmes.

Le groupe SYDO rassemble plusieurs équipes universitaires et des centres de documentation sur le thème du traitement automatique des langues et son application aux systèmes d'information spécialisée. Il comprend:

- Le laboratoire d'informatique documentaire de l'université Claude Bernard à LYON - France.
- Le centre de recherche en linguistique et sémiologie de l'université LYON 2 - France.
- Le centre de recherche en informatique pour les sciences sociales de l'université des Sciences Sociales de GRENOBLE - France.
- Le département de linguistique française de l'université de FRIBOURG-Suisse.
- Le centre de documentation CANCERNET à VILLEJUIF - France.
- Le centre de documentation MERLIN-GERIN à GRENOBLE - France.

1 LE POINT DE DEPART DU MODELE:

Un système d'indexation automatique ne peut se concevoir que sous l'hypothèse suivant: les unités du discours utilisées pour décrire le document doivent obligatoirement en faire partie. Le procédé consistera donc à découper le document en unités à définir, et à sélectionner parmi elles celles qui sont les plus pertinentes pour la description du contenu. L'analogie avec l'indexation manuelle, qui utilise une proportion assez

Automatic indexing: a summary

Professor Gerard Salton

Department of Computer Science,
Cornell University,
New York

1. THE INDEXING ENVIRONMENT

In information retrieval, the stored documents and records are normally identified by sets of terms or keywords that are collectively used to represent the document content. The task of assigning the terms to the individual documents is known as *indexing*. The indexing task is obviously crucial for retrieval because failures in the indexing policy immediately lead to retrieval failures. Indeed if the indexing is insufficiently exhaustive - that is, if the chosen index terms do not properly reflect all the subject areas covered by a given document - it may not be possible to retrieve a document when it is needed. On the other hand, when the assigned terms are too broad and insufficiently specific, it may not be possible to reject a document that is clearly extraneous. Retrieval performance is often measured by the ability of the system to retrieve the items wanted by the users (the *recall* factor) and at the same time to reject the extraneous items that are not wanted (the *precision* factor). A highly exhaustive indexing which uses reasonably specific terms for document content representation is believed to lead to high recall as well as high precision.

At the present time two principal indexing strategies are used in operational retrieval environments:

- a) In most situations, the indexing is performed manually by trained indexers, or subject experts, who assign to each document terms that may be freely chosen or may be taken from a controlled list of acceptable terms. Typically between 5 and 15 distinct terms are then assigned to each item for content representation.
- b) In some systems, an automatic so-called *full text indexing* system is used where all the words (except for a few common function words) included in a document, or document excerpt, are collectively assigned as index terms.

The quality of the manual indexing is dependent on the experience and background of the individual indexers. The full text indexing system is not subject to the same variability; however, the assumption that each text word is equally important for content representation is subject to question.

Techniken für den Zugriff auf numerische Daten in alphanumerischen Referenz-Datenbasen

Dieter König

Bundesanstalt für Strassenwesen,
Cologne

1. EINFÜHRUNG

Wenn sich ein Wissenschaftler, ein Manager aus Industrie, Handel oder Verwaltung mit einer neuen Fragestellung auseinandersetzen hat, kann er sich heute an einen Dokumentationsdienst wenden, der ihm entsprechende Veröffentlichungen und Forschungsberichte unter Angabe von Titeln, bibliographischen Daten und Kurzreferaten nachweist. Zur genaueren Information ist der Interessent gehalten, die gewünschten Angaben den Volltexten selbst zu entnehmen. Eine alternative Möglichkeit der Informationsbeschaffung bestünde darin, sich an eine Datenbank zu wenden, die bereits die gewünschten Daten direkt enthält.

Für viele Entscheidungen und Planungen werden Aussagen in Form von tabellarischen Übersichten benötigt, um z.B. Angaben über Verhaltensweisen, Trends, Materialeigenschaften oder Konstruktionsdaten zu erhalten. Das Bedürfnis hierfür ist groß.

Entsprechende Datenbanken stehen jedoch oft nicht zur Verfügung oder enthalten keine Angaben zur gewünschten Fragestellung, da beim Entwurf der Datenbank niemand an gerade diese Fragestellung gedacht hat. In der Literatur sind die gesuchten tabellarischen Übersichten aber meist vorhanden. Die Beschaffung und das Studium der entsprechenden Volltexte ist jedoch zeitraubend und oft schwierig.

2. VON LITERATUR-DATENBASEN zu FAKTEN-DATENBASEN

Als Ausweg bietet sich für Dokumentationsdienste an, künftig nicht nur wie bisher Veröffentlichungen und Forschungsberichte mit bibliographischen Angaben, Kurzreferaten, Sachgebieten und Deskriptoren zu speichern sowie nachzuweisen, sondern auch *tabellarische Übersichten mit Fakten* von allgemeinem Interesse, die in der dokumentarisch erfaßten Literatur vorkommen, zu speichern und bei Bedarf nachzuweisen.

Ein solches Vorhaben soll numerische oder nichtnumerische Datenbanken, wie sie heute systematisch auf vielen Gebieten angelegt werden, nicht überflüssig machen. Datenbanken werden gemäß den gewünschten Zielvorstellungen entworfen, wobei die einzelnen Felder für die gesamte Datenbank gemäß der Aufgabenstellung einheitlich festgelegt und systematisch gefüllt werden. Eine Fortschreibung erfolgt in der gleichen Struktur wie die Erstanlage der Datenbank und in der Regel kontinuierlich.

Session 3

INFORMATION TRANSFER AND DELIVERY

New technology and R & D information: an international comparison of trends

Professor A.J. Meadows, A. Singleton
and M. van der Lem

*Primary Communications Research Centre,
University of Leicester*

INTRODUCTION

Most prognoses for the 1980s predict that new information technology will rapidly expand and diversify. Countries in Western Europe and North America are therefore planning for significant changes in the communication of information during this period. However, planning tends primarily to be on a national basis, so that international differences are not precluded. For R&D communication, which is particularly dependent on the international exchange of information, national divergences in the development of the new technology might have an adverse impact in two ways. Firstly, if the rate of introduction of information technology differs markedly between countries; secondly, if the same information is handled by different types of device in different countries. The aim of this project is to see whether any such divergences are likely to occur during the 1980s.

Forecasting the effect of new technology on R&D communication has major difficulties of its own. One special reason for this is that the technical innovations sometimes arise within the R&D environment, whereas, at others, the application to R&D information represents a spin-off from a totally different area of communication. For example, the growth of computer networks for information retrieval has mainly derived from R&D needs. But the development of word-processing systems has been aimed predominantly at the office: their growing usage for R&D purposes represents a spin-off from this. In yet other instances, R&D communication may involve the further development of technology originally intended for a different audience. Thus, videodiscs are currently being developed for entertainment and education: but their further application to the storage of large quantities of alphanumeric information is mainly of concern for R&D purposes. So, in examining the future of R&D communication, the forecaster must range far beyond the needs of the R&D community itself. Some compensation occurs in the present project, because it relies on a differential analysis. There are reasons for supposing that forecasts of differences in the development of innovations, especially as between countries, are somewhat more reliable than (say) forecasts of the rate of development of a specific innovation within a single country.

Planning the future: computer-based modelling and the dynamics of information systems

Professor B.C. Vickery and Dr. R.G. Heseltine

University College, London

The use of computer-based modelling techniques as an aid to research and policy-making is increasingly common, not only in commercial organisations but in governmental and academic settings as well. At University College London we have been trying to apply one particular methodology, known as System Dynamics, to problems in the library and information field. In this paper we want mainly to talk about the lessons which can be drawn from work of this kind, and about its general relevance, rather than focussing narrowly on the methodology itself. It seems appropriate, however, to begin by giving some indication of what the technique entails.

System Dynamics modelling has two main purposes. In the first place, it is used to clarify the existing structure and operation of a system. How is the system organised, what influences are at work in it, and how is its behaviour determined? Secondly, it is used to suggest ways of *improving* that behaviour, either through changes in the system's structure or through changes in the policies used to control the system. System Dynamics is not, therefore, a forecasting tool in the conventional sense. It is not a means of predicting the future. Although it is intimately concerned with behaviour over time, it is primarily an aid to system-design and policy-making.

Very briefly, a System Dynamics model has two main forms: a *diagrammatic* form and a *mathematical* form. So far as the first is concerned, we use a particular kind of diagram known as an influence diagram. Drawing this diagram is the basic means of specifying the existing structure of a system and the influences at work in it. The technique itself is very straightforward. Following certain standard procedures, the diagram is developed by incorporating all the variables which are present in the system, and connecting these by arrows which indicate the direction of influence or causation. Some of these causative links are obvious, especially if they involve physical flows of people or materials. Other links are less obvious, particularly those concerned with the policies used to control the system, which are often not explicit. This lack of explicitness is rather characteristic of our conventional understanding of systems. The influence diagram, by contrast, is intended to be a very explicit representation. To achieve this, to bring out the essential processes at work in a system, the technique needs to be applied in a very rigorous and disciplined way. But if this can be done, then the influence diagram is capable of functioning as a highly effective form of system description.

Informed enterprising

R.H. da Silva

*NOBIN,
The Hague*

INTRODUCTION

It would be wrong to assume that interest in the information problems of Small and Medium sized Industries (SMI) in the Netherlands is new. It has long been recognized that when it comes to the use of external documented knowledge and experience, SMI may be disadvantaged by comparison with large firms. For a variety of reasons access to relevant material held by the main public information services, libraries and documentation centres has been problematical. This problem has been highlighted by information professionals and during the seventies it became a key issue. The gap between the creation of scientific information and its subsequent application by managers is a much studied phenomenon. Various publications have revealed the complexity of the problem (Havelock 1966). In 1976 it was decided to focus on the information needs of SMI in the Netherlands. By the beginning of 1977 the Ministry of Economic Affairs had commissioned NOBIN to perform an exploratory study as part of a project for the advancement of information facilities for small industry (PBIB).

The aims of the research were:

- to gain a wider familiarity with the phenomena governing the SMI information market including the views of actual *and* potential users;
- to gather knowledge about practical ways of eliminating barriers to the transfer of external recorded information to SMI;
- to establish feedback routines in order to bring about changes at the suppliers side when necessary.

It was intended that the study would establish priorities for further action. By the end of 1980 this preliminary phase was concluded. It was discovered that within the economic system as a whole, relevant information as provided by intermediaries and in principle available to smaller firms hardly ever played a role of any importance. Secondly, it was posited that in order to achieve a better understanding of the potential of external information, its significance should be measured in circumstances where firms were forced to consider major changes in their production processes in order to cope with changes in the environment. Once

Technical aspects of document delivery using broad band techniques and satellites

Jørgen A. Richter,

*CENTERNET/NEUCC,
Technical University of Denmark,
Lyngby*

FTP - THE BASIC CONCEPT

Whenever information is stored in digital form, automatic retrieval and delivery services seem easy to implement - if not in practice then in theory. The new generation of all-digital transmission systems, where it has become clear that fibre-optics are technically and economically superior to analogue, coaxial techniques, opens up new possibilities. The new architecture of mass storage computer systems where fast electronics and autonomous I/O processors can relieve the central cpu makes bulk data transfer feasible.

In theory the overall system performance should today meet the end-users' anticipated or real-time requirements. In practice the following calculations, shown in Figure 1, illustrate where the major problems are to be found.

Assume three documents are to be delivered at a remote location. The first being a EURIM 5 paper. It contains typically 56 lines/page with 64 characters/line and occupies 6 pages (2500 words). The transfer protocol suppresses redundant information and will see a file of size 150 Kbit. The second is a LANDSAT scene. It is a four band image covering 2300 x 3200 pixels. The total amount of data is 235 Mbit. However, new satellites will soon be in orbit offering images in 7 bands covering 6000 x 6000 pixels. The total amount of data here is 2 Gbit. Please notice that the phrase 'interactive analysis' is used in these applications.

| Line Speed | EURIM 5 | LANDSAT | NEXT LANDSAT |
|--------------|-----------|----------|--------------|
| 1200 bit/sec | 2 min | 54 h | 462 h |
| 9600 bit/sec | 15,6 sec | 6,7 h | 58 h |
| 56 Kbit/sec | 2,7 sec | 1,2 h | 10 h |
| 1 Mbit/sec | 150 msec | 3,9 min | 33 min |
| 6,4 Mbit/sec | 23,4 msec | 36,7 sec | 5,2 min |
| 34 Mbit/sec | 4,1 msec | 7 sec | 59 sec |
| 140 Mbit/sec | 1 msec | 1,7 sec | 14,3 sec |

Figure 1: One way transmission delay

Video discs in information technology: actual and possible applications

Dr. Fredrik Backlund

*Swedish Delegation for Scientific and
Technical Information,
Stockholm*

BACKGROUND

The video disc is a slim, rapidly spinning circle of plastic which brings sound and colour pictures to the TV-screen. Its appeal is a heady compound of science *fact* and science *fiction*. Put a disc on your player and you have made your own choice of TV-programme. Add a stereo audio track and you can enjoy the experience of a musical performance that combines richness of sound and sight. Use the disc as a data storage device, and you have several hundred - and in the very near future thousands - of books on a single record. Hook the player up to a computer and you can select exactly the picture frame you want.

At this point of the narrative I have to assure you that I am not a seller of video discs and I won't become one. Stripped of all the often repeated rhetoric surrounding it, the video disc seems to offer three features not found in any other audiovisual medium.

First: the raw material out of which discs are manufactured is far less expensive than the film or magnetic tape used to display images or sound.

Second: since the information density is some 5 to 10 thousand times denser than microfilm, the finished disc is so compact that it can be carried, mailed or stored far more easily than can other forms.

Finally: an optical video disc lends itself to rapid and precise retrieval of a single frame of information which can be displayed on a screen for as long as the user wants.

A SWEDISH APPLICATION

Video disc systems offer a unique possibility of storing and retrieving information consisting of *text* and *pictures*. Databanks covering, for example, patents and technical descriptions can be set up and, in fact, are already on the market.

The neglected resource

Margaret Slater

*Aslib Research and Consultancy,
London*

That deliberately enigmatic and hopefully curiosity-provoking title (the neglected resource) refers primarily to the unenviable and marginal situations of many library-information services within parent firms. It could equally describe desultory application of research effort to an apparent problem in our field. Hundreds of user studies must have been executed by now. Comparatively few projects have explored non-usage.

Research into different topics sometimes uncovers evidence of sub-optimal use. Recently, I undertook work on ratios of special library-information staff to users - actual and potential. The aim of this exercise was the provision of background information for manpower planning, at national and corporate levels. A subject, one might think, somewhat removed from non-use? Yet during this survey of 655 organisations revelations occurred, relevant to unsatisfactory usage and its possible causation.

Respondents from the industrial-commercial sector defined the size of their potential user market, on average, as 40% of the total workforce of their organisation. Under half this group (46%) were actual users. The term 'usage' also covered various degrees of frequency, success and relevance to professional life. According to library-information managers, much usage is minimal, superficial, sub-optimal. Not only under-use, but also pointless over-use, misuse and sheer abuse occur. Why should this happen? Evidence from this survey indicated various possible reasons. Perhaps the target population is not always correctly or sufficiently defined? Only 41% of respondents had a tight, clearly outlined definition of their potential user population.

Causes of non-use may be more subtle and intrinsic to the user's personality or professional role. Such aspects of neglect are harder to investigate. The non-user becomes remarkably elusive - perhaps one simple pragmatic reason why research has concentrated on the user instead? Investigating the non-user, one feels rather like those anthropologists who trailed the retreating Krena-Krori Indians through the Amazonian rain forests. Hoping for contact with this lost tribe, all the pursuers ever found were ashes, abandoned cooking pots, or a fugitive glimpse of a distant shadow flitting between trees. What I am providing here is analogous: some insights, painfully acquired, from small samples of sometimes uncooperative potential users.

Utilisateur, producteur, serveur: un ménage à trois pour le meilleur ou pour le pire?

Bernard Marx

*Direction des Bibliothèques, des Musées et
de l'information scientifique et technique,
Ministère de l'Éducation Nationale,
Paris*

1. INTRODUCTION

L'image du ménage à trois pour caractériser l'ensemble du processus de la recherche documentaire en conversationnel est bien sûr une vision simplifiée de la situation réelle. Il y a au moins un quatrième intervenant, le producteur public ou privé du réseau de télétransmission; sans compter l'intervention des Pouvoirs Publics aux différents niveaux du transfert de l'information scientifique et technique: production, diffusion, transmission et utilisation.

Quelles sont les caractéristiques du meilleur fonctionnement des relations producteur-serveur-utilisateur?-des informations validées fournies régulièrement et rapidement, des recherches effectuées par un logiciel adapté à la nature de l'information, une coopération des organismes concernés pour des documents d'aides à la recherche complets et pour l'organisation efficace des formations.

Sans faire la description du pire fonctionnement et sans réaliser une revue exhaustive des différents services, il convient de situer quelques exemples de difficultés rencontrées par l'utilisateur parmi les services proposés:

- conditions d'utilisation
- documents d'aide à la recherche
- formation des utilisateurs,

de les analyser par rapport aux responsabilités spécifiques ou communes des principaux partenaires (producteur, serveur, utilisateur) et de proposer l'amélioration des services actuels.

2. CONDITIONS D'UTILISATION

2.1 EVOLUTION DE L'INDEXATION D'UNE BASE

C'est un élément de la responsabilité directe du producteur de la base. Les thésaurus imprimés et les listes de descripteurs n'indiquent pas toujours les anciens termes d'indexation. Lorsque le thésaurus de la base n'est pas en ligne, l'utilisateur rencontre une difficulté pour utiliser les différentes formes

The social context of information transfer: an analysis of facilitating and inhibiting factors

Professor Mark van de Vall and Nelleke Schoemaker

*Institute for Social Policy Research,
University of Leyden*

1. INTRODUCTION

We begin with a rather trivial statement: the road between the production of information and the utilization of that information is long and complicated. Even extreme simplification results in five different areas of activity, each with two fields of specialization. Each of the specializations is supported by its own package of knowledge, skills and instrumentation:

Model (1): Fields of Activity and Specialization in Utilizing Information

a. Research b. Technology c. Assessment d. Production e. Merchandising

| | | | | | | | | |
|---------|---|-------------|---|-------------|---|------------|---|------------|
| Basic | → | Transforma- | → | Testing | → | Retooling | → | Marketing |
| and | | mation and | | and Cost- | | and | | and |
| Applied | | Development | | Benefitting | | Production | | Servicing. |

As the information moves from area a (research) to area e (merchandising), several trends occur in the process of utilization: (i) the range of alternatives becomes narrower; (ii) emphasis upon the practical value of the information increases, (iii) the cost-structure becomes more differentiated; (iv) investments show a tendency to increase in size. A well-known illustration of (iv) is that it costs about \$30,000 to develop a *workable* jet-engine but between \$500,000,000 and \$100,000,000 to produce a *marketable* jet-engine (Klein 1958, quoted in OECD 1972: 20).

2. THE STRUCTURE OF INFORMATION TRANSFER:

In each of the five fields we find utilization of information, facilitated by knowledge transfer. Although each with its own specific content, the basic structure of those utilization processes is roughly identical, resulting in the following model:

Changes in occupation and profession in information work: the impact of new communication technologies on information work.

Dr. Thomas Seeger

*Lehrinstitut für Documentation,
Frankfurt am Main*

1. INTRODUCTION

The title of my paper is broad, so to avoid generalisations of a gross kind I shall restrict myself to a small number of aspects of this topic; aspects which relate to developments in the Federal Republic of Germany within the past decade. This will be followed by a short assessment of the effects of new technologies on the information profession. All statements are strictly oriented to the situation in Germany, in recognition of the fact that discussions of this kind are different and probably more advanced in other countries.

I am not going to add to the thousands of statements on likely future developments. Instead, I would like to give you some background information on trends in information policy in Germany and summarize the quality of prognostic statements produced and made in several forecasting studies relating to technological developments for Germany. These predictions on future developments in the field of information work will be compared with recent trends and developments.

2. THE TRADITION OF INFORMATION AND DOCUMENTATION

Information and documentation in Germany developed quite independently from library and archive work. Different from other countries, the various facets of information work had very little in common, because each sub-field insisted on its own traditions and historical merits - institutional egoism playing a major role.

This separatism worked as long:

- as there was only little, or at least steady, change in the institutions of information work
- as changes could be easily anticipated and accomodated by organizational structures
- as the core of intellectual information handling procedures (indexing, cataloguing etc.) were not severely affected.

Session 4

MULTILINGUAL AND INTERNATIONAL SYSTEMS

Multilingual information systems backed up by machine aided partial translation can provide the information needed for the 80's

Professor Peter P. Canisius,

*Bundesanstalt für Strassenwesen,
Cologne*

1. INTRODUCTION

Information in future will have to look different. In the past, we learned how to *handle* information. The present must be used to shape it into a form which can be *accepted by any user* - especially those outside the scientific and technical environment. Only if information is widely distributed and ultimately accepted is it true 'information for development'. And this is relevant for developing and industrialized countries alike.

I have said on other occasions that I think information should aim at serving the three E's, namely

- * Experts (Science and Technology)
- * Economy (Industry)
- and
- * Everyone.

So far, I think, we have reached less than 50 per cent of these end user groups. We seem to provide much for the S & T sector, little for industry, and almost nothing for everyone. And this particularly because information systems lack the quality of being 'multilingual' in the sense that they fail to get across or speak and appeal to the aforementioned groups.

2. FROM MULTILINGUAL TO 'POLYGLOT'

Therefore, I would like to go *beyond the meaning of the word multilingual* into an area not only referring to several languages, but where information is 'understandable and acceptable' in a broader sense. There are of course a number of multilingual information systems, in the narrow sense, in existence. One or two of them I shall briefly refer to later. But since the term 'multilingual', in the past, was used extensively for data bases which did not really '*speak*' to a large user community, I would like to call the systems I am aiming at '*polyglot*'. While

Machine translation's big ten years: 1980-1990

Loll Rolling

*Commission of the European Communities,
Luxembourg*

Fully automatic translation is impossible. In other words, it is impossible for a machine to do exactly what a human translator does.

One may also say that it is impossible for human beings to fly - nevertheless, if necessary we can be in New York this evening without having to spread our wings, thanks to Boeing who have designed fine aircraft and a pilot who has been properly trained. But since the 747 cannot land on the roof of our hotel, we still have to take a taxi to get us to our destination, supplementing the 'automatic flight' by a more traditional activity.

Machine translation, then, which is not entirely automatic, has nevertheless made its breakthrough. While its initial appearance, during the Fifties, may have been due to the curiosity of some researchers who wanted to see what else one could make the computer, this new toy, do, nowadays it is imposed on us by the pressures of the market.

THE TRANSLATION MARKET

The study of the translation market, completed in 1981 at the European Commission's request, shows that in one year 150 million pages were translated by 175,000 translators, professional or amateur, that their turnover was of the order of 18 thousand million FF, and that these figures are increasing by 9 to 10% per annum

And that 50 million pages, translation of which would have been desirable, could not be translated owing to a lack of qualified translators.

How is this market made up? Who are the users? This is shown in the following table (fig. 1).

If the international institutions, from the European Community to the United Nations, have the largest translation services and the most qualified translators, the major market is nevertheless the industrial one. Hundreds of thousands of pages of publicity brochures and maintenance manuals have to be translated before an industrial firm can hope to sell a pharmaceutical product, a domestic iron, or a jet fighter across the linguistic borders.

TITUS IV: système de traduction automatique et simultanée en quatre langues

J.M. Ducrot

*Institut Textile de France,
Boulogne-Billancourt*

INTRODUCTION

TITUS IV est une méthode de traduction automatique qui a été spécialement conçue pour le traitement multilingue de bases de données dans les domaines scientifiques et techniques. Son utilisation permet de constituer une collection documentaire dans laquelle les informations sont introduites soit en allemand, soit en anglais, soit en français, soit en espagnol.

Ainsi les informations stockées peuvent-elles être récupérées et traduites automatiquement dans l'une des 4 langues du système.

Afin d'éviter les difficultés inhérentes à la complexité des langues et à leurs ambiguïtés, le système repose sur une méthode particulière de traduction automatique dite 'à syntaxe contrôlée' (controlled syntax), qui autorise des formes d'expressions courantes, naturelles, grammaticalement classiques, mais dont la structure doit obligatoirement satisfaire à des critères linguistiques restreints et prédéterminés.

Le langage à syntaxe contrôlée utilisé pour la méthode TITUS IV est constitué de 2 éléments de base. Ainsi pour chaque langue, et en parfaite correspondance entre elles, on a déterminé (Fig. 1):

- a) *un sous-ensemble du vocabulaire*, qui contient toutes les notions spécialisées d'un domaine et une partie du corpus du vocabulaire basique de la langue considérée. Bien entendu, le vocabulaire spécialisé est variable selon le domaine d'application, alors que le vocabulaire basique est pratiquement commun à tous les domaines.
- b) *un sous-ensemble de toutes les règles syntaxiques* régissant une langue. Bien qu'en nombre réduit, les règles syntaxiques acceptées par le système TITUS IV sont tout à fait naturelles et les plus classiques dans chaque langue.

An alternative strategy for steady growth towards high quality translation networks

A.P.M. Witkam

*BSO/Automation Technology b.v.,
Utrecht*

1. INTRODUCTION

This paper has not the pretension to present a proven, ready-made or easy solution. Instead, amidst the persistent search for breakthroughs in the difficult field of computerised translation, it intends to contribute to the scientific debate by pointing out a rather new and largely unexplored direction.

In machine translation (MT), but also in data-base enquiry, advanced word processing and natural language programming systems, the analysis of the source text is the crucial process, responsible for parsing and disambiguation.

For this purpose, conventional MT systems initially relied on only grammar and dictionary, the grammar being limited to morphology and syntax. Later, semantic elements were added to existing systems (SYSTRAN) or included in the conception of new ones (EUROTRA). Besides, surprising results were obtained with a computerised disambiguation dictionary for English (5).

However, these developments have not lead to full-automatic and high-quality translation of a general range of natural language texts. The problem is that our computer systems, unlike men, do not generally have knowledge of the world, the context or the situation, to resolve all hidden source text ambiguities.

In order to get along under the present state-of-the-art, some MT systems (notably SYSTRAN) rely on human post-editing to enhance their output quality. Other systems operate successfully by virtue of specialisation on a narrow subject (METEO) or by constraints on the input sentence patterns (TITUS; MCE) (6). Still others (EUROTRA) hope and wait for the future results of artificial intelligence, to attain perfection.

The strategy advocated by BSO/Automation Technology after several years of study in this area (11), exists in a combination of automatic and interactive disambiguation during the source text analysis phase. At text entry on a specially equipped word-processor, the system first tries to parse and disambiguate a sentence as far as possible. Any residual ambiguities (structural as well as lexical) are then presented to the

Closing Session

Closing address

Professor J. Meyriat

*Président, Association Française des
Documentalistes et des Bibliothécaires
Spécialisés*

Mr Chairman, Ladies and Gentlemen.

EURIM is now nine years of age. It has gone through a number of European countries and it is coming back to its cradle, which is why I have the honour of making the closing address. Last year you entrusted this task not to a researcher but to the President of a national association for information transfer, and it is as a French professional that I want to speak to you this morning. The first thing I would like to say is that the success of this Conference is due to a very large extent to the fact that it is based on interaction between professionals and the researchers of the universities and industry, and so I would like to stress the need for and the absolute utility of this interaction. Of course each of us has to have a main point of interest, a main pole of interest, but in our field of information this interaction, this constant exchange is basic. Research cannot be dissociated from practice; from previous EURIM proceedings I have found that thirteen papers were presented by French speakers, and the very clear majority were managers or people responsible for information services who reported on their experience. This I think is an illustration of the very close link between professional activity and research. There is no doubt that a researcher should never stray from the realities of professional life, and I would like to stress the additional requirement, that professional practice must always be vivified, animated, by research, fed by research, be this basic research, applied research, or development studies, as a certain number of people call their work. If this is not done then the profession becomes routine and sterile, and the countries that do not bend to this effort become dependent on concepts, methods, techniques which are exogenous, and this of course on a long-term basis is a risk.

This leads me to my second point, which relates to the usefulness, which I think is very great indeed, of a forum for research in Europe, and I would like to stress the European characteristic of EURIM even though this is already implicit in our terms of reference. Certainly the problems arising out of the management of information systems are universal, world-wide problems, and research relating thereto knows no borders; and in any event it would be puerile to make reflections independent of what is being done in the other major regions of the world - *inter alia*, of course, being the U.S., where the advances in a number of fields is undeniable; a