FID/CR News 30

The ISCCR Conference

The 5th International Study Conference on Classification Research (ISCCR'91) was held in Toronto, Canada, June 24-28, 1991. In spite of a war in the Persian Gulf and a world economic recession, the conference was attended by 55 participants from 10 countries, including representatives from six classification research organizations from North America and Europe. 40 papers were presented. Two papers prepared for the conference but not presented have been included in the proceedings.

The theme of ISCCR'91 was "Classification Research for Knowledge Representation and Organization". Research was broadly interpreted to include both theoretical and empirical research and participants included educators and practitioners as well as researchers. The papers fell into three general categories, defined in the original call for papers as: general principles and policies, structure and logic in classification, and empirical investigation.

Classification research itself was the subject of two papers in the first plenary session of the conference. In her keynote address, "Classification: Prospects, Problems and Possibilities", Elaine SVENONIUS identified five forces - philosophy, technology, sociology, economics and politics - which direct and influence classification. Selecting from this vast topic, Svenonius examined one of these forces - philosophy - in detail. Her primary focus was the 20th century philosophies of logical positivism, linguistic analysis and systems analysis with reference to modern classification research. In conclusion, assuming the continued influence of these forces, she speculated on the future of classification research. Svenonius' paper established a framework for the remainder of the conference. As deliberations progressed, it was evident that virtually all of the research presented at the conference has been influenced in one way or another by one or more of the five forces. A second paper which examined the broad concept of research was Robert FUGMANN's "Illusory Goals in Information Science Research". Fugmann recognized that, by its very nature, information science involves some processes whose course and goals cannot be successfully determined in advance of action to be taken, but require a posteriori decisions. Examples of such indeterminate processes were identified as: prior determination of users' needs, and the expression of general concepts in natural language. The viability of search algorithms being applied successfully to individual concepts was recognized but is was pointed out that recognition of meaning in general concepts in natural language presents difficulties because of paraphrasing and the unpredictability of the wording of such concepts. In summary, Fugmann warned that algorithmization in information science should be applied only to those processes which are determinate. At the same time, there is need for intensification of intellectual approaches to information science research in those areas which cannot be determined a priori.

Two papers dealt with the notion of fundamental categories and the origins of modern classification systems. Francis MIKSA's "The Concept of the Universe of Knowledge and the Purpose of LIS Classification" concluded that even though a classification system may begin as an independent universe of knowledge, when applied and developed in conjunction with documents such classification systems ultimately grow and develop on the basis of literary warrant. In a more precisely directed paper Jean PERREAULT's "Essay on the Prehistory of General Categories (II)" described an analysis of the contributions of G.W.Leibniz and Conrad Gesner to the foundations of modern document classification.

As might be expected, the well known general classification systems were well represented in the conference deliberations. Two papers were presented on DDC and three papers each on UDC, LCC and the Colon Classification. In all cases the influence of both technology and economics was evident. Much of the research on these systems is related to improving their effectiveness, either through improvements in vocabulary and structure, or through the development of computerized systems for the manipulation of schedules. The application of computers to DDC was the subject of two papers. Julianne BEALLS's presentation on "Editing the Dewey Decimal Classification Online: The Evolution of the DDC Database and Editorial Support System" described the DDC Division's online system currently being used to edit and produce the successive editions of DDC. For the future, this system ensures more accurate editing and speedier revision resulting in an increasingly more up-to-date system. Diane VIZINE-GOETZ's paper on "The Dewey Decimal Classification as an Online Classification Tool" focussed on a different aspect of computerization. Her research centred on the database design, user interface and classification record enhancements of a prototype online system which could become an online tool for classifying documents. This research, being carried out at OCLC, utilizes the Decimal Classification Divisions's Editorial Support System. Continuing with the theme of computerization, Rebecca GUENTHER's presentation on "The USMARC Format for Classification Data: Development and Implementation" described the machine-readable format, developed and presently being tested by the Library of Congress. The format will enable LC to convert the LCC schedules for machine manipulation by classifiers and online catalogue users. The format has been developed to accomodate both DDC and LCC and it is anticipated that it will be adaptable to other systems such as UDC, BBK and the National Library of Medicine Classification.

Mary MICCO's "Suggestions for Automating the Library of Congress Classification Schedules" recommends that the LC move towards a hierarchical notation for easier manipulation of the LCC schedules in computerized systems. Also concerned with more efficient handling of LCC online was Karen Markey DRABEN-STOTT's paper on "Shelflisting Using Expert Systems". Drabenstott, Leslie RIESTER and Bonnie DEDE have built a prototype expert system using a section of the QA Class in LCC. Their research demonstrates that such a system has promise. However, to be effective in practice, it would need to be integrated with other components of an online cataloguing system, and the LCC schedules and the book numbering system would need to be in machine-readable form.

While DDC and LCC have many inherent weaknesses and imperfections, they are generally assumed to be reasonable healthy intellectually, financially and administratively. This is much less true of UDC. Two papers on UDC focussed on efforts to improve the maintenance and revision process and to restructure the schedules. As chair of the UDC Management Board, Alan GILCHRIST, in his paper "UDC: the 1990's and Beyond", outlined the problems facing UDC and described some of the proposed solutions. Changes in the administrative and financial structure, as well as in the editing and revision of schedules, are already taking place. Since the ISCCR conference, FID has moved one step closer to a possible new era for UDC, and a public announcement of the plans for the future of UDC is imminent. Following from the proposed changes, Nancy WILLIAMSON's paper on "Restructuring UDC: Problems and Possibilities", responded to a recommendation from the Task Force on UDC System Develop - ment that UDC be restructured into a fully-faceted system accompanied by a thesaurus. Her paper described very preliminary research and recommended a pilot study to test the viability of using the facet structure of the *Bliss Bibliographic Classification* as a guide to the internal reordering of topics in the UDC schedules. A final paper on UDC, by Blake WILSON, "An Evaluation of the Systematic Botany Schedule of the Universal Decimal Classification (English Full Edition, 1979)", described a comparative study of the Engler-Diels taxinomic system and the relevant UDC schedules. Recommendations for improvement of the UDC schedules were made.

Because facet analysis is now clearly recognized as a fundamental principle basic to various kinds of information systems, Ranganathan's facet principles and chain indexing were discussed in numerous contexts throughout the conference. However, three papers focussed directly on the Colon Classification. Shabahat HUSAIN's presentation on "Notational Techniques for the Accomodation of Subjects in Colon Classification 7: Theoretical Possibility vis-a-vis Practical Need" was concerned with hospitality of the CC notation and its ability to accomodate new subjects. Two papers on chain

indexing were technology driven. A theoretical paper by William STILES examined possibilities for automating the chains, while Elaine SVENONIUS, Songqiao LIU and Bhagi SUBRAMANYAM conducted empirical research on the "Automation of Chain Indexing". A prototype system developed at the University of California, Los Angeles, features an automatically generated chain index using DDC. This paper focussed on the algorithms used to create the indexes and the problems encountered.

Other general classification systems were examined in two papers. Margaret ANDERSON posed the question "An Islamic Classification Scheme - is it Necessary?". The work of Ziauddin Sardar in two areas - classification and the unity of the Islamic world - was examined. Discussion centered on the validity of Sardar's arguments for an Islamic classification scheme and analyzed the system he had outlined. In another context, Lei ZENG, at present in Pittsburgh, described "Research and Development of Classification and Thesauri in China". She described and analyzed three lines of research and development - the improvement of schedules for practical use, the exploration of strategies for solving problems in indexing languages, and the establishment of fundamental theories for all methods of knowledge organization.

Thesaurus design and development is a very productive area of research. Five papers were testimony to the growing interest in this area of knowledge representation and organization and the importance of classificatory structure in indexing languages. Two papers focussed on vocabulary. Lorna REES-POTTER addressed the question "How well do Thesauri Serve the Social Sciences?". Her conclusion: not very well. Among the problems singled out were soft vocabulary and the nature of semantic relations in thesaurus design. In this same area Michele HUDON presented a paper on "Term Definitions in Subject Thesauri: the Canadian Literacy Thesaurus Experience". The trust was on changing directions of thesauri from database dependent tools to semantic tools independent of specific collections. Hudon suggested that the objective in thesaurus development should be a more informative tool without increased complexity. Her paper examined ways in which the work of terminologists might solve the problem of complexity. The components of a term bank were compared with those of a thesaurus, addressing the kinds of definitions needed, the use of definitions to replace complex structure, and possible guidelines needed. Three papers dealt with research on thesaurus display. Timothy CRAVEN described research on "Concept Relation Structures and their Graphic Display", outlining two algorithms for the automatic production of graphic displays, which have been tested informally. Hanne ALBRECHTSEN described "PRESS: a Thesaurus-Based Information System for Software Reuse". An ESPRIT project, PRESS (The Practitioner Reuse Support System) features a thesaurus with a semantic network of indexing terms for

browsing and retrieving subject information of reusable software concepts. PRESS also implements a semi-automatic strategy for thesaurus compilation as well as a semi-automatic approach to subject indexing. In the third paper in this group, Raya FIDEL studied the search behaviour on online searchers as a means of shedding light on "Thesaurus Requirements for an Intermediary Expert System" that mediates between online databases and end users. Among the major concerns were searchers' decisions and the selection of search keys. Fidel's criteria for such systems would be provision for advice to end users on how to select search keys, indication of the usefulness of search keys, a display of useful combinations of natural language and descriptors and switching languages for multilingual databases.

Special needs for classification in very diverse fields was recognized in seven of the presentations. A theoretical approach to fiction analysis was described by Clare BEGHTOL in "Toward a Theory of Fiction Analysis for Information Storage and Retrieval". Beghtol's research concerned the primary literature of fiction and the secondary literature about fiction. Using four techniques she focussed on fictional and critical warrants. James TURNER described his research on subject access to moving images in "Structure in Data in the Stockshot Database at the National Film Board of Canada". ISCCR'91 also broke new ground in classification research in the field of information management. Nasser SWAYDAN investigated "Subject Organization in Archives", especially the possibilities of a special classification system for these collections. In the related area of records management, Monica SCOTT and Flavia FONSECA described early research on the handling of office documents in "Methodology for Functional Appraisal and Development of a Functional Thesaurus". This research arose from a decision by the World Bank to revise its methods for appraisal of its records to focus on the function, rather than the content of records series. A thesaurus of functional vocabulary is required to carry out the appraisal. The Scott/Fonseca paper examined differences between their concept of a functional thesaurus and the traditional tool. In this same general area, Barbara KWASNIK and Corinne JORGENSEN's paper on "The exploration by Means of Repertory Grids of Semantic Differences among Names of Official Documents" investigated ways people name and classify documents in office situations. Repertory grid techniques were used to discover the constructs used to distinguish a document of one name from a document of a different name and to determine the dimensions along which people decide to assign a name to a document. Closely related was Marlene ROCKMORE's work on "Structuring a Flexible Faceted Thesaurus Record for Corporate Information Retrieval". Rockmore described the construction of thesaurus records for a thesaurus/indexing management system which allows a knowledge engineer to design flexible faceted thesaurus structures which can be very powerful in capturing semantic

categories and disambiguating terminology.

Symptomatic of the current economic situation, only one paper which was concerned with reclassification in libraries. Lynne HOWARTH's "Factors Influencing Policies for the Adoption and Integration of Revisions to Classification Schedules" presented preliminary research from a survey of Canadian public and academic libraries on factors which may influence cataloguing managers' decisions to incorporate classification revisions into their systems.

New directions in classification were exemplified in four papers. Rebecca GREEN's "Insights into Classification for the Cognitive Sciences: Ramifications for Index Languages" was primarily concerned with basic level categories and frames and their relevance to indexing language development. She observed that implemented together these two devices render indexing languges more flexible and more powerful. Green also presented a second paper, "The Expression of Syntagmatic Relationships in Indexing: Are Frame-Based Index Languages the Answer?" This study looked at three aspects of the topic - the features desirable in a system with syntagmatic relations, a method for developing frame based indexing languages and case studies. Examples were drawn from passages in the *Bible*. Some problems were encountered but a high degree of success had been demonstrated. Also frame-based, Susanne M.HUMPHREY's paper on the "Use and Management of Classification Systems for Knowledge-Based Indexing" described research on access to biomedical information taking place at the National Library of Medicine. A system called MedIndEx (Medical Indexing Expert), an interactive knowledge-based prototype for computer-assisted indexing of the MEDLINE database was described. MedIndEx is designed for domain-independent application and for portability across software platforms. Closely related was research by Crawford REVIE and Godfrey SMART describing "The Construction and Use of Faceted Classification Schema in Technical Domains". Part of the SIMPR (Structured Information Management: Processing and Retrieval) project under the ESPRIT II research initiative, this project was concerned with the indexing and classification of large textual document bases. The paper focussed on the classification itself and the tools used to build it. The practical use of SIMPR was discussed with reference to application in a specific technical domain.

While a wide range of problems in the representation of knowledge were discussed throughout the conference, two papers turned to problem areas in themselves. Alan WILSON's paper on "The Hierarchy of Belief: Ideological Tendentiousness in Universal Classification" described classification as a construction of dynamic and often competing hierarchies based on sociological perception. Using Martin Allor's definition of ideology as power exercised through representation, Wilson used the major classification systems to substantiate claims that government classification systems, such as

LCC and BBK exhibit the greatest degree of ideological information. In contrast, DDC and Colon exhibit ethnocentricity. Another problem, that of compatibility, was discussed in two papers. Linda SMITH's "UNISIST Revisited: Compatibility in the Context of Collaboratories" analyzed past efforts to develop better tools for controlling and converting natural and indexing languages in science and technology. Following from her analysis, Smith identified a need for further research in three areas - the improvement of compatibility between users' language and the language of information systems, the means for providing links among different kinds of information objects, and the extent to which compatibility can be achieved through automatic means as opposed to human efforts. From another perspective, Corentin ROULIN discussed problems in developing "Sub-thesauri as Part of a Metathesaurus". His research derived from experience in managing a multilingual European Education Thesaurus, and concentrated on indexing languages for restricted fields. A functional definition for the term sub-thesaurus was proposed and a review of the elements relationships and subsystems, as well as methods of presentation, constructioon and use of such a tools were described.

Finally, three papers focussed on knowledge representation and organization from particular perspectives, Pauline COCHRANE's paper "Indexing andSearching: the JANUS or Proteus of Information Retrieval?" confronted the question of whether or not the processes of indexing and searching are really so similar that the same tools can be used effectively for both processes. Various information retrieval models were reviewed to show how these processes have changed over time and how such changes may affect the design of classification schemes and thesauri. Online use of such systems in conjunction with computer graphics, hypertext and expert systems may require some adjustment. Comparisons were made in the use of tools by both intermediaries and end-users. In her paper on "The Basis of a New Universal Classification System Seen from a Philosophy of Science Point of View" Ingetraut DAHLBERG looked at the philosophy of science as a basis for a theory for systems of knowledge and the constitution of knowledge fields, as well as the criteria for the identification of such fields. The paper discussed studies already done and explored further possibilities for the use of these theories in the construction of a new universal classification system to provide better organization than is possible with the present general systems. Finally, Eric de GROLIER's paper "Towards a Syndetic Information Retrieval System" provides constructive criticism of past research on switching devices and their problems and offers some possible solutions, taking into account progress in technology, particularly hypertext.

ISCCR'91 concluded with a panel discussioon chaired by Ia McILWAINE with panelists Timothy CRAVEN, Ingetraut DAHLBERG, Bernd FROHMANN, and Lorna REES-POTTER. Each presented their personal views of the conference and exchanged comments with the participants. Following the panel discussion Eric de GROLIER, the only participant in all 5 of the FID/CR Study Conferences, presented a summary and conclusion. He began with a statistical comparison of the 5 conferences, concluding that approximately 50 participants was the optimum number of a conference of this kind. Then de Grolier addressed the question "What is new since 1982? (since the Augsburg Conference). Cited were technological advances in online catalogues, CD-ROM products and hypertext, and the effect of increased use of personal computers. The impact of these advances is still in process and we are still striving to achieve effectiveness. De Grolier cited financial crises and the decline of western economy as causes of the end of "30 glorious years" of information retrieval. New inter-governmental programmes, while promising, have some limitations. To de Grolier the obsolescence of the old systems is becoming more apparent because of online systems and scientific progress which has resulted in new objects and new theories. Also cited were the turmoil in the social sciences, in particular the end of the ideologies, the collapse of established theories, social disorder, and the end of the disciplines. Together these changes could have a profound effect on the future of information systems. At the end of his presentation de Grolier "Proposed Recommendations for Determining Fields in Need of New Studies". He stated that research is needed into:

1   The most efficient use of new technologies;
2   The construction of practical tools for serving as "bridges" between different systems (specialized and general);
3   The means of avoiding ethnocentrism;
4   The integration of classification and terminology research (especially in social/human sciences);
5   The relations of classification with logico-mathematical models and numerical taxonomy; and
6   The restudy of notational systems.

As an addendum to these recommendations Ingetraut DAHLBERG proposed that, as the basis for future study, there be an analysis of the recommendations from the four previous International Study Conferences to determine where progress in classification research stands.

While these recommendations were not debated at the conference, they provide food for both thought and action over the next 5 to 10 years. With an analysis of progress on previous recommendations and de GROLIER's projections it is to be hoped that significant useful research can be achieved before a 6th International Study Conference takes place.

The essence of success in a conference is in its organization and participants. The Organizing Comittee and the conference participants fulfilled their roles superbly, as did the sponsors of the various social activities - Forest Press, FID, and the Faculty of Library and Information Science and The Library, University of Toronto.   Nancy Williamson

232