Reports and Communications

Pertinent Research at OCLC

In January 1990 we received the "Annual Review of OCLC Research, July 1988-June 1989", 68 pages full of most interesting information. In his introduction Martin DILLON, Director of the Office of Research of the Online Computer Library Center, Inc. in Dublin, Ohio stated that the Office focusses at present on four strategic areas: (1) enhancing use of the OCLC Online Union Catalog; (2) investigating the requirements of the library of the future, particularly in an academic setting: (3) increasing the productivity of catalogers; (4) digitizing data, both to preserve library materials and to transfer information from paper form into automated retrieval systems. - At present the OCLC Online Union Catalog has a database of more than 20 million bibliographic records. Many of the projects performed in the Office of Research aim at enhancing the use of this database. But there are also projects investigating the "Library of the Future" and the possible increase of the productivity of catalogers. In the following the nine (out of 25) pertinent contributions are listed with their project investigators and the abstracts to the project reports given in the Annual Review. Of interest is also the Distinguished Seminar Series. Among the 6 seminars the one given by Dagobert SOERGEL should be mentioned. It tells about his research project "Improved Access to LC Classification and LC Subject Headings for Online Searchers and Catalogers". The responsible Research Advisory Committee members ought to be congratulated for having provided valuable input indeed in setting the course for the OCLC research activities during this period.

Stuart L.WEIBEL, Research Scientist: **Applying neural networks to classification problems.**
Connectionist systems, often referred to as neural networks, represent a promising method for pattern recognition and classification. This report describes the application of neural networks to the problem of automated classification of document layout only objects as text or extraneous noise. These methods are compared with the performance of a conventional statistical method of classification and the advantages and disadvantages of each are discussed.

Roland HJERPPE, Visiting Scholar: **Bibliographic control and document architecture in hypermedia databases.**
This project focusses on the need and potential (1) for designing components and structures (formats) for electronic documents that enable algorithmic derivation and generation of a variety of bibliographic representations that can be used in "catalogs" of collections of electronic documents, and (2) for designing these catalogs as adaptable instruments integrated with the other tools available to the users or producers of electronic documents.

Martin DILLON, Director, OCLC Office of Research: ·
Enhanced bibliographic retrieval.
This project evaluates the retrieval effectiveness of contents information in bibliographic records. The first stage of this project examined table-of-contents information in a database of predominately technical monographs. The results showed that the inclusion of this information significantly improved retrieval effec-

tiveness. Further research will examine table-of-contents information in other subject areas, as well as other sources of contents information such as abstracts and back-of-the-book indexes.

Nicholas BELKIN; Tefko SARACEVIC, Rutgers University, School of Communication, Information, and Library Studies: **Design principles for third-generation Online Public Access Catalogs taking account of users and library use.**
This project will establish design principles for Online Public Access Catalogs (OPACs) on the basis of empirical study of library users and library use. Our major assumption is that OPACs are tools to support people as they see information. We must understand how people intend to use information, how they seek it, and how users' characteristics relate to these issues. Based on this data, we will suggest functions that OPACs should and could perform in various circumstances, identify the resources and the information about users such systems would need, and propose some principles for overall OPAC design.

Karen Markey DRABENSTOTT, University of Michigan, School of Information and Library Studies; Diane VIZINE-GOETZ, OCLC: **Increasing the accessibility of the Library of Congress Subject Headings in online bibliographic systems.**
OCLC and the University of Michigan engaged in a collaborative research project to investigate automated techniques to link subject terms entered by online catalog users with the library's controlled vocabulary. This report describes comparisons between assigned subject headings from a large bibliographic file and LCSH-mr (machine-readable Library of Congress Subject Headings) and comparisons of patron entered subject access points and the library's controlled vocabulary, phases 2 and 3 of the project, respectively. The final project report will include recommendations for subject searching improvements to guide online catalog users from their subject access points that are not expressed in terms of the catalog's controlled vocabulary to LCSH-mr and its syndetic structure. It will also provide display guidelines for LSCH-mr, its syndetic structure and linked assigned subject headings.

Clement YU, University of Illinois at Chicago, Department of Electrical Engineering and Computer Science: **Learning term weights.**
Learning and clustering procedures to weight terms for retrieval are described. Experimental results using two document sets show that a method of learning term weights based on the nonbinary independence model gives better performance than the inverse document frequency method and that learning from related queries, instead of from arbitrary queries yields improved retrieval performance. A formula to incorporate the frequency of terms in queries is presented.

Nancy J.WILLIAMSON, University of Toronto, Faculty of Library and Information Science: **The Library of Congress Classification in the Computer Age.**
Research leading to a detailed content analysis and profile of the Library of Congress Classification schedules is presented, including a description of the design, methodology, and preliminary findings of research in progress. Final results of the study are expected to aid the Library of Congress in the development of a MARC (MAchine Readable Cataloging) format for its classification schedules and in the conversion of the schedules into machine-readable form.

Gerard SALTON, Cornell University, Department of Computer Science: **A syntactic approach to automatic book indexing.**
Automatic publishing systems are now widely used to produce books and documents of many types, and the text of these materials, available in machine-readable form, can be used for automatic processing. This study describes automatic methods to generate back-of-the-book indexes based on a syntactic analysis of the text and a phrase-generation system that identifies meaningful index entries.

Roy R.LARSON, University of California, Berkeley, School of Library and Information Studies: **Enhancing topical searching using classification clustering.**
Early investigations of user interaction with online library catalogs showed a dramatic increase in subject searching when compared with card catalog use studies. However, subject searching also caused the highest rates of search failure and presented many difficulties for users. A method of enhancing topical subject searching using classification clustering is discussed.

Preliminary Programme TKE'90

The Second International Congress on Terminology and Knowledge Engineering, 2-4 Oct.1990 in Trier, Germany promises a lot. Again, as already in the first conference of 1987 a big number of Organisations and Institutions support this conference which will – if its preliminary program holds what it seems to promise – become a major event! There will be the following nine keynote lectures: A.DZINCHARADZE: The role of terminology in creating knowledge bases. – E.OESER, University of Wien: Terminology science and knowledge theory as a prerequisite of knowledge. – A.M.TJOA, R.WAGNER: Basic conceptual elements of knowledge based systems. – W.RAUCH: Information science in an information society. – A.MELBY: Terminology and technical documents: the importance and limitation of formal systems. – M.SCHAAR: Multilingual information and knowledge management. – D.WALKER: Lexical resources for the management of massive text-files. – G.ENGEL: New professional profiles in knowledge engineering and knowledge transfer.

Furthermore, 8 sessions are foreseen with altogether 85 papers. 7 workshops will be held as preconference events, most of them have already been mentioned in the note on the 2nd Call for Papers, TKE'90, published in Int.Classif.89-3, p.162. The Session papers will be listed in the following:

Session 1: Terminology, Knowledge Theory and Knowledge Engineering New Applications. Chair: E.OESER.
O.LASKE: Acquisition of knowledge for expert systems: unsolved problems (working title). – N.P.PESHKOVA: Different types of discourse structure as a means of knowledge representation. – J.HAAS: Treatment of uncertainty in social-science expert systems. – G.BUDIN: Scientific knowledge structures. – A.A.STOCHIK, S.NIZHNY, S.: Some clear and unclear aspects of research of relationship. – K.AVERBUKH: Term as a subject of study and a tool for knowledge fixation and knowledge transfer. – B.SÖRENSEN: 'Level Structuring' of knowledge and its application to an expert system domain with a particular view to the perspectives for terminology work in general. – S.D.SELOV: Terms, ternability and knowledge. – F.A.CITKINA: Comparative terminology theory: problems, goals, methods, applications. – B.Y.GORODETSKY: Cognitive aspects of terminological phenomena. – S.V.GRINEV: Terminology and knowledge theory. – M.F.PESCHL: Some cricial reflections on symbolic knowledge representation. – T.PETROVA: Gnosstic aspects of the term. – K.URAZBAEV: Space (exploration) terminology and the integration of science and engienering sublanguages.B.I.GUREVITCH: The descriptive approach to visual knowledge representation and visual processing.

Session 2: Knowledge-based Systems. Chair: W.KLAS
K.BAKARSIC: Science citation index for the field of archeology hermeneutical and informatics aspects of SCI based systems of regional bases. – W.BENN: Modelling multiple paradigm support in a KBMS. – R.FUJIWARA: Intelligent design of control systems combining knowledge-based systems with conventional design methods. – M.LINSTER, B.GAINES: Supporting acquisition and performance in a hypermedia environment. – I.MEYER, D.SKUCE: Computer-assisted conceptual analysis: an essential component of a terminologist's workstation. – J.NIGGEMANN: Analysis and representation of neuroanatomical knowledge. – G.OTMAN: Semantic networks in terminology: the case of AI. – M.A.Partyka:The application of the Quine-McCluskey algorithmm of self-assigning minimization to knowledge-based systems. – J.WALLMANNSBERGER: Hypertext approaches to terminological information processing.

*Session 3: Natural Language Processing and Knowledge Engineering.*Chair: B.B.RIEGER
B.ENDRES-NIGGEMEYER: A cognitive model of abstracting. – Ch.DEFRISE/S.NIRENBURG: A language for representing text meaning. – T.BULOT: Générateur de dialogue et optimisation de la consultation télématique. – G.KLOSE, K.LUCK: The representaiton of knowledge in LILOG. – M.TARNOWSKI: Text-Understanding – extracting knowledge from written texts. – M.GERLACH: Terminological reasoning in the natural language consultation system WISBER. – J.-U.MÖLLER: How to communicate expert knowledge. – H.SÖRENSEN: The use of knowledge-based frames for terms in EUROTRA. – I.SEREBRYAKOVA: Non-neural lexical units in scientific texts and the problem of their human and machine-as-

sisted human translation. – A.SCHILLER, P.STEFFENS: A two-level environment for morphological descriptions and its application to problems of German inflectional morphology. – Y.UEDA, S.AHLEN, K.KOGURE: Declarative control of generation using typed feature structures. – T.OSADA, S.MATSUYAMA, M.NAKADATE, H.NAKAI, T.ISHIBE: An analysis of drug intneractions using descriptions in drug package insert documents; a study on standardizing therapeutic words.R.L.BUCHAN: Retrospective indexing (RI) – A computer-aided indexing technique. – M.T.Genuardi: Knowledge-based machine indexing from natural language text knowledge base design, development and maintenance. – N.J.I.MARS, H.Van der VET: Using abstracts for semi-automatic knowledge acquisition. – S.SIMONIAN: Conceptual description of the linguistic information base of modern Armenian.

*Session 4: Documentation Languages and Ordering of Knowledge.*Chair: W.SCHMITZ-ESSER, B.LUTES-SCHAAB: An online thesaurus for fact retrieval. – T.ROGOZHNIKOVA: Individual information thesaurus: associative fields and a model of semantic development. – F.SARRE, J.MITTERMEIER, U.GÜNTZER, G.JÜTTNER: Evaluation of the thesaurus generating information retrieval system TEGEN with regard to learning behaviour and user acceptance. – F.KOVACS: Terminology and conceptual system for creating patient records. – R.SUNEETI:Thesaurus in law: some problems of construction. – S.SEETHARAMA: Towards establishing concordance among medical classification systems.–P.STANCIKOVA: Bilingual thesaurus VODOINFORM and its new functions in application of CDC/ISI software (micro- and mainframe). – R.VOLHO-LOPES: Automated access to multilingual information.

*Session 5: Electronic Dictionaries.*Chair:T.YOKOI.
L.BELIAEVA, R.PIOTROWSKI, S.SOKOLOVA: Principles of linguistic automata and their information bases design. –J.-M.MAES: The intelligent dictionary project. – R.MAYER, D.GEER, K.-H.HANNE: On how to bring Hypertext to termbanks. – S.MIIKE, S.AMANO: The structure and function of the EDR concept dictionary. –A.YEZHOV: Organization of lexicographic data on the basis of a lexical unit oriented approach. –P.BOUILLON, L.TOVENA: Word formation and computational dictionaries. – Kim DEOK-BONG, Hyun-Jae IM, K.CHIO, G.KIM: KOCP: an extended OCP for Korean and English. – L.LEMNITZER: A conceptual schema for a multifunctional lexical database: a proposal. – J.C.MERLO, deS.ZITZER: Terminological elcctronic dictionaries as tools for access to knowledge. – A.A.POLIKARPOV, L.KOLODYAZHNAYA: A system for compilation and analysis of electronic philological dictionaries using a personal computer. –J.RAY, M.NADKARNI: Electronic dictionaries– an Indian experience. – Y.TANAKA, S.YOSHIDA: Preparation of a concept dictionary.

Session 6: Information Management in Organizations. Chair: H.CZAP
H.CZAP: Construction and representation of concpets in enterprises. M.E.SHACKLETT: Integrating information systems into corporate strategic planning and plan execution. – H.BERTRAND-GASTALDY, L.C.PAQUIN, L.DUPUY: The need for information and knowledge management. – D.DAOUST: The evaluation of sociolinguistic and terminological change in a commercial enterprise.

Session 7: Computer Support in Technical Communication. Chair: S.-E.WRIGHT
H.EISELE: Les frustrations de terminologue. – U.HEID, G.FREIBOTH: Terminological and lexical knowledge for computer-aided translation and technical writing. - R.G.HEIN: Hypertext: indispensable information systems for communicators. – H.-D.LUCKHARDT: Automatic terminology support in computer-aided and machine translation. –G.M.SHREVE: Requirements analysis, empirical research and prototyping in the software engineering of workstations for computer assisted translation. – S.-E.WRIGHT, C.LARSON: Building the universal engine: implementation of transpaltform MTX. – S.E.DIAMESSIS: Ongoing efforts in Greece. –S.SIRAYEVA: Compound term engineering. – L.L.NELJUBIN: Terminographic text formalization.

Session 8: Terminology and Knowledge Transfer Tools. Chair: C.GALINSKI
W.DILGER: Transferring knowledge bases among hybrid knowledge representation systems. – A.KUKULSKA-HULME: Speed understanding of an unfamiliar domain. – M.OSADTSCHUK: Utilization of semantic frames in studying and teaching terminology. – M.I.DIEGUEZ, A.M.BURDACH, P.I.CABRERA, B.C.G.GARBARINI, V.P.HÖRMANN, R.LAZO: A terminological microbank for Chile and Latin America. – C.GALINSKI: Knowledge transfer to/from countries with languages using non-European scripts. – F.GAUDIN: La socioterminologie et les discours d'experts. – L.GUESPIN: Socioterminology facing problems in standardization., – CH.HOSBAYAR, HASBAGAN: A study on Mongolian terminology for physical concepts. –D.M.ZIYADULLAEVA: Thematic dictionary on hydrotecnical construction. – W.NEDOBITY: Methodological and practical aspects of computer-assisted knowledge transfer to developing countries. – J.SANS: EOQL: Glossary of terms used in the management of quality. – Wubin SU: China National Technical Committee on Terminology for Standardization (CNTCTS).

There will be an accompanying exhibition. The proceedings arc to be available to participants at the beginning of the Congress.

For further information contact: The TKE'91 Congress Secretariat, Dipl.Kfm.Thomas A.Ochs, P.O.Box 3825, D-5500 Trier.