Reports and Communications

44th FID Conference and Congress on "Information - Knowledge - Evolution"

The 44th Conference and Congress of the Fédération Internationale de Documentation took place from Aug.24 to Sept.1, 1988, in Helsinki, Finland, in Alvar Aalto's Finlandia Hall, hence in a most noble setting in the best spirit of Finnish architecture. The organizing committee chaired by Ritva Launo and supported by sub-committees for specific tasks, each headed by a Finnish chairperson, had done exemplary work. The result was a big conference attended by almost 700 participants, which, down to the tiniest detail of the working and supporting programs, was marked by rarely experienced friendliness and solicitude. In addition, lady participants from Central Europe, certainly not overly indulgenced in this respect so far, for once experienced a conference at which - as the moderator of the final discussion panel put it - men and women were absolutely equal.

It is only within limits that the following report tries to present an overall picture of this FID Congress, focussing as it mainly does on that part of the event which is of interest from the point of view of computer linguistics and information science alike, with subjects of linguistic interest receiving special attention. Besides them, heed is also given to more general discussion points of importance for the working areas bordering on information science and computer linguistics.

The caliber of the personalities present at the inaugural session was fully in keeping with the ambitious overall theme of the conference. After Finnish Prime Minister HARRI HOLKERI, the chairperson of the organizing committee and FID President MICHAEL HILL, the next speaker to take the floor was ALEXANDER KING, chairman of the Club of Rome. His topic was the transition to an information society, as thoroughly different from the present-day industrial society as the latter is from its agrarian predecessor. Besides the population explosion and the greenhouse effect - the case of a worldwide temperature increase - he also counted the social consequences of information technology among the mainsprings of change. The final address was delivered by a Finnish philosopher of venerable age, emeritus professor SVEN KROHN, who characterized human knowledge in a manner which even AI researchers would not have found far removed from their way of thinking.

Content Analysis and Document Description

The session on "Content Analysis and Description of Documents" was subdivided into three parts. A first partial session was devoted to linguistic aspects, a second one to problems and trends in classification and indexing, while a third special session had as its subject "Industrial product documentation -Language aspects". Since not all speakers were able to actually read their papers, the Proceedings were also resorted to in writing this report.

Linguistic Aspects in Content Analysis

In her keynote address, Amy J.WARNER from Madison, WI, proposed a pinpointed research effort in information retrieval which would integrate linguistic research results. This assures her of the enthusiastic backing of her European colleagues, who, working in the same direction, have seen new prospects opening up to them which, over and beyond information retrieval, encompass the field of information science in its entirety. Looking, in linguistics, beyond such fields as kernel and system linguistics, they prefer to concentrate on active special fields such as computer linguistics, text linguistics and cognitive linguistics, all of which are particularly accomodating to the interests of information science. In a given concrete case, a stepped-up transatlantic exchange of views would certainly be useful for giving impetus to a novel scientific approach.

J.NIEMISTÖ and H.JÄPPINEN reported on the difficulties that automatic indexing finds itself confronted with in a strongly inflexional language like Finnish. A morphological analysis is called for. The MORFO system breaks down compound words into their components and reduces inflected forms to a canonical form. In the case of homographs, all interpretations are generated.

Likewise in connection with automatic indexing work for free text retrieval, Gerda RUGE advocated an improvement of retrieval using Boolean operators and such contextual operators as ADJACENT, NEXT or NEAR by conducting an additional search using dependence structures in nominal phrases, such as are offered by the SIEMENS company's COPSY system. The documents retrieved can be divided into relevance classes on the basis of the extent of agreement between the nominal phrases of the search question and the documenta. In a test conducted, COPSY assigned 85% of the documents to the correct relevance classes, with the remaining 15% being found in neighboring classes.

The procedure proposed by Carmen CAZARES for the automatic indexing in Spanish of Russian scientific-technical texts stems from her desire to make Russian data banks accessible to Latin-American users. A morphological analysis of the Russian texts supplies the basic forms of significant words, which then by means of a dictionary are replaced by Spanish words. If necessary, several alternatives are offered in the case of homonyms. Through this indexing procedure, retrieval and document content assessment becomes possible in Spanish. Varying with the subject field, the system allocates from 93 to 99% correct Spanish indexing terms.

The contribution by Ahmad Ali AL-FALAHI et al. from Iraq describes a simple procedure by which advanced developments of Zipf's Law are utilized for computer-assisted indexing.

Problems and Trends in Classification and Indexing

In their introductory overview, Peter INGWERSEN and Irene WORMELL distinguished between three types of research approaches used by information science: system-oriented, user-oriented, and cognitive ones. The most customary retrieval technique utilizes Boolean operators and is based on exact agreement between search string and string sought. The weaknesses of this procedure are well known. Improvements are being

sought by using, among other things, more precise indexing procedures, such as PRECIS, or intelligent user interfaces. In expanded Boolean retrieval procedures and probabilistic retrieval models, indexing and enquiry terms are weighted. Such procedures are far more efficient. Equally good results may be achieved with cluster models. The text representation procedures used are determined by the given practical requirements. With the uses made of information rapidly changing, new requirements arise as to the representation of and access to documents. The main problem in document representation is the different interpretation of concepts by different persons. System design must be oriented to the information problems of the users. Particularly unclear thematic retrieval questions are hard to handle.

In a Boolean retrieval operation, Miranda Lee PAO observed that thematic indexing and indexing on the basis of the literature cited produce wholly different retrieval results, with literature on the basis of the literature cited leading to a high precision. It appears advisable to use the cross-section of both indexing procedures when very high precision is desired.

Cognitive models and procedures were resorted to in two contributions to explain content analysis processes. Brigitte ENDRES-NIGGEMEYER proposed a descriptive framework for content analysis which is based on two adapted, more general models: a model for the writing of factual texts and a model for the understanding and abstracting of texts. In addition to a research program she was able to formulate first elementary abstracting operations. Irmeli HOVI investigated among 35 test persons how they go about classifying documents. On the whole, the rough subdivision of the process into three phases (analysis of the document, identification of its central concepts, and expression of these concepts in terms of the classification used) as per ISO Standard 5963 conforms to reality. The classification system that will be used influences the representation that is made of the document. Already existing classifications often complicate the intellectual process by offering too many alternative classes without their different suitability for the given case becoming apparent.

A group of papers dealt with the UDC. Päivikki KARHULA described a Finnish project which, as a follow-up to the work of Irmeli HOVI, aims at a better adaptation of the UDC to use in online data banks. Klaus LOTH and Herbert FUNK presented the ETHICS system for UDC-based literature retrieval as employed by the ETH, the Zürich Institute of Technology. Eugeniusz SCIBOR and Ina SHCHERBINA-SAMOILOVA explained the new UDC revision strategy. Five co-ordinating revision committees are each responsible for a broad range of subject fields. In addition to these revision committees, paid experts are likewise working on UDC revision. This revision is taking place step by step so as to avoid a rupture between the existing UDC and a revised one. Under the promising title "An expert system for automatic UDC number assignment" Yukio NAKAMURA and Tetsuya ISHIKAWA presented a system for computer-assisted UDC use.

Masaaki TONOSAKI described medical professional journals on the basis of their bibliographic structure, evaluating, in so doing, particularly the columns (original papers, reviews, etc.) occuring in them. The journals are grouped together into clusters, so that a typology of these professional journals is obtained.

Henning SPANG-HANSSEN pleaded for more research and development in the field of alphabetic index preparation, with the UDC subject index serving as his most important example. It is only with improved methodology that computer-produced indexes will be able to attain the information quality of those prepared by man.

It was an impressive experience to hear M.M. MOSHOESHOE reconstruct the laborious process by which a library system was brought into being and subject description as a librarian's task was introduced in her home-country Lesotho. A small country where the first scientific libraries came into being only in the 1960's now has progressed to the point where it is linked up with supranational library networks.

Using as her example a series of conferences organized in Dubrovnik, Yugoslavia, Neva TUDOR-SILOVIC depicted the informatory effects scientific congresses may produce.

A background for the current efforts to arrive at improved methodology and more international co-operation in content analysis, Antje SCHUREK's contribution on UNESCO's General Information Program was most interesting.

Industrial Product Documentation - Language Aspects

Hans KARLGREN reported on the efforts undertaken by the FID Committee on "Linguistics in Documentation" to place linguistical questions in a broader context. Two years ago the committee started to occupy - and still occupies itself with the subject "Technical Documentation". Without accompanying information material, complicated products such as paper factories or software systems will not be operable. The production of such user information means task-oriented text-writing under professional conditions for practical utilization. We are dealing here with a highly interesting, special case of knowledge transfer. A case study conducted in a large Stockholm enterprise brought to light an awareness of the problem but only few actual improvements in the procedures employed. In particular, increased computer use was not leading to better methods.

Kolbjorn HEGGSTAD stressed the necessity of adapting software products and their documentation - as is done with other products, too - to existing local circumstances, of "localizing" them as it were. Particularly under difficult conditions, knowledge is more easily and thoroughly acquired in one's mother tongue. Why should one be afraid of not seeming to be international enough? It would be more useful to develop one's own language and terminology to the point where a local EDP culture can be represented in it.

Oyvind HAALAND contributed a well-presented case study: The documentation on the Oseberg offshore platform northwest of Bergen, Norway, comprises more than a million A4 pages. For reasons of safety and cultural identity, "Nor-english" had to be avoided as a working language. A Norwegian terminology had to be developed. The company entrusted with this task did the necessary translating work, conducted courses in technical writing, developed a Norwegian oil terminology and checked and revised the documentation and in-

struction material. From the point of view of quality assurance and avoidance of extra costs, this investment in documentation has proved worthwhile.

The panel discussion on Language Engineering was conducted with an all-Scandinavian panel. Henning SPANG-HANSSEN did not find a national EDP language necessary: the problem rather was to keep up to date at all times; the (linguistically) exotic was easier to sell; bilinguality was a normal phenomenon in a great many countries; where resources are short, the preservation of linguistic identity was one of the things that had to yield to more important objectives. To this, "clever localizer" HEGGSTAD replied that an exotic brand would only then be easier to sell if the product meets the local requirements as to functional effectiveness. From experience gathered in Sweden Hans KARLGREN contributed the following: A nation of 8 million people was too small to have its own special language for each technical field. Therefore on should not export one's own nationalism into other countries. Starting out from Luther's translation of the Bible into German, he described how the linguistic history of Central and Northern Europe might well have developed along more positive lines: Just suppose that, as basis for the written German language Luther had not selected, as he unfortunately did, a High German dialect but rather Low German, which, as the Hanseatic language, also served as working language in all of Scandinavia. Then this Low German-based written language would have remained the standard language of the educated, i.e. the literate, throughout the German-speaking world as well as in all of Scandinavia. Nor would Finnish, a rather artificially revived language, have been able to firmly establish itself, since then it would not have been Swedish, but rather the type of German thus taking shape that it would have had to compete with - and probably succumbed to. In the young USA, the vote on whether English or German should become the official language would not have been narrowly carried by the English language, but, with the support of the US citizens of Scandinavian descent, by the German one. Under those circumstances Scandinavia would never have been confronted with the dilemma of whether or not to develop an EDP language different from English.

Man as an Information Processor

Matti BERGSTRÖM discerned, in man, a 4-stage mechanism in the brain for process control in information processing: Man's "value capacity" controls the information processing proper; the latter, in turn, controls the energy flow, through which, finally, matter is acted upon. If man's education does not provide for adequate systematic exercise in this value capacity, the result is a specific form of disability which renders the individual unable to acquire an overall view of situations and to take decisions on the basis of such a view in any given situation. Apparently constituting neither information nor energy, the "value capacity" is connected with the holistic capacities of the right brain half. It also controls man's motivation and emotional capabilities and thus his or her ethical conduct.

Most instructive was Thomas J.FROEHLICH's contribution on relevance, a central concept in information science. Basing himself on contemporary philo-

sophers (Schulz, Merleau-Ponty, Habermas, Heidegger, Gadamer, Taylor) he showed that there is no absolute foundation of knowledge. Rather, we are born into a certain practical - historical and social - situation from which our knowledge stems and into which it flows again. On this basis we develop a general concept of relevance. More developed relevance judgments are the outcome of a successful systematic exercise in the determination of relevance as more specially related to life, e.g. in retrieval from information banks.

Electronic Knowledge

Under the title "Wholly new forms of encyclopedias", Linda C.SMITH presented the main characteristics of hypertext systems. Although not new as a concept, hypertexts are now gaining new actuality from the fact that systems for their electronic representation are now commercially available. While printed texts in principle presuppose linear reading, a hypertext is based on an interlaced structure of the text, which the reader can follow along self-chosen paths or along ones pre-arranged by the author for different reasons. Production, distribution and use of hypertexts still need to be investigated more precisely.

Rapidly Growing Information Systems

The "Telecottage" concept, most successful in Scandinavia, was explained in several contributions (Henning ALBRECHTSEN, Leif BRISFJORD, Inez SPERR BRISFJORD). This concept stands for local computer and information network stations which, primarily in rural areas, offer infrastructural, training, consultation and related services. In their respective municipalities they facilitate decentralized EDP utilization, especially by small enterprises or private persons. Thus they contribute to the social and economic animation of the given region, particularly if it is a remote one. The organization of the telecottages varies, but is always based on the principle of self-administration.

Panel Discussion on "Overcoming the Barriers to Information - Local and Global Information Networks"

As panelists for the concluding panel discussion on the subject of information transfer and information networks, Peter CANISIUS had rounded up four ladies and four gentlemen from Africa, Latin America, North America and Europe. Two examples must suffice to illustrate the style of the discussion, which in remarkable fashion brought out the different personal points of view of the participants, coming as they did from a great variety of walks of life. Marjatta OKKO, who claimed to rank first in seniority among the Finnish professors of information science, had assumed the task of calling the subject into question. Accordingly, she called attention to the highly useful channeling function of information barriers. Later on she was to offer the valuable suggestion of splitting off a scientific conference from the FID conference. M.M.MOSHOESHOE, president of the Lesotho professional librarians' association, described, with all the emphasis desired, how in her everyday professional life questions whether book transport by donkey is feasible or whether the worst information barrier did not consist in the population's illiteracy rate of 85%. On the whole, a spectrum of opinions of rare

richness in facets revealed itself, thus once more illustrating, in concentrated fashion, the advantages of the conference before the moderator drew a final resume of the event. The next FID Congress will take place in Havana, Cuba, in 1990.   Brigitte Endres-Niggemeyer

Prof.Dr.B.Endres-Niggemeyer, FH Hannover, FB BID,
Hanomagstr. 8, D-3000 Hannover 91

Gesellschaft für Klassifikation eV, Herbsttagung 1988

On Dec.2-3, 1988 the Section on Data Analysis and Numerical Classification (SEK-DA NK) of the German Society for Classification held its traditional "Fall-Meeting" with the following 10 papers presented: DEGENS, P., WOLF, K.: Kritische Untersuchung einiger Verfahren zur Bestimmung von additiven Baummetriken. - VACH, W.: Kleins-Quadrate-Approximation von Baummetriken. - DRESS, A.: Eine kanonische Zerlegungstheorie für Metriken auf endlichen Mengen. - KOHLSCHE, A.J.: Numerische Klassifikation bei Hirntumoren (Schmitt-Problem). - DEICHSEL, G.: Cumulative Logits für Kontingenztafeln mit geordneten Responsekategorien - ein Ansatz mittels generalisierter linearer Modelle. - FEGER, D.H.: Analyse von Merkmalsmustern. -HERDEN, G.: Clustermethoden for ordinal skalierte Daten. - WILLE, R.: Wissenserwerb mit Methoden der Formalen Begriffsanalyse. -TÜSHAUS, U.: Entscheidungsunterstützung und Wissensrepräsentation. - ALDAG, R., STECKNER, C.: Praktische Demonstration des PC-EXPRESS Datenbank- und Datenanalysesystems.
For further information please contact: Prof.Dr.M.Schader, Universität der Bundeswehr Hamburg, Postfach 70 08 22, D-2000 Hamburg 70

SoftStat 89

5th Conference on the Scientific Application of Statistics Software, 2-6 April 1989, Heidelberg, FRG.

The organizer of SoftStat 89, the Zentrum für Umfragen, Methoden und Analysen (ZUMA) e.V. has issued its announcement and preliminary program for its 5th Conference. 104 papers are listed in 24 sessions. The conference includes also panel discussions, poster sessions and 17 software demonstrations. For further information write to: ZUMA, Postfach 12 21 55 - B2,1 D-6800 Mannheim 1.

BMDP Data Analysis Course

Prior to the SoftStat 89 conference there will be a Data Analysis Course at the University of Heidelberg, FRG on March 30-31, 1989. Topics include Multiple Linear Regression, Repeated Measures and Factorial ANOVA, Frequency Tables and Log Linear Models, Correspondence Analysis, Factor Analysis, Discriminant Analysis, Cluster Analysis, Data Screening, Missing Data, and the BMDP Data Manager. To register call or write to Statistical Software Ltd., Cork Technology Park, Model Farm Road, Cork, Ireland, Tel.: +353 21 542722. Registration fee 700 DM, before Feb.1 : 600 DM.

2nd International Workshop on Data, Expert Knowledge and Decisions

In order to strengthen connections between economics and related areas of research, expert knowledge research, data analysis and decision support a workshop has been announced to be held at Hamburg, Sept.3-5, 1989. The program Committee comprises P.Arabie, A.Bultez, E.Diday, W.Gaul, F.J.Radermacher, and M.Schader. For further information please contact: Prof.Dr.M.Schader, Institut für Informatik. Universität der Bundeswehr, Holstenhofweg 85, D-2000 Hamburg 70.

Terminologie und Wissenstechnik

This is the title of a seminar held in two blocks on Nov.18-19, 1988 and January 13-14, 1989 at the Austrian Standardization Institute, Vienna. The four lecturers of the first seminar and their topics are: E.OESER: Terminologie und Wissenstechnik als Anwendungsgebiet der Wisschaftstheorie. - H.BÜHLER: Interlingualer und transkultureller Wissenstransfer aus der Sicht der Translatorik (angewandte Übersetzungswissenschaft). - W.NEDOBITY: Wissensordnung aus terminologischer und informationswissenschaftlicher Sicht. - C.GALINSKI: Wissensdarstellung in der Terminologie. The seminar is to be summarized by a panel discussion followed by a general discussion. The topics of the second block are devoted to applications and new professional images. Speakers will be a.o. representatives from industry. For further information contact: Mr.Ch.Galinski, Infoterm, Postfach 130, A-1021 Wien.

International Conference on Terminology Standardization and Unification in Theory and Practice

This conference is to take place on March 13-17, 1989 at the Hotel El Mechtel, Av.Ouled Haffouz, Tunis-Belvedere. It is organized by the Tunesien National Institute for Standardization and Industrial Property in cooperation with Infoterm. Topics of the conference are:
- Theory and methods of international and regional terminology regulating bodies. - National terminology standardization and unification activities. - Company level terminology unification. - The users: Needs, access, training. -International coordination of terminology standardization. - New tools for assisting terminology standardization. - The conference is to take place in the mornings. In the afternoons workshops are planned on the following topics: Coordination of terminology work in the field of ethnicity research. - Teaching material for terminology training. - Computer-assisted knowledge transfer to developing countries. - Congress languages are Arabic, English, and French. Simultaneous interpretation will be provided. Congress fees: USD 100 with 50% reduction for contributors and students. For further information please contact: Mr.Zuohir Marrakchi, I.N.Nor.P.I., P.B.23, 1012 Tunis-Belvedere, Tunisia.