KO Reports

Challenges in Knowledge Representation and Organization for the 21st Century: Integration of Knowledge Across Boundaries

Seventh ISKO International Conference
Granada, Spain – July 10-13, 2002

Nancy Williamson

Participants at the Seventh International ISKO Conference, Granada, Spain, July 10-13, 2002 spent a very full and busy four days. They listened to an impressive range of papers covering much ground and moving from general fundamental theories through many individual research projects demonstrating the application of new and old methods of organizing knowledge, and stretching from facet analysis to artificial intelligence in the struggle to cope with the tremendous flood of information pouring towards us through the Internet. The theme of the Conference was "Challenges in Knowledge Representation and Organization for the 21st Century: Integration of Knowledge Across Boundaries." Papers were presented in sixteen sessions over the four days. The published proceedings contains 80 papers, a few of which were not presented because the authors were unable to attend the Conference.

In her keynote address entitled "Conceptual Universals in Knowledge Organization and Representation," Rebecca Green drew our attention to the importance of universal concepts in Knowledge Organization and Representation, stressing that it is the concept, rather than the terminology, that remains constant across languages and cultures. On the assumption that where universals exist knowledge integration is most likely to occur, her paper explored "a representative inventory of semantic and lexical universals" that should be accounted for in any information system. Indeed, she set the scene for the subsequent presentations, touching on the problems of vocabularies both actual and artificial, (i.e., controlled vocabularies, including classifications), on the need for a clear analytical approach and the possible potential for switching languages. Green also highlighted the challenges presented by cultural differences. All of these themes were revisited throughout the conference. In conclusion Green stated that problems of comprehension and retrieval associated with lack of physical structure on the web have caused information professionals to return to traditional principles, but she saw this as only a partial answer. "The development of effective information systems hinges on understanding social and cognitive environments as well as knowledge of organizational structures and technological capabilities."

The opening session included four papers on "Theoretical Models and Universals in Knowledge Organization and Representation," all dealing with aspects of cognition. Jack Andersen addressed the problem of "Ascribing Cognitive Authority to Scholarly Documents: On the (Possible) Role of Knowledge Organization in Scholarly Communication." His concern was the examination and discussion of how, and to what extent, knowledge organization as an epistemic instrument in scholarly communication, can contribute to cognitive authority of scholarly documents. In her paper on "Augmenting Human Capabilities: Classification as Cognitive Scaffolding," Elin Jacob sought "to extend the notion of the classification scheme as a culturally-transmitted knowledge tool by emphasizing the cognitive value of the schemes internal patterns of relationship." In doing so, she considered classificatory structures in terms of their constraints, selections and expectations. In this same session, Clare Beghtol examined "Universal Concepts, Cultural Warrant and Cultural Hospitality." Her analysis started from the prem-

ise that "regardless of linguistic and other domain boundaries or cultural traditions" access to information can be examined on the basis of how cultural universals are implemented in specific cultures at specific times and places." "Time," and its implementation in calendars, was used to illustrate how this concept is treated in systems of knowledge organization. "Hospitality" was interpreted broadly and a "theoretical principle of cultural hospitality" was derived to provide the theoretical framework for decisions about types of access that could be "globally useful and ethically balanced." Time was also a factor in the final paper in this group. In "Subject Ontogeny: Subject Access through Time and the Dimensionality of Classification," Joe Tennis based his presentation on the understanding that knowledge changes, making constant revision of classification a necessity. He sees this as the opportunity to "add dimensionality to classification" and a chance to use classification schemes to explain conceptual warrant, to explain the shift from disciplinary to multidisciplinary knowledge, and to serve as a tool for domain analysis. Some applications of subject ontogeny were proposed, including metadata preservation models, online access tools and interoperability frameworks. The author sees subject ontogeny as an addition to, or an enhancement of, existing classification systems – acting as an interpretive layer to serve the needs of information workers.

The second session of the conference focused on "Epistemological Foundations for Knowledge Structures and Analysis." A paper by Nuno Silva and João Rocha entitled "Merging Ontologies Using a Bottom-Up Lexical and Structural Approach" took the view that classical integration approaches do not meet current requirements and sought to propose new strategies. The authors suggest an "alignment process" as opposed to the merging of data sources. This process, applied at the ontological level, is described in six phases – normalization, similarity measuring, bridging, representation, transformation and negotiation. "Ontological Analysis of Literary Works of Art" by Gillola Negrini and Patrizia Zozi focused on the way that ontological structures can aid the understanding and modelling of works of art. The starting point was an analysis of the authors' own work on the *Thesaurus of Italian Literature,* that used categories derived from the work of Ingetraut Dahlberg and some of its significant aspects were highlighted. Following from this, the theories of Nicolai Hartman and Roman Ingarden were examined in a search for commonalities between the Dahlberg ontology and the two theories, hoping for some easy solutions. Jarmo Saarti, in his paper on "The Analysis of the Information Process of Fiction: a Holistic Approach to Information Processing," challenged the concept of "aboutness" in subject catalogues. It compared the work of the author with previous studies of the information process by such authors as Clare Beghtol, Annelise Pejterson and Jutta Austin. Findings suggested that users vary in the search elements that they use, establishing the need to include documents in their totality in the information process. Systems have to be multi-faceted to meet all the needs of various users. The final paper in the group by NairYumiko Kobashi, Johanna W. Smit and M. de Fátima G.M. Tálamo on the "Constitution of the Scientific Domain of Information Science" attempted to delimit the terminology of information science on the assumption that the "central" terms of the discipline should form the core of the discipline and the peripheral terms should be derived from knowledge areas with which information science interfaces. The Information Science discipline was described as one that is "in development" and does not yet have clearly defined boundaries or interfaces. The authors state that the terminology borrows from a number of domains including logic, administration, linguistics, computer science, sociology, communications, and cognitive science, as well as librarianship. Procedures were set up for detecting terms. Based on the results, it was concluded that central terms come in large measure from librarianship and constitute the field of "Information Organization and Representation," while terms from other areas showed little adaptation. In the majority of cases the terms were adopted as is, but "their original conceptual structure" was abandoned. Further, it was determined that "the interdisciplinary character of Information Science is clearly upheld in the majority of discourse ... from the 1990s." Information science is described as a social science, rarely presenting itself as a science. While the authors feel that it is not a full-fledged science, it appears to aspire to intervene "practically in society through optimization of information flow."

A major segment of the conference focused on "Models and Methods" for knowledge organization and retrieval. Perhaps this should not be surprising since, after all, this is what the conference was about. Five sessions were devoted to this general topic, each dealing with a different aspect – representation, tools and systems, retrieval, the electronic environment (i.e., the Internet), conceptual relationships and information systems. In the session on "representation," five papers were presented, each dealing with an unusual situation. Anita

Coleman discussed "A Classification of Models" themselves. As stated by the author, our systems of knowledge organization do not provide a unified location for works about "models." This paper presented a framework for such a class of one category of models (i.e., scientific models) using facet analysis. Gain Piero Zarri addressed the "Indexing and Querying of Narrative Documents, a Knowledge Representation Approach" in a paper that described a system called Narrative Knowledge Representation Language (NKRL). The system takes into account narrative multimedia documents (e.g., news stories, corporate documents, legal texts, and representations of patients' medical records) and uses knowledge representation principles and some high-level inference tools. Jeremy Shapiro's "Interdisciplinary Knowledge Integration and Intellectual Creativity" explored and analyzed the "use of culturally prominent metaphors, symbols, archetypes, myths and narrative patterns as metadata, as a method of facilitating the discovery and retrieval of information and the integration of knowledge across both disciplinary and cultural boundaries." A Universal Cultural Symbol Thesaurus is described as a potential subject language for a lexicon of metaphors and symbols that could be used to classify information objects. Theoretical background is given, the metaphorical and symbolic metadata are explained and the thesaurus described. The final paper of the session, on "Images and Words" by Catalina Naumis Peña, focused on a new resource for indexing images. The differences between traditional indexing and digital audiovisual information were explored and the use of the thesaurus as a tool for image indexing re-evaluated. The author proposes Erwin Panofsky's principles as the basis for a new thesaurus for image indexing.

In a "Models and Methods" session on "Tools and Systems," five papers were presented with particular emphasis on the need for tools and systems to handle digital resources. A presentation by Maria Inés Cordeiro and Aida Slavic examined "Data Models for Knowledge Organization Tools" and considered the need for such tools to be "fully disclosed and available" in the open network environment. The place and value of traditional library and information science tools – classification schemes, thesauri and authority files – and their interoperability was explored. Some of these tools are already available; others need to be adapted and reshaped. In contrast, Vanda Broughton, in a paper entitled "Facet Analytical Theory as a Basis for a Knowledge Organization Tool in a Special Subject Portal" described a project being carried out at University College London (UCL), which focused on the way classification schemes using facet analysis principles can be applied to the organization of digital resources. The UCL system builds on the classificatory principles developed by the Classification Research Group and derives its methodology from the *Bliss Bibliographic Classification*. The nature of classificatory structures, the application of faceted classification and the facet analytical methodology are explained. Facet analysis was the basis recommended by many at the conference, though not everyone put the same interpretation on this much-used and frequently misunderstood term. Broughton gave a clear definition in her paper as follows: "It is taken to mean that rigorous process of terminological analysis, whereby the vocabulary of a given subject is organized into facets and arrays, resulting in a complex knowledge structure with both semantic and syntactic relationships clearly delineated." In this same session, Stella Dexter Clarke addressed the problem of "Planning Controlled Vocabularies for the UK Public Sector." Public sector information is being made available through an e-Government Interoperability Framework that is based on a set of standards, one of which is a "a small and simple taxonomy," a controlled vocabulary known as the Government Category List (GCL). Improvement in subject access is needed and six thesaural options (or models) were discussed here. In a similar vein, Widad Mustafa el Hadi focused on the need for tools to satisfy access to new media in the form of Natural Language Processing (NLP) techniques. In this context, the author surveyed the "recent trends of terminology acquisition and management in relation to information retrieval." New trends, new techniques and most particularly the needs and means for building ontologies were discussed. It was suggested that some testing of existing tools on a representative corpus is needed before drawing up the ontology. In the final paper of the session Hur-Li Lee and Allyson Carlyle gave a presentation on "Academic Library Gateways to Online Information: A Taxonomy of Organizational Structures." The authors describe a preliminary analysis of schemes applied by academic libraries in arranging their electronic resources. Most of the gateways considered were created locally and do not conform to the well known universal schemes. When the websites in academic libraries in 10 countries were examined, there were many differences and few recognizable patterns found. Cultural differences were the exception to this. No final conclusions could be drawn but there were implications for future re-

search. Two findings did emerge: among academic librarians very simple schemes of organization were considered appropriate; and there is a tendency for libraries of the English speaking world to make more effort to organize than their counterparts in other parts of the world.

Closely related to "Tools and Systems" of knowledge organization was a session of "Models and Methods" in "Retrieval." The first paper by Gerhard Riesthuis and Maja Žumer on "The Functional Requirements for Bibliographic Records and Knowledge Organization" dealt primarily with descriptive cataloguing data, as opposed to subject cataloguing. The catalogue is viewed as an interconnected network with special attention to access points and relationships between entities and the changes they will bring to the new catalogue model. A paper by Snunith Shoham and Rochelle Kedar on "The Subject Cataloguing of Monographs with the Use of a Thesaurus" presented the findings of a study of indexing procedures with the use of a thesaurus for post-coordination. The article outlines numerous subject cataloguing mistakes attributed to the use of pre-coordinate systems – specifically LCSH – and explores the alternative of a post–coordinate system. Findings indicated that post-coordination resulted in an increase in the number of descriptors assigned and the longer the item catalogued the greater the number of descriptors. Mistakes in the assignment of descriptors tended to be for topics with less than 20 percent literary warrant in the item. Other errors could be attributed to the lack of sufficient scope notes and the lack of familiarity of the indexers with the post-coordinate type of indexing. In the paper on the "Application of the Cantor Set Theory in Making Decisions about Collections Development" by Ana Pérez López, Mercedes de la Moneda Corrochano and Ángel Moros Ramírez, Cantor Set Theory is used as a decision-making tool in collection development. It takes a "structuralist approach" and depends on the identification of underlying structure of the subject provided by relations between demand, knowledge and patterns of publication. The resulting structure is then used to formulate a collections development policy. The application of the theory proved useful as it aided in the identification and representation of the core of knowledge, the peripherals and the degree of interdisciplinarity in particular thematic areas. Hemalata Iyer and Jeanne Keefe described an exploratory study to assess the value of "WordNet as an Auxiliary Resource to Search Visual Image Database in Architecture." Carried out at the Rensselaer Polytechnic Institute, Troy, New York, the *Art & Architecture Thesaurus* provided the terminology and the study revealed problems encountered by "lay users" and observed how WordNet was useful. In the final paper of the session, a paper by Douglas Tudhope, Ceri Bindings, Dorothee Blocks, and Daniel Cunliffe on "Representation and Retrieval in Faceted Systems" discussed the retrieval potential of faceted thesauri and XML representations of fundamental facets. The work described is related to the "FACET" project, collaborative research between the University of Glamorgan in Wales and the UK National Museum of Science and Industry.

Another "Models and Methods" session that focused on "Conceptual Relationships" contained five papers, three of which were domain related. In their paper on "Structured Models of Scientific Concepts for Organizing, Accessing and Using Learning Materials' Terence Smith, Marcia Zeng and the ADEPT Knowledge Organization team "introduce a 'strongly-structured' model of scientific concepts that provide the foundation for a knowledge base." The domain-related system is described and applied in a digital library environment. Another domain-related presentation on "Reference Linking in Economics: the CitEc Project" by José Manual Barrueco and Vicente Julián Inglada described a software system that would provide for the automatic linking of each work cited in a document with its electronic full text. CitEc is the agent that automates the process. Also, domain-related, "Equivalence in Tillett's Bibliographic Relationships Taxonomy: A Revision" by Allyson Carlyle and Lisa Fusco, challenged Barbara Tillett's definition of equivalence as being too restrictive by excluding relationships among items "that may, based on contexts of use, act as equivalent." A revised definition is offered. Closely related to the Carlyle and Fusco paper was the final presentation of the session by José Antonio Fraís and Ana Belén Ríos Hilario entitled "Visibility and Invisibility of the Kinship Relationships in Bibliographic Families of the Library Catalogue." This work also focuses on the study of the *"Functional Requirements for Bibliographic Records."* As such, it is a detailed analysis of that document and is more closely related to descriptive cataloguing than to the representation of topical subjects. Jonathan Furner described "A Unifying Model of Document Relatedness for Hybrid Search Engines" that focused on providing a system that would give information seekers the opportunity "to exploit multiple sources of evidence of document relatedness in the process of search." The proposed system is graph-theoretical and would give users control over multiple ways in which document collections might be ranked and re-ranked for retrieval.

In the final of the six sessions on "Models and Methods" the focus was on 'Integration in Information Systems." At the most general level Rebecca Green, Carol Bean and Michèle Hudon discussed "Universality and Basic Level Concepts." In doing so, they questioned whether a concept's hierarchical level affects its universality across schemes for representing and organizing knowledge. As a basis for analysis, equivalents were drawn from a bilingual thesaurus, a pair of biomedical vocabularies and two ontologies. Findings suggested that equivalent concepts occur most often at basic levels as opposed to either subordinate or superordinate levels. As a consequence, the authors recommended that attempts to integrate knowledge representation and organization tools should be made at the basic level. The remainder of the papers in this session focus on specific systems. In "Chronotope and Classification: How Space-Time Configurations Affect the Gathering of Industrial Statistical Data" Grant Campbell used Mikhail Bakhtin's theory to investigate how representations of space and time affect the first six classes of the *North American Industry Classification System.* Closely related to Campbell's work was the presentation by Fernando Elichirigoity and Cheryl Knott Malone, entitled "Representing the Global Economy." In conjunction with the North America Free Trade Agreement (NAFTA) the three countries involved – Canada, Mexico and the United States -- had developed a unified *North American Industry Classification System* to replace separate economic classification schemes used in each country. The authors took the novelty of this situation as an opportunity to explore the implications of the creation of such a tool, to explore the role of economic theory and political context as they affected classificatory theory and to ponder the fact that the organizing data about industries "serves an essentially different purpose than schemes for organizing library materials." In a paper entitled "Corporate Thesauri – How to Ensure Integration of Knowledge and Reflections of Diversity," Marianne Lykke Nielsen and Anna Gjerluf Esau compared three methodologies for constructing thesauri. The methods used were literary scanning, word association tests, and the involvement of subject expert groups. Their analysis indicated that the methods are complementary and should be used together in constructing the same thesaurus. In her paper on 'Knowledge Integration and Classification Schemes," Nancy Williamson used the three major universal schemes in a preliminary analysis on the handling of the domain "environmental sciences." This work in progress is concerned with problems of primary location, interdisciplinarity, degree of scatter, terminology and structure. "Semantic Views over Heterogeneous and Distributed Data Repositories," by M.V. Hurtado, L, García and J. Parets was concerned with providing users with integrated access to data located in different repositories. The authors propose a two step process which would preserve the semantics of the data stored. A first step would provide an integrated schema or semantic overview and an integrated terminology as a basis for describing the content of the repositories. In the second step the schema and terminology would need to be enriched to allow for "evolutionary" aspects of the system.

A session on "Organization of Integrated Knowledge in the Electronic Environment" brought together five presentations on the Internet. A paper on the "Organization of the Information about Health Resources on the Internet" by J.A. Salvador Oliván, J.M. Agnós Ullate and Mª. J, Fernández Ruíz, focused on the fact that the usual search engines are inadequate to meet the needs of users. As an alternative they propose the design of a medical information retrieval system that uses MeSH as the controlled language and fields structured primarily on the basis of the Dublin Core. Criteria for quality and credibility were provided and the goal was a system that would be an improvement on other search engines and subject gateways. A "Practical Method to Code Archive Finding Aids in Internet" by Eduardo Péis, Antoni Ruiz, Francisco José Muñoz Fernandez, and Francisco de Alba Quiñones dealt with another specialized retrieval problem. Taking advantage of the need for specialized handling, the authors propose the use of the Encoded Archival Description (EAD). Based on SGML this tool "makes information interchange possible." In contrast, Marthinus van der Walt's presentation described "An Integrated Model for the Organization of Electronic Information/Knowledge in Small, Medium and Micro Enterprises (SMMEs) in South Africa." This paper explored the "feasibility of using a business process model as a framework for the integrated organization of electronic information in the context of business enterprises in general" and in particular one-owner businesses and firms with fewer than 100 employees. Three contexts were examined – document creation, document collection and information retrieval systems – based on surrogates. Findings suggest that this approach is "theoretically feasible." Finally, Roberto Poli presented a paper on "Framing Information" that distinguished between semiotic, semantic and ontological classifications.

Another area of concentrated interest was artificial intelligence, for which there were three sessions. One can conclude from an examination of these sessions that much active work in this area is occurring in academic communities in Spain where all but two of the papers on this topic were produced. One paper came from each of Singapore and France. The Spaniards appear to be a closely knit research group since some authors are involved in more than one of the papers. In the first session on "Applications of Artificial Intelligence Techniques to Information Retrieval" three papers were presented. Christopher Khoo, Karen Ng and Shiyan Ou carried out "An Exploratory Study of Human Clustering of Web Pages." Here the purpose was to find out how human beings cluster web pages naturally. Based on ten queries, a group of web pages were retrieved using the Northern Light search engine. The results were then sorted by three human subjects into "meaningful" categories. Findings indicated that different users clustered the same web pages quite differently, creating different categories. The goal of the authors is to develop automatic methods of categorizing web pages based on the categories that the users create. Stéphane Chaudiron, Majid Ihadjanene and François Role presented a system designed to provide for "Authorial Index Browsing in an XML Digital Library." The system proposed would offer a possible delivery format for controlled vocabularies and might "offer a single interface to different controlled vocabularies and classification systems." Xavier Polanco set out a proposal to use "Clusters, Graphs, and Networks for Analyzing Internet-Web Supported Communication Within a Virtual Community." Co-citation analysis was used in retrieval from the websites of countries of the European Union and examined using graph theory and social network analysis. Further research, in the framework of pattern recognition and exploratory data analysis will be based on graph theory. Pedro Cuesta, Alma Gómez and Francisco Rodríguez investigated "Using Agents for Information Retrieval" with a focus on intelligent information retrieval agents. Such agents are seen as a "key technology" for the future of the Internet. The functions of such agents are described as assisting users in finding useful relevant information, managing and overcoming difficulties associated with information overload, informing users of new relevant data that has been published, and carrying out tasks independently or working in a co-ordinated way with other agents. A prototype system called MASIR is introduced and described.

In the second session on "Applications of Artificial Intelligence," four papers were presented on topics relevant to aspects of the field. All contributors come from a group of researchers at the Universities of Granada and Extremadura. An "Evaluation of the Application of Genetic Algorithms to Relevance Feedback" prepared by Cristina López-Pujalte, Vincente P. Guerrero Bote and Félix de Moya-Anegón described different genetic algorithms that could be applied to relevance feedback using documents found in literature. A vector space model was used and comparison was made with a traditional feedback algorithm previously used in a study by Gerard Salton. The experiment was carried out using the Cranfield collection. Two other papers were presented by Félix de Moya-Anegón on behalf of colleagues. "An Inductive Query by Example Technique for Extended Boolean Queries Based on Simulated Annealing-Programming" addressed the problem of the user who does not know the terminology of a subject needed to express a valid query. The system described would automatically derive queries from a set of relevant documents supplied by the user. A paper entitled "Graphic Table of Contents (GTOC) for Library Collections" introduced the application of UDC codes for subject mapping. Multi-dimensional scaling (MDS), cluster analysis and neural networks were the methodologies used. A conclusion describes the viability of the application of such an approach. In the final paper in this session on artificial intelligence, Luis de Campos, Juan Fernández-Luna and Juan F. Huete provided a brief survey of the literature on applications and models for "Managing Documents with Bayesian Belief Networks." These tools deal with the problem of "uncertainty" in information retrieval.

In the third session on "Applications of Artificial Intelligence," there were four papers, all prepared by groups of researchers from the University of Granada, and the Universities of Jaén and Extremadura in Spain. Included were papers on "Using Neural Networks for Multiword Recognition in IR;" "Artificial Neural Networks Applied to Information Retrieval" and two papers using fuzzy set theory: "On the Evaluation of XML Documents Using Fuzzy Linguistic Techniques" and "Fuzzy Logic for Measuring Information Retrieval Effectiveness."

Two sessions focused on "Integration of Knowledge in Multicultural Domain- Oriented and General Systems." Significant work is being done in this domain and the eight papers on this topic were presented in two sessions. The papers took a variety of approaches to multi-

lingual problems. "Knowledge Organization in a Multilingual System for the Personalization of Digital News Services" by Antonio Garcia Jiménez, Alberto Díaz Esteban and Pablo Gervás was concerned with sending periodic news selections to subscribers of digital newspapers by means of electronic mail. It describes a system called "Hermes" that applies existing techniques from the field of text classification and categorization. Based on user profiles it selects items from different newspapers in different languages (in this case Spanish and English) – that is, a kind of automated bilingual SDI system which also allows for relevance feedback. Similarly, in "Knowledge Representation of Gender Studies on the Internet," Maria López-Huertas and Mario-Guido Barité Roqueta focus on retrieval of information in a variety of languages. One domain, "gender," was chosen and eight search engines, two international, one international affiliate and five local (in Spanish) systems were used. The research took two approaches – qualitative and quantitative. Analysis based on retrieval on the word "gender" was carried out, using the structural and semantic contextual analysis of the search engines. Not surprisingly, the international search engines were superior. And some search engines performed inconsistently. Victoria Frâncu, in her paper on "Language-Independent Structures and Multilingual Information Access," was concerned with the need for the creation of adequate tools to enable efficient retrieval of information across "geographical, linguistic and cultural boundaries." She suggested a classification system, such as UDC, as a solution to the problem. The paper is a case study that describes in detail a multilingual access system that uses two approaches to information retrieval languages – the mapping of a traditional classification system onto an interdisciplinary thesaurus and the use of either UDC codes or UDC-based descriptors in various languages. Languages included in the project were English, French and Romanian. In their paper on "Models for Collaborative Integration of Knowledge," Annelise Mark Pejtersen and Hanne Albrechtsen introduce a cognitive engineering approach to modelling collaborative integration of knowledge in work domains. A generic means-ends model was used to provide a theoretical foundation, and a decision task model was used in decision making. The problem of the integration of knowledge was explored in a web-based study of collaborative film indexing. Richard Smiraglia's paper took a somewhat novel approach to the investigation of multiculturalism. His paper, "Crossing Cultural Boundaries," provided perspectives on the popularity of creative works (known variously as opera, oeuvres, works). The investigation was based on evidence that the popularity of works contributes to "mutation and derivation" of their intellectual content and that the more popular the work, the more likely change will be observed over time. In this context, Smiraglia studied a sample of best-selling works (fiction and non-fiction) from 1900 to 1999. This is a work in progress, but preliminary results suggest that his basic assumption will be supported. In "Some Patterns of Information Presentation, Organization and Indexing for Communication Across Cultures and Faiths," A. Neelameghan and Hemalata Iyer see devices such as indexes, glossaries of terms and presentation of text in different languages and scripts as aids to communication across cultures, faiths and linguistic boundaries. Their research used documents from the spiritual domain.

As might be expected, the conference also included a session on "Epistemological Approaches to Classification Principles" in which there were several papers on "design and construction." The presentation by Birger Hjørland addressed the state-of-the art in "The Methodology of Constructing Classification Schemes." He postulated that previous methodologies neglect the empirical basis of classification and provided evidence for the necessity of historical and pragmatic methods. His presentation covered the methods of constructing classifications from Ranganathan to the design of ontologies in computer science. Further he referred to the recent 'paradigm shift' in classification research. Hope Olson, Juliet Nielsen and Shona Dippie discussed "Encyclopaedist Rivalry, Classificatory Commonality, Illusory Universality." In it they describe the cultural construction of classification as seen through the work of two French Encyclopaedists, Jean d'Alembert and Denis Diderot, and the work of Samuel Taylor Coleridge. As presented at the conference, their work is part of a larger study which is exploring the cultural construction of classification and constitutes an example of the first phase of the project which is "a deconstruction developed from relevant texts." The texts are being encoded and when the encoding has been fully developed "it will allow analysis of texts from multiple perspectives." In this same session, a paper on "Evolving Paradigms of Knowledge Representation and Organization" by Jain Quin, provided a comparative study of classification, XML/DTD and ontology in which she based her comparison on "how they specify concepts, build data models and encode knowledge organization structures." Differences and relations in knowledge organization were examined and four paradigms were generalized – inte-

grative and disintegrative pragmatism, and integrative and disintegrative epistemologism. One of the findings in the comparison was the indication "that more recent approaches to knowledge representation and organization are developed using the foundations established by precursors." Significantly, this finding mirrors comments and results from a number of presenters at this Conference. In this same session, I.C. McIlwaine's paper, "Where Have All the Flowers Gone? An Investigation into the Fate of Some Special Classification Schemes," took the conference back to the roots of classification theory. The Classification Research Group (CRG) was a driving force in the creation of "a framework and body of principles that remain valid for the retrieval needs of today." Highlights of this paper were the history and background of the CRG and its special schemes and fundamental principles that remain valid. In contrast to this, Jens-Erik Mai asked the question "Is Classification Theory Possible?" Mai is "Rethinking Classification Research" and questions "the assumption that bibliographic classification theory can resemble scientific theories." In doing so, he argues that expertise is not achieved through the correct application of classification theory but that it is the more practical experience and adjustment to context that operates in the real world.

Perhaps the most unusual session of the conference was the session on "Professional Ethics. Users and Information Structures. Evaluation of Systems." It was unusual in two ways. First of all "ethics" is a topic that has seldom, if ever, been a major topic for discussion at conferences of this kind. Will this be a new trend? One can only surmise that ethics has become an issue because of globalisation, national and international sharing of databases, the Internet and advances in technology. Secondly, unlike the other sessions, this one dealt with three distinct topics – ethics, users and evaluation. As for the mix of topics, it is clear that there were insufficient papers focused directly on these topics to warrant their own independent sessions. Three of the papers focused on the ethics issue. "Ethical Aspects of Knowledge Organization and Representation in the Digital Environment" by J. Carlos Fernández-Molina and J. Augusto C. Guimarães aimed to identify the principal ethical values related to the representation and organization of knowledge and to find out to what degree these values are addressed in the codes of the professional organizations. Findings suggested that ethics are addressed but not in sufficient detail and that there is need to deal with the problems that have come about with generous sharing of information and the advent of the Internet. A second paper, "Knowledge Organization from a 'Culture of the Border,'" presented by Antonio García Gutiérrez dealt theoretically and metaphorically with multiculturalism and the problems of transcultural ethics. Then Christopher King, David Marwick and M. Howard Williams, presented a paper entitled "The Importance of Context in Resolving of Conflicts When Sharing User Profiles." Their concern is that while user profiles have been an important factor in the everyday life of information retrieval for some time, such profiles are being "assembled by information providers with no regard to future sharing" of this information. Their paper discusses the problems involved and a prototype has been constructed to examine their ideas on this topic. Two of the papers in this session focused on users. "Ariadne's Thread: Knowledge Structures for Browsing in OPACs" by Carmen Caro Castro and Críspulo Travieso Rodríguez presented a discussion on subject searching with particular reference to "how users' expressions match subject headings" and to determine whether "knowledge structure used in online catalogues enhances searching effectiveness." In a somewhat different look at users, Ali Asghar Shiri, Crawford Revie and Gobinda Chowdhury presented a study "Assessing the Impact of User Interaction with Thesaural Knowledge Structures." This was a pilot study to determine the extent to which a thesaurus-enhanced search interface to a web-based database aids end users in their selection of search terms. Finally, Linda Banwell explored the issues surrounding "Developing an Evaluation Framework for a Supranational Digital Library" as seen through the TEL (The European Library) project. The project, funded by the European Union, is a project to set up a co-operative framework for access to the major national, mainly digital collections in European national libraries. As described by Banwell, the focus is on "consensus building, and also includes preparatory technical work to develop testbeds, which will gauge to what extent interoperability is achievable." Her paper describes the work on the project to date and seeks "to establish what are the key drivers, priorities and barriers encountered in developing such a framework."

While the conference was presented in what were intended to be clearly defined sessions, these sessions were far from being mutually exclusive and many topics, problems and issues often spilled over into other presentations and discussions, in that they permeated the whole conference. At the beginning of the

conference twelve participants were invited to act as rapporteurs or relators. At the end of the conference their views were brought together as a summary and presented in a wrap-up by Ia McIlwaine. In conclusion, a selection of those comments are included here. Clearly, the thesaurus predominates as the favoured tool for retrieval and the concept of ontologies is becoming more and more prominent. However, as McIlwaine indicated, "under every good thesaurus there should lurk a classification scheme." And while a number of papers "showed an awareness of tensions between classification as a culturally specific activity and classification as an activity based on universal concepts," classification, in particular facet analysis, is still fundamental to knowledge organization and representation of knowledge. Not surprisingly the problems of interdisciplinarity, the need for interoperability and the shortcomings of the traditional approaches to knowledge organization were all in evidence and highlighted by several contributors. The desirability of capitalizing upon the expertise of all branches of the information world, and the experience of those in libraries, archives and museums was stressed. The representation of what Elaine Svenonius calls "aboutness" and the challenge of coping with creative thought and theoretical innovation in the humanities and the social sciences were explored. The possibility of combining classificatory approaches with a mark-up language, such as XML, was investigated and its potential as a tool to assist with problems posed by interoperability between systems was explored. Also, artificial intelligence was seen as an important player in future information systems and, of course, the Internet was not forgotten.

The theme of many sessions included the words "Models and Methods for Knowledge Organization" which, to some extent, could be claimed to be the underlying intention of the whole conference. As described by Ia McIlwaine in the wrap-up, another predominating feature was the reporting of individual projects that have been carried out by discrete groups of people or individuals. Of course, proper attention has been paid to the work of others in the field, but the prevailing problems of retrieval in a networked environment do call for some corporate activity across international boundaries and this is not yet very evident. Many of the enterprises that we heard about have rejected the tried and tested approaches in favour of ontologies, taxonomies and approaches favoured by computer scientists.

In summary, the principal themes that emerged are the fundamental need for facet analysis and a plea for "back to basics," together with the use of categories and a systematic approach to the design of retrieval tools. The general preference is for the thesaurus as the favoured tool for subject retrieval, and the sad lack of up to date standards for the creation of retrieval tools has been commented upon. Related studies in Artificial Intelligence and in lexicography and the need to draw upon others working in parallel fields, and to make our methods known to them, were emphasised, especially in the design of ontologies for accessing data on the Web. XML has featured as a useful tool on a number of occasions. Social and ethical problems and the needs of users from different backgrounds and at different levels of sophistication were emphasized. Overarching all of the papers there was the challenge of effective retrieval of information from the Web and the need for work on interdisciplinary fields and across different cultural divides in order to resolve the problems posed by the organization of resources in a networked environment.