

## The Sixth International ISKO Conference

To conclude the Sixth International ISKO Conference held at the University of Toronto, Canada July 10-13 2000, the organizing committee asked three people to present a summary of papers and themes discussed during the conference. The resulting presentations are printed here.

JEM

### Two Recurring Themes: Universality and Clustering

Rebecca Green

University of Maryland  
College of Library and Information Services  
College Park, Maryland

The introduction to the proceedings of the conference insightfully notes that the papers presented “focus on the need for improvements and refinements in knowledge organization principles and practices in the effort to fulfill the *conflicting desiderata* for both dynamism and stability” [p. xi; emphasis added], thus echoing the theme of the conference. This review of the conference begins by exploring that conflict.

Dynamism bespeaks continuous activity and change. Typically, we may think of change occurring over time, so that there are differences both in what we knew in the past, what we now know, and what we will know in the future, as well as in what we needed to know in years gone by, what we need to know at present, and what we will need to know in times to come. On the one hand, we need to identify and maintain that core of our knowledge base and users’ needs that remains constant and stable over time, but we also need to identify what has changed and continues to change so we can add to and/or transform our organizational tools and systems as necessary.

However, rather than explore dynamism and stability diachronically by looking at change and constancy across time, this review focuses on exploring change and constancy synchronically, across discourse communities based in culture, discipline, or function, as reflected in differences and similarities across languages, ontologies, vocabularies, metadata schemes, and indeed, any type of knowledge organization scheme. The question then becomes one of the extent to which we can

generalize across these various boundaries. In other words, what in our systems, especially our conceptual systems, is universal? And where concepts are not universal, can we – and if so, how can we – integrate conceptual systems that are not fully commensurate one with another?

The papers presented at the conference largely disavowed the notion that conceptual universals abound and opted instead for culturally specific concepts. For example, in her keynote talk Hanne Albrechtsen espoused ecological classification systems, systems ‘designed to be used within a particular context or environment ... that do not strive for universality across domains’ (p. 1). Alexander Sigel’s paper downplayed the problems of indexer inconsistency by noting that variant indexings of a document may just reflect ‘alternative conceptualizations’ deriving from the viewpoints of different groups. And Hope Olson, among quite a few others, asserted the cultural specificity of classifications, calling into question the universality of some of our most fundamental – and perhaps cherished – principles of classification.

No sooner do we recognize the relative separateness of various contexts, however, than we start working to put all the pieces of Humpty Dumpty back together again. Many of the presenters at the conference have discussed the challenges of establishing equivalences across natural languages, across controlled vocabularies, across ontologies, as well as postulating particular approaches for doing so.

For example, the talk by Widad Mustafa el Hadi concerned cross-language information retrieval, a task that assumes a user will query the retrieval system in one language, while wanting to retrieve documents in another language. A vexing shortcoming noted is that polysemous terms may be translated incorrectly when only shallow linguistic processing is done.

Gerhard Riesthuis addressed the guidelines for multilingual thesauri. He illustrated equivalence problems in trying to map between the vocabularies of different

languages, including problems of inexact equivalence, partial equivalence, single-to-multiple equivalence, and non-equivalence. His examples pointed to a rejection of the hypothesis that descriptors in any one language should have equivalent descriptors in all other languages covered by a multilingual thesaurus.

Victoria Frâncu discussed related issues involved in the design of an interdisciplinary multilingual thesaurus (covering Romanian, English, and French), based on an abridged version of the UDC. The equivalence problem between languages is dealt with in part by substituting the problem of how each language represented in the thesaurus can be made compatible with the underlying classification scheme. The low level of specificity reflected in the system's preferred terms helps alleviate some of the problem of cross-linguistic integration.

Several of the talks addressed issues of equivalence within a single language, although within the context of translation or transliteration. Lynne Bowker, for example, took on translators and terminologists alike in addressing terminological variation in the literature of medicine. She noted that translators must be sensitive to subtle differences between seemingly synonymous terms in one language and not enforce a principle of standardization upon them when translating into another language.

In a well-crafted study Clément Arsenault took up the vexing issue of word segmentation in two Romanization systems for Chinese, Wade-Giles and *Hanyu pinyin*. Not only do the schemes differ in how various sounds are rendered in a Roman script, but there are also differences in word segmentation. An evaluation of known-item searching in an OPAC by native language searchers who had to "translate" between Chinese characters and the various Romanization systems revealed that users do not use the three of them equally effectively.

Mapping between vocabularies would seem to be easier within a single language. However, even within a single language and a single subject domain, there can be problems of vocabulary integration. Carol Bean explored the alignment of English language medical terminologies. In cases where users were unable to locate terms in the Unified Medical Language System (UMLS) equivalent to terms from other vocabularies, they mapped most often to broader terms, but also mapped at times to narrower terms. Bean analyzed the specific semantic types involved in mapping down. Hierarchical/Subsumption relationships are thus seen to play a role in equivalence relationships, echoing a point made in Frâncu's paper.

Robert Kent's paper addressed conceptual integration in ontologies. He stated that the stable aspect of ontologies is based on their types and constraints, while the dynamic aspect is based on instances and community-specific classification relations between instances and types. Ontology sharing – resulting in a common, generic extensible ontology – is based on the terminology and semantics common to multiple community ontologies.

As suggested before, the papers presented at the conference tended to promote the distinctiveness and specialization of language and classificatory structure across domains. But this was not the case in every instance. Indeed, some of the papers quoted earlier regarding the need to design specific classificatory structures for distinct user groups also explicitly called for investigation of the constant elements of those structures. Again hearkening back to the keynote talk, Hanne Albrechtsen called for 'empirical analysis of the stable or invariant structures of a particular domain' (p. 1).

But why should we assume that only particular domains have invariant structures? Why not pursue empirical analysis of the stable or invariant structures of any and all domains? If we are willing to adopt the assumption that our language reflects (and also molds) our thought, then the empirical investigation of linguistic universals should become a benchmark against which to answer the question that underlies our theme: To what degree can we build stable and invariant knowledge organization schemes that cross the various boundaries of culture, language, discipline, function, *etc.*?

Winfried Schmitz-Esser proposed the development of an ontology, whose global semantic validity is challenged on the one hand by volatility with regard to *instances*, and on the other hand, by the detection of newly discovered or acknowledged *universals*. These universals are 'concepts or themes or topics that are common knowledge in civilized communities' (p. 84); further, multilingual ontologies are made possible by the ability of modern languages to express new and unknown concepts clearly and unequivocally on the basis of existing universals. (However, it is not clear how this model could accommodate itself to the social sciences and the humanities, where the set of concepts that constitute 'common knowledge' is rather more limited than in the sciences and technologies.)

Clare Beghtol's paper on taxonomic and partonomic relationships – division by kinds and parts – also explored the universality question. After first noting the predominance of the assumption that classification schemes are culturally based, culturally biased, and non-

universal, and then noting emergent efforts to develop practical methods for integrating diverse points of view, she asserted our need to '[investigate] possible conceptual universals that can be used as foundation elements in classification systems for any culture or domain' (p. 313). She then cited the work of Goddard and Wierzbicka, who have 'identified eight categories of semantic and lexical universals [ - 38 specific ones - ] that obtain across a large number of [diverse] languages.' Some of these specific universals are very broad, e.g., DO, HAPPEN, NO, KIND OF, but some seem more specific, e.g., TWO, UNDER, AFTER, THINK. For the most part, the universals are relational in nature. The substantives in the list, are either pronouns (I, YOU) or very general (SOMEONE, SOMETHING, PEOPLE). This suggests that we could profitably exert effort in trying to craft more universal approaches based on relationships, a point echoed by Schmitz-Esser's presentation. It also suggests that in moving from earlier assumptions of culturally neutral, universally applicable classification schemes to assumptions of culturally biased, non-universal schemes, we may have overreacted.

Turning our attention from classification to clustering, we note in both cases an equal concern with the formation of some sort of equivalence sets. Classification consists of pre-identifying a set of categories into which objects of interest are sorted; the same classification scheme can theoretically be used for many different sets of objects. Clustering, in contrast, consists of grouping objects of interest on the basis of their characteristics; the resulting categories emerge from the objects, so that different sets of objects generate different categorizations.

At least two of the papers concerned the clustering implicit in the notion of bibliographic works. Richard Smiraglia gave an epistemological characterization of works, likening them to Saussure's signs, with ideational content (the signified) conveyed by symbols (the signifier). Works were also characterized as members of a social canon, with contextually interpreted meanings. Allyson Carlyle & Joel Summerlin took the categorization process a step further, examining the degree to which manifestations of works could be grouped into subclusters automatically on the basis of, for example, keywords in designated fields of the MARC record. In the sample of records they examined, the vast majority of records could be sorted into smaller groups on the basis of language, medium, presence of illustrations, etc. Admittedly, the process they described is really a classification process, but it does operate on clusters, i.e., works, and reflects a common use of clusters in online

searching, namely, as a device to help users cope with large retrieval sets, by presenting such a set clustered in some number of smaller retrieval sets of related items.

This is also the strategy of the CATHIE system, presented by Majid Ihadjdene et al. CATHIE facilitates filtering of retrieval sets by suggesting to the user those terms - subject headings, DDC captions - most commonly associated with the retrieval set. This prompts the user to reformulate his or her query, in essence by selecting subsets of the retrieval set, clustered on one or more of the assigned subject headings or DDC classes.

More elaborate techniques for forming clusters of documents were presented in the paper by Xavier Polanco and Claire Francois. They used an axial k-means algorithm to group records based on their keywords or index terms, into overlapping clusters. They then adopted the use of a multilayer perception and related components analysis to transform the clusters into a two-dimensional map exhibiting relatedness between the clusters and thus reflecting a higher-level organization governing the clusters.

Igor Jurisica discussed the use of case-based reasoning as a dynamic approach to knowledge organization. Case retrieval systems can be optimized by organizing the case base into context-based clusters. Clustered cases are thus more likely to be relevant to the same incoming problem descriptions.

Félix de Moya-Anegón & María López-Huertas presented work in which author, journal, and article co-citation data were used to form clusters of specializations within biotechnology, thus yielding a dynamic view on the conceptual structure of the field.

Gobinda Chowdhury et al. used co-word analysis to form clusters of keywords capable of functioning as an associative thesaurus. The standardization process involved in the pre-processing of their data, in which keywords were mapped manually to descriptors in controlled vocabularies, will need to be automated - an area of ongoing effort - if this method is to be of general usefulness.

Research reported by Rebecca Green used clustering to group the senses of different verbs that evoke a common relational structure or frame. A subsequent step attempts to identify the roles associated with the frames for use in developing classification schemes or ontologies.

Michael Buckland et al. presented ongoing work from a UC, Berkeley project on the generation of entry vocabulary indexes to knowledge organization schemes; this project is based on the premise that sub-domain vocabularies are distinct. The clustering in this

research is akin to the integration of medical vocabularies taking place in the UMLS project of the U.S. National Library of Medicine, though in this project the “vocabularies” being integrated are loosely structured sets of natural language terms, not controlled vocabularies, and the integration is more implicit than explicit. Statistical association techniques are used to find domain-specific correlations between words and phrases found in titles and abstracts of documents and the index terms or subject categories assigned to those documents. Which terms or categories of a knowledge organization scheme best represent a natural language query can then be predicted on the basis of these associations. Application of the method does require access to prior indexing and classification assignments, but such data should be readily available in precisely those cases where it is most needed.

In addition to the seemingly ubiquitous clustering theme, there were several other recurring themes among papers treating new information technologies. The Internet, of course, was everywhere. Although search engines are a critical part of the Internet experience, they did not receive extensive treatment during the conference. (As a group, we seem not to be much impressed by them.) Exceptions included Martin and der Walt’s evaluative survey of South African search engines, search directories, and portals and Vanda Broughton’s comparative discussion of retrieval systems based on classified, linguistic, and symbolic approaches, in which she proposed several potential applications of faceted classification to retrieval on the Web.

Another theme echoed in several papers concerned the benefits of mark-up language, with XML (Extensible Markup Language) being fully embraced. (1) For instance, Joan Mitchell and Diane Vizine-Goetz’s paper noted that records in the Dewey editorial database have been marked up using a proprietary set of meta tags. On the one hand, the mark-up helps guide the production of printed output. On the other hand, it can be used to partition the data into subsets appropriate for specific tasks. (2) Grant Campbell discussed possible points of intersection between mark-up languages and principles of classification. The association between XLL (Extensible Linking Language) and facet analysis would seem to represent a fruitful collaboration. (3) XML also played an important role in the paper by Elizabeth Davenport & Howard Rosenbaum, in which mark-up is used to capture the ‘shape’ of documentary genres in a system for organizing situational knowledge. (4) And Lynne Howarth proposed the develop-

ment of a tool that would provide front-end search assistance to XML-compliant metadata schemes.

Another theme addressed in a small handful of papers concerned interfaces. In addition to papers that addressed the filtering of online retrieval sets, already mentioned, there was a paper by Jon Dron *et al* on COFIND, a collaborative bookmark engine or resource base.

Last, but by no means least, we come to the work of Horacio Saggion & Guy Lapalme on the automatic generation of summaries, based on analysis of the relationships between human-generated abstracts and the texts they represent. A novel feature of their system is the mixture of indicative and informative elements. A user presented with an indicative abstract may request further information and is then presented with text spans that elaborate the topic.

It is not surprising that many themes would run throughout a conference such as this. Indeed, what does perhaps surprise is how many of the papers touched on, whether openly and explicitly or only by way of assumption, some of the same themes. By way of summarization, a large number of papers treated the issue of universals vs. cultural specialization. Here no extreme answer will be suitable. Therefore there is the need for further exploration to determine where the most realistic in-between positions are. Additionally we should note what will likely be a growing trend to develop categories out of the data rather than impose pre-established categorizations on the data.

---

## Reflections on Papers at the 6th International ISKO Conference

Jens-Erik Mai

Information School, University of Washington  
Seattle, USA

---

There seems to be one theme that penetrates all the papers we have heard at this conference, namely the urge for finding and developing new approaches to studying and understanding knowledge organizations. In my comments here I will try to summarize my impressions of some of the themes discussed at this conference. I will focus mainly on the papers delivered in the two tracks named “Theories of Knowledge and Knowledge Organization” and “Culture, Language, and Communication in Knowledge Organization.” I will,

however, not comment on all the papers in these tracks, nor will I limit myself to these tracks. I apologize in advance for any misinterpretation I might make of the papers and presentations and for sometimes focusing on some minor aspects of the papers.

In her keynote address, Hanne Albrechtsen cites Svenonious' (1992) keynote paper at the "5<sup>th</sup> International Study Conference on Classification Research," where she argues that classification research, to a large degree, has focused on the development of practical principles for classification scheme design. The underlying assumption of many of the papers that I will comment on is that these principles by and large have been developed sufficiently and have been exhausted for further theoretical research. The title of this year's conference, *Dynamism and Stability in Knowledge Organization*, suggests that the focus of the conference was on the circumstance rather than the techniques and principles. That said, I want to acknowledge the papers that have focused on different aspects of faceted classification schemes (such as Priss, Devadason, Broughton) and the papers concerned with multilingual thesaurus construction (such as Mustafa el Hadi, Riesthuis). They have added tremendously to our knowledge about these important systems and techniques. The title of the first ISKO conference, *Tools for Knowledge Organization and the Human Interface*, suggests that the focus of that conference was on techniques and practical principles.

The argument that more research on the processes and circumstances of classification is needed was raised a number of times during the 90s, when future research in indexing and classification was discussed. The American Society for Information Science's Research Committee (Shaw & Fouchereaux, 1993) identified in the early 90s areas within library and information science where more research was needed. One of these areas was "what are the cognitive processes involved in indexing and classification?" Milstead has likewise discussed the need for more research in indexing. She noted the importance of research into the mental processes involved when an indexer decides what a piece of information is 'about'.

Many of the papers at the conference have in fact been concerned with the wider context of indexing and classification. However, the majority have not asked what the mental and cognitive processes are – but instead have suggested that more research into the importance of studying the social environment and gaining knowledge about this area before we start using the practical principles and techniques is needed.

Breitenstein in particular explores the boundaries of classification theory. The author suggests that classification theory has been heavily influenced by modernism's belief in a 'grand plan.' A plan where everything is organized and which is out there somewhere only waiting to be discovered and displayed in a classification scheme. The author therefore stresses two other frameworks for studying knowledge structuring. The first is culture studies, where focus is on the reality of situated knowledge. The second framework is individual experience and the fact that the individual is the smallest element of action. Classification cannot be studied or understood without including knowledge about these two frameworks.

In a closely related paper Jacob argues that the belief in the 'grand plan' or grand narratives is based in a modernistic scientific tradition which has been under attack in various disciplines during the past few decades, under the heading post-modernism. The paper repeats the argument that this post-modern criticism of modern science was raised by the American pragmatics a hundred years ago. She suggests that traditional classification theory suffers under the ideologies of modern scientific thought and that pragmatism could be used as the framework for a more 'modern' theoretical framework for classification theory. It is quite interesting to note the growing interest in relativistic theories within classification research. Much of this post-modern criticism was raised 25 years ago and one therefore only wonders why it has not reached the classification research community before.

However, it seems as though things are changing now. A number of presenters focused on new frameworks for classification theory. One group of researchers, Pejtersen and Albrechtsen, suggest an approach which they call "ecological work based classification schemes" that is based on perception theory. The aim is to analyze the invariant structures of a complex work domain and the information needs of its actors in order to produce a transparent and structured information environment. Another group of researchers, Cardoso, Bemfica and Borges, suggests using the concept of auto-poiesis as a meta-theoretical approach to understanding knowledge, knowledge organizations, and social organizations. In such an approach, the myth of absolute objectivity will fall and thus provide room for a new understanding of knowledge and knowledge organization.

In a related paper, Christensen discusses how the French sociologist of science Latour's view of science and the production of truth influences thinking in

knowledge organization. The author offers an alternative definition and purpose of knowledge organization, which I like very much. He says that "Knowledge organization is the organization of statements in a discussion, and should reflect not only what is said but also who says it and why. It is the librarian's task to deliver this transparency" (p. 310). This definition emphasizes that the social aspects of knowledge organization cannot be ignored!

Another group of papers has been concerned with understanding and enhancing existing classification schemes, and with how to build strong classification schemes. Fallis and Mathiesen, for instance, defined a set of consistency rules for classification schemes. They suggested that these rules could potentially form the basis of measuring the degree of consistency of a classification scheme. I find the paper interesting because it re-phrases the already known rules and principles for development of classification schemes and offers a new vocabulary for talking about consistency in classification schemes. Although I liked the Beanie Babies example I would have liked to see how the consistency rules could be applied to a bibliographic classification scheme. In a related paper Beghtol discussed the problems of consistently identifying the different kinds of subdivisions that are applied throughout a classification scheme. Her paper argues that any theory of classification should accommodate the potential ambiguity that follows from the fact some concepts are not mutually exclusive. The observation that concepts are not mutually exclusive should form the basis for any theory of classification. If we have no way of handling meaning and concepts, rules of consistency will only be of little help.

The Internet has created a new environment for organizing a different media. Hudon notes that the Internet community so far has not regarded the 'traditional' methods of construction of classification schemes and thesauri "appropriate and/or necessary" (p. 35). From the examples given in the paper it is obvious that many of these web sites would benefit from an introductory course in knowledge organization! However, Williamson suggests that the standard guidelines for development of thesauri need to be updated. The presenter felt that the present guidelines are inadequate for the digital age. Neelameghan argued that Ranganathan's principles are general enough to also include other media than the books he speaks of in his five laws – there is no reason to invent new theories and new methods for knowledge organization just because we face new media. The argument that the practical principles that were devel-

oped and have been polished during the last century are strong enough to face the challenges of the 21<sup>st</sup> century is good news. It should however be clear that there is a huge challenge in making the principles and reality meet, especially when the reality is as dynamic as the Internet. My hope is that our graduates actually do go out and get those high paid jobs in the information industry and start implementing the principles that we have taught them!

One major factor for this to happen is that we develop a theory for practice to support the development of knowledge organizations; this theory of knowledge organization should somehow comply with the criticism that has been raised toward to the traditional theories of classification. I believe that the main focus for this criticism has been the traditional theories' belief in neutrality and objectivity. Solomon actually summarized some of this criticism and related it to the problems of organizing material on the Internet. The presenter suggested a very interesting 'stew' of theories that could form the basis for developing a new theoretical foundation for knowledge organization that focuses on contextualization.

Campbell notes that the principles of building classification schemes have received attention when designing markup languages. It is suggested that the principles of classification can be used not only to organize documents but also to make the internal structure of the documents visible. The same paper, however, indicates that we need to redefine our concepts of "document" and "collection" in the light of new technologies. I am not too sure that is necessary – I don't think that there is such a great difference between paper-based collections and documents and electronic collections and documents. Or in other words, the difference isn't great enough to undertake a conceptual exploration of these concepts. In a related paper Davenport and Rosenbaum argued that by incorporating genre analysis into XML codes, a retrieval mechanism that allows for retrieval of sets of activities and therefore gives access to situational knowledge is possible.

My feeling is that the field is faced with a dilemma. It must focus either on developing new features and techniques or on understanding the socio-cognitive circumstances for the development of classification schemes. My hope is that these approaches will merge in the near future.

I also want to emphasize the historical papers we have heard. We were, for instance, reminded of the debate of whether Bliss or Ranganathan invented the notion of synthesis in faceted classification by La Barre.

Furner showed just how influential the Classification Research Group has been on knowledge organization and information science in general.

Before I finish I would like to go back to theme of the conference, *Dynamism and Stability in Knowledge Organization*, and ask what we have learned during the past few days. I would say that we have learned that of course there is dynamism in knowledge organizations. It is inherent in the very definition of a knowledge organization. And no, there can never be any stability in knowledge organizations. But we can work at understanding the conditions that create this dynamism--not to control it--but to understand it, so that we can build theories and knowledge organizations that are based on insight and not on some dream about an objective universe of knowledge that is out there waiting to be discovered and displayed. This, however, furthermore suggest that future research agendas within knowledge organization should focus on studying the social and cultural aspects of knowledge organization and show how insight into this could enhance theory and practice.

---

## Closing remarks for the Sixth International ISKO Conference

Ia C. McIlwaine

University College London  
School of Library, Archive and Information  
Studies, London, United Kingdom

---

About two years ago Nancy Williamson and I conducted the investigation into the developments that had taken place in research on subject analysis over the preceding ten years (McIlwaine & Williamson, 1999), referred to by Hanne Albrechtsen in her keynote address. Many of the conclusions we made then still appear to be valid at the end of this conference. We were of the opinion, after an extensive examination of both the journal literature and papers presented at a range of relevant conferences, including those of ISKO, that there was great interest in the thesaurus as a means of retrieval, and that the principles underlying thesaurus construction might form a major contender as the most discussed topic. We also noted that many of the papers offered to international conferences discussed local research projects, of great interest to the community for

whom they were designed, but without relation to the greater outside world and the use that might be made of them more generally. Finally, we noted a predominance of contributors from the world of academia, with much less participation from practitioners in the field.

The past few days have not greatly altered the overall impression that we reached then. The keynote speaker pleaded for more co-operation – meetings such as this is one way of stimulating it. There remains scope for greater co-ordination of results towards more generally applicable solutions to the problems of retrieval in the twenty-first century, both in terms of reaching more universally applicable solutions and in a mixture of the “real world” with that of the school of library and information studies. It is encouraging, however, to note several pleas for more co-ordination and coherence in arriving at more far-reaching and universal solutions.

We tend to be too parochial in our outlook, and not only in relation to individual projects, as this is endemic to our discipline as a whole. We should take greater note of what is going on in other fields, such as Biology, Pathology and Pharmacy, to mention but three relevant disciplines where much work on the organization of knowledge takes place. Several speakers have suggested that they are different, and not relevant to our concerns. This is not entirely true. There are many people interested in classifications and in resolving the problems relating to both organization and retrieval who are working in different disciplines and from whom we can learn much. It is therefore good to have this brought to our attention by some of the participants here in Toronto. The classification of patents has always had close connections with bibliographic classification, but it is the first time in a number of years that we have heard them discussed at a conference such as this. The example of ornamental design (Rademaker) as developed by the US Patent Office provides a number of parallels with the more general problems encountered by information workers today. The division of the field into various facets such as properties, entities, function, etc. had a familiar ring to many of us.

Similarly, we saw another example in the account of the procedures undertaken by the UK Treasury when it re-organized the system of its files in the 1920s (Craig) in order to cope with what was at that time seen as the “information overload” in the material being produced by central government, which called for a radical restructuring of procedures after the First World War. Interchange of ideas with people working in different though related fields is vital to maintain a “living” approach to solving the problems of the 21<sup>st</sup> century and

to ensure that there is a sharing of ideas and an avoidance of reduplication of effort. We should not forget the past, for a favorite pastime today seems to be reinvention of the wheel and we have been reminded by several speakers, for example Furner, of many pioneers who laid the foundations for our present work.

We have heard a paper dealing with the background to the Colon Classification and the problems of personal relationships between Ranganathan and Bliss (La Barre), two of the principal figures in the development of a proper scientific structure with an accompanying vocabulary to label the concepts that underlie our field. Bliss and Ranganathan indisputably between them provided a grammar and terminology but the terms that they used were not the same and this has led to confusion for future generations. Nevertheless, their influence remains as strong as ever, as Furner's survey of citations of Classification Research Group members showed, and Priss drew attention to the underuse of facets by people working in the retrieval field. Campbell, too, drew a parallel between the Colon Classification and XML.

Terminology is of vital importance to those working in the information field. Bowker emphasized this in her paper. We have heard a plea for updated standards for the creation of thesauri and an outline of the proposals to bring this about from the chairman of the working group at present drafting this revision (Riesthuis). Thesauri have figured in many contexts, not just as an aid to web searching, but for the organization of collections of other media such as the moving image (Hudon). There are clear parallels in the two situations, as there are also, of course, with the organization of patents for ornamental design (Rademaker). The use of a thesaurus within a limited field overcomes many of the problems of terminology because there are shared perceptions and understandings, but these vanish as soon as one widens the scope of the database concerned. Frâncu highlighted some of the issues of dealing with a more generalized situation in her paper where she told us about a multi-lingual thesaurus in Romanian, French and English based on the structure of the UDC which she argued provides a good starting point, since it has great facility for the expression of relationships. Greater problems arise when one is not only in a multilingual, but also a multi-script environment. The paper by Arsenaull illustrated this point by discussing script conversion, and the migration from Wade Giles to Pinyin. His findings serve as a warning when we think we have found solutions because although we may manage to convert from one script to another successfully, we

may yet fail to convert the presentation of the groupings of letters accurately and meaningfully, with resultant problems for filing and searching.

The Internet, inevitably, dominates the thoughts of those who are seeking to improve methods of access to the information that floods our screens and to create some order out of the chaotic situation. Attempts to find solutions are based on two schools of thought. There are those who feel the sensible approach is to take as a starting point the tried and tested methods. This means the use of classification, either systematic, as in the traditional bibliographic schemes used in libraries for over a century, or of thesauri which, if properly constructed, are simply classifications turned round and reduced to an alphabetical arrangement. The second school of thought prefers to rely on words alone, and the matching of keywords through automatic analysis and the use of bibliometric methods in order to cope with the flood material which far exceeds anything that can be handled by humans. This view results in all the limitations imposed by the use of language as the key to the solution rather than structure and context.

The thesaurus has been demonstrated as successful in a range of specialisms. Its success becomes more limited, however, when one moves into the world of interdisciplinarity. A major problem with interdisciplinary thesauri is overlap and the need to distinguish the different meanings of the same term. A possible solution recommended both by Frâncu and Williamson is a hybrid approach using both classification and the thesaurus. Interdisciplinarity has been widely discussed. Buckland rightly notes that every community and every specialty has its own specialist vocabulary but that the situation becomes far more complex in a more general context. Similar difficulties arise when one attempts to accommodate newly arising fields of study, be they environmental science or tourism (McIlwaine) or homosexuality (Huber and Gillespie). Here, the issue is the fact that they are not really new, but are a different approach to what has always been there. Difficulties do not arise with the totally new, (examples such as electronics, computer science, etc. for instance) which are manageable because there are no preconceptions. It is the so-called new studies which are in actuality the result of a coming together of a number of discrete topics under one umbrella that are so difficult to manage. Huber and Gillespie clearly make this point when they list the different aspects in the body of knowledge that delineates gay and lesbian health care. There are cultural and linguistic implications as well as a range of disciplines from the social to the medical to be accommo-



dated. The role that culture and context play in the Knowledge Organization was touched on in Albrecht's keynote address when she discussed the conflicting roles of classification systems as facilitators versus barriers to the activities and evolution of information ecologies. Breitenstein, too, sees culture as framing approaches to the creation and the accessing of knowledge. The most unpredictable of all the elements that make up the approach framed through cultural background, is that of the individual which is "peculiar" in the best sense of the word, unpredictable and brings a unique perception to the discovery of knowledge.

The problems of interdisciplinarity are exacerbated by the Internet and the need to access the information that it makes available, at all levels and through such a diversity of aids. This diversity is very healthy; purists may scorn the lack of structure evident in the search engine and the crude categorization and massive amounts of hits that may be made. But such tools have their uses, in just the same way as categorization commended itself to many a British Public Library, while it would be hopeless for an academic institution. We have been alerted to the pitfalls as well the advantages of such search engines as Yahoo and Northern Light and their advantages in the papers presented by several speakers, including Hudon, Dextre-Clarke and van der Walt.

The thesaurus may seem to be the key to successful access for many, providing as it does the structured approach of a classification, if it is well constructed, together with the favoured mode of access via words rather than apparently meaningless notations. Not everyone, however, sees it in this way, whether it is used in conjunction with classification or independently. Ding, Chowdhury prefers a bibliometric approach, based on cocitation relationships of keywords, at the very least to support the thesaural approach and to provide the opportunity to eliminate "noise". This, however, still results in a thesaurus of sorts, though one arrived at through a different route from the traditional. Ohly, similarly, recommends an approach based on bibliometric principles to accessing and evaluating the information contained in state of the art reports, demonstrating his findings from his own work in the field of nutrition.

An approach that is heavily dependent on automated access unaided by human intervention is also advocated in the work being undertaken by OCLC with automatic classification, building on the second of the Desire projects. We have seen a comparison of results of two trials using DDC and the Engineering Index (Ardo *et al*). At OCLC an accessible taxonomy is being

superimposed upon DDC in an approach similar to that of the Open Directory Project now being used by Yahoo, but employing DDC instead. Dextre-Clarke brings an open mind to both solutions and compares Knowledge Management Systems which retrieve information "automatically" with the traditional thesaurus and rightly points out that the real challenge is to find a viable solution.

The effectiveness of search engines and newly developed methods for retrieving information from the Web is discussed by a number of participants, including Hudon, who takes the opposite view from the likes of Chowdhury and Ohly and argues that these so-called classifications fail miserably, and that it is better to rely on the tried and tested. We have been updated on the current activities of several of these tried and tested classifications. Both the Dewey Decimal Classification and the Universal Decimal Classification have been discussed in several papers and the ways in which they are being used and developed to adjust themselves to changing circumstances have been outlined. The value of the analytico-synthetic approach is demonstrated in the work on the decomposition of numbers. This is something that is being attempted with both schemes, and we saw the success of Pollitt's experiments with the Area notations in DDC. Both these classifications, sharing as they do, a common root, have much work to do before this can be totally achieved, since so many compound concepts have been originally enumerated in the tables and then notated in a manner that suggests to the machine-retrieval device that they are simple concepts rather than compounds. Until these compounded notations have been segmented, in a manner not unlike that needed for the segmentation of Chinese characters (Arsenault) it remains quite impossible to unravel the constituent elements easily.

The success of the work currently being undertaken at OCLC demonstrates the abiding value of the long-standing retrieval tools, such as classifications originally designed for the arrangement of books on shelves. Their adaptability to the present day has been demonstrated through the application of the Dewey Decimal Classification to automated classification of information (Vizine-Goetz), as a means of encompassing the vastness of the Web and the complexity of the information contained therein. The embedding of DDC (or any other) classification number in the metadata is a way of handling the present issues, but although the retrieval is automatic, the original embedding needs human intervention and consequently time.

There have also been pleas for consistency in approach (Hudon), and emphasis on the need to avoid the inconsistencies that are inherent in schemes like DDC and LCC (Fallis), which are at the root of the present need for laborious decomposition. Consistency can only be ensured via standardized procedures which immediately present problems in specialized contexts (History has shown that – hence the range of general and special traditional classifications to be found in our libraries). A major factor in knowledge organization is the need to set a standard which inevitably will always be out of date, will reflect its origins and the perceptions of the society from which it originated and which depends upon a context to make sense to its users. Hanne Albrechtson emphasized this in her opening address when she equated classification schemes with information infrastructures and said that they must be ecological, which she interpreted as meaning that they were inevitably linked to particular contexts and environments. Hope Olson continued in much the same theme, demonstrating her point by reference to Feminism and challenging the reliance upon a hierarchical approach that is evident in all the traditional schemes of classification.

Hierarchy is another concept that has been constantly referred to throughout the past few days, with its defenders and its attackers. A number of papers refer to the essential nature of hierarchy for retrieval purposes, Olson, Chowdhury, Ohly, Carlyle, Bean,

Schmidt-Esser and Vizine-Goetz to mention but a few. Campbell argues that mark up languages give hierarchy prominence and that we are witnessing the reintroduction of the old dichotomy between analytico-synthetic and enumerative classifications, since markup languages permit searching through tags so the internal structure and organization of electronic documents need to be given just as much attention as the relationship of one item to another. Ibadjadene, on the other hand, presents the opposite view in his rejection of the value of an hierarchical approach. He claims that such an approach results in the failure of such search engines as NorthernLight to achieve their intended goal, because they are following essentially the same methods as those traditionally used by classification schemes. He rightly notes that the use of keywords simply produces information overload, and by way of obviating this he proposes a filtering method by which the user is offered words and makes his or her selection so as to ensure greater relevance.

On a final note, one cannot but agree with Solomon who stresses the need for a sound theoretical understanding to underlie all our efforts, together with a realization of the value of change. Perhaps he has squared the circle with his identification of the process that we are all endeavoring as moving from Structures (Knowledge Organization) through Actions (Information Retrieval) and back to Structures.