



**ISKO UK**  
The UK chapter of  
International Society for Knowledge Organization  
<http://www.iskouk.org>  
e-mail: [info@iskouk.org](mailto:info@iskouk.org)



## Report: ISKO UK KOKO Event, 5 November 2007

### Ranganathan Revisited: Facets for the Future.

ISKOUK held its second KOKO (KONnecting KOMmunities) event from 14:00 - 20:00 on November 5th. entitled *Ranganathan Revisited: Facets for the future*. The intention was to explore the current status of faceted classification from both theoretical and practical viewpoints. The event was sponsored by Factiva from Dow Jones and the venue was provided by the School of Library, Archive and Information Studies (SLAIS) at University College London UCL). Eighty-one people attended.

#### Meccano, molecules, and the organization of knowledge: The continuing contribution of S.R. Ranganathan. *Vanda Broughton*

Vanda, lecturer at SLAIS and ISKOUK Chairperson, provided an account of the origins of faceted classification in the work of the eminent Indian scholar and librarian S. R. Ranganathan in the 1930s and described how its influence persists today. Ranganathan himself derived inspiration for his Colon Classification from Meccano, which he came across in a London toy shop whilst studying at UCL in 1924. Vanda, on the other hand, proposed that the molecular model is perhaps a better representation.

The principles of faceted classification continue to contribute to an increasing degree to knowledge organization frameworks, not only in the Library and Information Science (LIS) domain, but also in the domain of digital information. After presenting examples from several e-commerce Web sites, Vanda went on to demonstrate how simple faceted structures commonly occur in general classification schemes such as the DDC (Dewey Decimal Classification) and LCC (Library of Congress Classification), and of course in UDC (Universal Decimal Classification) which is highly faceted.

Ranganathan's five fundamental categories (PMEST) were extended in the 1970s and 1980s by the Classification Research Group (CRG) to 13, but of the major classification schemes, only the second edition of the Bliss Classification (BC2) makes extensive use of them. However, the understanding of faceted classification and the ways in which it is used, can differ greatly among different communities.

The formal principles work best with science & technology, i.e. the 'hard' disciplines, whereas 'soft disciplines' and other areas of application can be problematic. Vanda concluded with a description of how faceted classification techniques were applied in the FATKS project carried out at UCL {JDoc 63(5), 2007, 727-754}.

#### “Classic” vs. “freely” faceted classification. *Claudio Gnoli*

Claudio Gnoli of the University of Pavia in Italy and Chair of ISKO Italy, explored the relative merits of classic 'faceted classification' (FC) and 'freely faceted classification' (FFC). In classic FC, the facets (and their relationships) which might be combined to express a compound subject, are restricted to those prescribed as inherent in the subject area. FC is therefore largely bounded by and restricted to a specific subject area. At the other extreme, free classification (as in the



## ISKO UK

The UK chapter of  
International Society for Knowledge Organization

<http://www.iskouk.org>  
e-mail: [info@iskouk.org](mailto:info@iskouk.org)



Web or folksonomies) allows the combination of values from multiple, disparate domains where the relationships among the elements are often indeterminate, and the semantics obscure. Claudio described how punched cards were an early example of free classification, and cited the coordination of *dogs : postmen : bites* as one where the absence of defined relationships made the semantics ambiguous.

Claudio identified himself as being a member of a band of 'heretics' prepared to regard classification as applying to 'phenomena' rather than the 'disciplines' which were the focus of traditional classification theory. This stance recommends an approach combining the best of totally free classification and classic FC, which he dubbed 'freely faceted classification'. In FFC, facets from different domains of phenomena could be freely combined and the adoption of rules governing citation order could help to make meanings clearer. He proposed that Integrative Level Classification (ILC), whereby phenomena are organized in order of increasing complexity (e.g. atoms > molecules > cells > organisms) could serve as a useful framework.

The increasing interdisciplinarity of IR needs was illustrated by a screen shot of the BioAcoustic Reference Database (BARD). While the notation might be expressive, it was not intuitive. The adoption of mark-up conventions for notation offers a possible solution to the problem of expressing semantic relationships, but the resulting character strings are not intuitively meaningful. In BARD, the obscurity of the notation is ameliorated through the use of PHP scripts to synthesize verbal descriptions which are displayed alongside the classmarks of items in search result sets.

Claudio concluded that FFC has great promise, but a number of problems remain to be resolved in 'taming complexity'.

### **Faceted Categorisation for the Corporate Desktop: Visualisation and Interaction using Metadata to Enhance User Experience.** **Mark Stapleton and Matt Adamson**

Mark Stapleton and Matt Adamson began their presentation by describing how Dow Jones' Factiva range of information services processed an average of 170,000 documents every day, drawn from over 10,000 sources in 22 languages. These documents are categorized within five facets: Company, Subject, Industry, Region and Language. The digital feeds received from information providers undergo a series of processing stages, initially to prepare them for automatic categorization and then to format them ready for distribution. The categorization stage is able to handle 98% of documents automatically, the remaining 2% requiring some form of human intervention. Depending on the source, categorization can involve any combination of 'Autocoding', 'Dictionary-based Categorizing', 'Rules-based Coding' or 'Manual Coding'.

Matt described how the output from the source processing is presented differently in three different products for different user groups. One group comprises information intermediaries such as librarians, who conduct research using a variety of business information products as their primary function, and who support a number of departments. A second group consists of individuals whose primary function is not research, but who nevertheless need ready access to quality, up-to-date information in order to make decisions. A third group comprises those who need to know what is happening, although not necessarily in great detail, and who expect information to be available when they need it.

The Factiva delivery platform offers a number of innovative capabilities, including alert, search,



monitor, cluster, discovery and analytics. A search can start with any of the five facets offered by the interface. Matt provided a simulation of a search starting with his own company, Dow Jones, illustrating how further basic information on each item in the result set was available via a pop-up window. For information intermediaries, Factiva's interface focuses on supporting the construction of complex queries, but for the second group of users, the emphasis is on the presentation of results in order to provide additional context and aid exploration and discovery. The conventional listing is enhanced by brief abstracts and enriched through the display of statistical summaries in chart form, together with graphical links leading to related information items tagged under other facets. For the third user group, information is provided to individuals by various 'push' technologies according to their interest profile.

Mark concluded the presentation by describing how Factiva need to ensure their products continually evolve in order to meet changing user expectations. He quoted from a Gartner report which highlighted how the knowledge worker of the near future will require to build his/her own custom 'portal' aggregating and personalizing a variety of sources and services. The role of the centralized information source would be eroded by increasing reliance on user communities, while a move to user-centric rather than technology-centric workflow and increasing use of consumer technologies such as social computing in the workplace would create a radically different working environment.

## Faceted Classification as an Intelligence Analysis Tool. *Jan Wyllie and Simon Eaton*

Jan and Simon are collaborating in the development of a collaborative web-based resource to be called The Energy Centre (TEC). TEC will allow the collaborative collection of clips relating to all aspects of the energy sector. The clips will be stored and organized in such a way that they are not only easily searchable, but can serve as the basis for content analysis - defined as 'a technique for systematic inference from communications'.

Jan began by explaining that it was while working as an intelligence analyst at the Canadian Trend Report in Montreal, that he learned about content analysis, a classic taxonomy-based intelligence research methodology. The insights into emergent trends which content analysis provides had proved of such value to Canadian government organizations that when he returned to England, Jan established Trend Monitor in 1984. Trend Monitor pioneered the publishing of intelligence reports on computing, communications and media, initially in association with Aslib.

Around the same time, Jan met Dr Tony Kent, one of the pioneers of free text software development, and became a director of his company Microbel where they produced and sold **Strix**, the world's first full function free text database designed to work on PCs. It was his work with Dr. Kent which gave Jan his insight into the ways in which such databases, 'multivariate analysis' and taxonomies could be combined to produce an intelligence analysis tool.

The collection of 'clips' from published media used to be highly effort-intensive, but with modern web-based tools such as Clipmarks, content items can now be collected easily from the Web. However, Clipmarks does not provide anything more than a very basic tagging system for categorizing clips, and seems uninterested in developing more comprehensive categorization facilities based upon taxonomies. Jan and Simon are therefore building TEC to provide intelligence in the energy sector based upon the comprehensive categorization, search and navigation facilities provided by Strix. Simon then presented a simulation of the interface being developed for TEC, which is based firmly upon a faceted approach where each facet is supported



by its own taxonomy.

## **AutoFocus: An Open-source Facet-driven Enterprise Search Solution. *Jeroen Wester***

In the final presentation of the afternoon, Jeroen Wester of Aduna described the main features of their open-source, facet-driven enterprise search solution, AutoFocus. AutoFocus is based upon and exploits the advantages of Semantic Web technologies, in particular RDF (Resource Description Framework), although a bewildering variety of related technologies - XML, SOAP, SKOS, OWL - are also employed. In addition to providing components for metadata-based data integration and cross-silo search and navigation in a single enterprise search solution, AutoFocus offers the advantage of being open-source, meaning that its source code is freely available for customization.

AutoFocus consists of two components. AutoFocus itself is a compiled desktop application which will scan and index documents on specified 'sources' and then provide facilities for searching them. Sources can include local file system folders, websites, IMAP email folders and remote Aduna AutoFocus Servers. The second component, AutoFocus Metadata Server, provides a means for an AutoFocus client to gain access to network resources. Significantly, AutoFocus Metadata Server not only offers standard content searching, but it will also index and make searchable key metadata fields embedded in source documents, such as those in the MS Office document Properties Sheet, and Acrobat PDF file XMP metadata.

After outlining the common problems associated with full-text search, Jeroen went on to characterize AutoFocus' approach as 'metadata-centric for metadata exploration & query formulation' and as providing 'information visualisation for search result exploration'. He continued with a simulation designed to demonstrate these distinctive features of the product, concentrating on AutoFocus' 'Cluster Map' visualization of search results.

Cluster Maps display search results as a sort of Venn diagram, where documents containing search terms X or Y are represented by two discrete clusters, while documents including both X AND Y are represented by a third cluster linked to the other two. The AutoFocus interface is user-customizable so that the clusters can display simply as a sphere showing the number of documents in each cluster, or can display each document as a sphere. Hovering the mouse over a document sphere pops-up a description of its location and title. Clicking a cluster displays a result list in the lower half of the window where any item can be clicked to open it in its native application. In search mode, AutoFocus can also display a list of suggested refinement terms derived from the other keywords indexed for that source.

Bob Bater, 2007-11-12

<<<<< >>>>>