An Overview of UKOLN Work Related to Subject-based Knowledge Organisation

Document details

Author	Koraljka Golub
Date:	16 July 2013
Version:	1.0
File Name:	UKOLN-report-semantics.doc
Abstract:	This report provides an overview of UKOLN work related to subject-based knowledge organisation.

Acknowledgements

Contents Page

An Overview of UKOLN Work Related to Subject-based Knowledge Organisation

1   Subject-based Knowledge Organisation

Knowledge organisation as a term is used by various communities to include anything from organizing the World Wide Web, including technologies of the Semantic Web, to creating bibliographies, catalogue records, archival searching aids and related. Subject-based knowledge organisation focuses on organising information and knowledge based on its topicality or aboutness. In indexing and abstracting databases as well as in library catalogues tools known as knowledge organisation systems are commonly applied, such as thesauri, subject headings and classification schemes. In Web 2.0 there are folksonomies whereby end users assign tags to documents they choose. In the Semantic Web, ontologies are used as a type of highly structured and detailed knowledge organisation system which is used to allow for logical inferencing.

The most recent UKOLN work related to subject-based knowledge organisation is an overview of knowledge organisation systems (KOS), its current usage and current issues. It forms a part of the Technical Foundations (http://technicalfoundations.ukoln.ac.uk/). The whole document is available at http://technicalfoundations.ukoln.ac.uk/subject/knowledge-organisation-systems.

The general purpose of KOS is to provide a means for organising information (ANSI/NISO Z39.19), through:

- translation of the natural language of authors, indexers, and users into a vocabulary that can be used for indexing and retrieval

- ensuring consistency through uniformity in term format and in the assignment of terms

- indicating semantic relationships among terms

- supporting browsing by providing consistent and clear hierarchies in a navigation system

- supporting retrieval

KOS play a crucial role in resource retrieval and discovery. They improve the effectiveness of retrieval by helping to handle the sheer mass of information and they provide knowledge-based support for end users who access information without the help of an intermediary. In comparison to free-text searching, there are many advantages to searching by KOS terms, such as the following:

- the most relevant search terms are selected, and relevant search terms which are not explicitly mentioned in a document may be added

- search terms are controlled, i.e. disambiguated, so that there is no confusion between terms that look the same but have different meanings

- search terms can come from semantically structured vocabularies – hence documents can be found by searching for synonyms, narrower, broader, and even related terms that may not be present in the document itself (semantic query expansion)

A well-structured KOS can be used as the knowledge base for an interface that can assist users with search topic clarification (e.g. through browsing well-structured hierarchies and guided facet analysis) and with finding good search terms (through query term mapping and query term expansion: synonyms and hierarchical inclusion).

Additional functions of KOS are to (Soergel 2003):

- help improve communication, support learning and assimilating information (e.g. through providing conceptual frameworks to help the learner ask the right questions, assist readers in understanding text by giving the meaning of terms, assist writers in producing understandable text by suggesting good terms, and support foreign language learning)

- provide the conceptual basis for the design of good research and implementation (e.g. assist researchers and practitioners with problem clarification)

- provide classification for action, classification for social and political purposes (e.g. classification of diseases for diagnosis)

- facilitate unified access to multiple databases
- serve as a source for data element definition and provide a conceptual basis for knowledge-based systems
- do all this across multiple languages

KOS may be used in a variety of applications. Their most prominent use is for improved information retrieval through searching, disambiguation, query expansion and reformulation, or browsing. Different KOS serve different functions, which is why more than one KOS should ideally be used in information retrieval applications. For example, classification schemes generally serve to group together topically related documents into classes and are thus better suited to subject browsing than other KOS; thesauri are used to denote a number of detailed topics and are thus better suited for searching (although examples of KOS which aim to integrate both functions exist). When considering adopting a particular KOS from a type of KOS, the subject indexing policy for the collection at hand needs to be considered: for example, the bigger the collection, the more depth the classification hierarchy should contain, and more detailed topics should be listed in a thesaurus; quality and maintenance (e.g. home-grown KOS on the Web often lack principles from international standards on design and development of KOS), etc.

Other uses include aiding in the general understanding of a subject area, providing "semantic maps" by showing inter-relationships between concepts, and helping to provide definitions of terms. KOS can help improve automated classification and indexing, semantic reasoning, text mining, and information extraction. Topical crawlers or harvesters can utilize KOS to define topics using the high-quality terms for those topics. KOS can also provide support for social tagging, and consequently improve information retrieval and knowledge organisation in Social Web applications.

Today KOS are used in a variety of contexts:

- in libraries: for shelf arrangement, information retrieval (both searching and browsing), and collection management (acquisition, circulation statistics, weeding)
- in museums and archives: for collection display, objects indexing and retrieval, and collection management
- in bibliographies, for subject information navigation
- in bibliographic databases (including repositories and subject gateways), for information retrieval
- in information services, for selective dissemination of information
- in journal articles (e.g. "keywords" or "index terms" in the abstract)
- in metadata (e.g. recommended as part of the Dublin Core element "subject")
- as a source for building various knowledge domain maps (ontologies) and other KOS
- in data mining
- in knowledge management

Examples of using KOS for improving the performance of automated subject indexing and classification also exist, and so do KOS as a feed for topical crawlers, as well as KOS as a source for social tagging (currently these are largely experimental but show considerable potential).

Major current research issues with KOS cover interoperability of KOS across various applications, exploring potential alternatives to manual subject indexing and classification, and improvements need for KOS in the digital, networked environment.

The fact that classification schemes use a system of notation to represent the hierarchical structure of concepts, where each concept is represented by a notation rather than a natural language term, provides the potential for interoperable search and browsing access to multilingual databases when the databases use the same classification schemes. However, if the KOS used in the databases differ in structure, domain, language, or granularity, the KOS will need to be transformed, mapped, or merged. Moreover, multilingual KOS mapping is complex because it involves translation of concepts, not terms, and there is often significant variation between languages. Different cultural perspectives also need to be integrated (e.g. the concept space of education in one country can be rather different

to that in a neighbouring country). On the one hand, communities develop KOS specific to their concepts, terminology, and needs; on the other hand searchers want to use a single search to find resources in databases serving different domains and accessed by different KOS, across which there may be no consensus regarding concepts, terminology, and knowledge organisation.

Apart from semantic interoperability, there also needs to be interoperability with applications: KOS should work with search engines, Content Management Systems, Web publishing software, etc. In order to do this they need to be made available in existing formats and protocols for data exchange, such as SKOS for representation of KOS in RDF in a simple way, and URIs for unique identification of the KOS, its concepts and terms. SKOS and URIs will allow KOS become Linked Data. While early adopters exist, there is a long way to go before the potential of these approaches is fully explored and implemented in practice.

Although it is very unlikely that there will be approaches that would entirely replace creating quality subject metadata by humans, there are two major attempts in current research and practice aimed at adding to subject metadata created by trained subject metadata specialists: social tagging using KOS as a basis, and automated or semi-automated means. Both approaches warrant further research:

1. Social tagging involves adapting KOS for end user tagging: it needs to be determined which modifications are most likely to make KOS more useful in this context. The changes may include more definitions, better displays and algorithms providing good automated suggestions. Motivation of end users for tagging also needs to be explored further, etc.

2. Although the vendors of today's research and commercial software sellers emphasise the high potential of automated tools for subject metadata generation, real evidence of their success is so far lacking. Software tools may be useful but only in very constrained subject domains; they are unlikely to improve with research because it is essentially "hard" artificial intelligence. The difference between reported high performance results and the reality is in part due to restricting the evaluation of these tools to comparison against existing or ad hoc metadata that serves as the gold standard in laboratory-like conditions which has inherent subjectivity problems in two areas: the correct interpretation of a document's subject matter; any evaluation of the tools is carried out in the context of a laboratory-like environment rather than a real operational system where the most commonly used measures are precision and recall. Although this issue has been discussed widely in the literature, mainstream research has not paid much attention, and published results are widely acknowledged nonetheless. However, existing human-assigned metadata cannot be used as a gold standard. For example, the classes assigned by algorithms, rather than by humans, might be wrong; alternatively, they might be right but mistakenly omitted during human indexing. Subject metadata creation involves determining subject terms or classes under which a document should be found: this goes beyond simply capturing what the document is about to what the document could be used for; algorithms might find such terms, given a good training set, but human indexers who are not well trained might miss them.

There are a number of areas in which existing KOS could be improved. One approach is to simplify complex KOS that are intended for use in the first instance by librarians and trained end users in a paper environment, for the benefit of non-specialists and for use on the Web. This should also include hierarchy browsing at different levels, hyperlinks for relationships, searching for compounds containing any combination of elemental concepts, adjustments for social tagging applications, etc. Replacing complex built-in concepts, which are present in some KOS, with a structure based on facets, would allow greater flexibility in building new specific concepts at the time of searching as required by the end-user and at the same time reduce the size of the KOS.

Another approach is to enrich one KOS with the benefits of other types of KOS. For example, enriching typical thesauri with hierarchical structure would enable their use both for searching and for browsing. Moreover, empowering end users in searching collections of ever increasing magnitudes, with performance far exceeding plain free-text searching, and developing systems that not only find but also process information, requires far more powerful and complex KOS: thus enriching thesauri with the characteristics of ontologies would be highly beneficial in such applications.

The slow maintenance and updating of some KOS is an issue for end-users who cannot find new concepts and terms or who cannot find out how to use them because of outdated structures, hierarchies and similar. A major reason why updating has been slow is that it would require re-indexing and re-classification of existing collections, which implies expensive re-shelving in libraries;

changing the structure would also cause problems for end-users as they would have to learn the new structures when browsing either online or in a physical collection.

KOS do not simply represent the information, but also construct that information. For example, while existing classification schemes are intended to be universal, they are actually culturally specific (e.g. the Chinese Library Classification, BBK in the former Soviet Union). In the Dewey Decimal Classification, the most widespread classification system in the world, regional variants had to be introduced as a compromise. In KOS there persists a historical bias on the basis of gender, sexuality, race, age, ability, ethnicity, language and religion, which limits the representation of diversity and effective library service for diverse populations. Now used globally and in interoperable systems, the KOS should be restructured in order to address these issues in a modern context: this once again implies re-classification and re-indexing efforts which are expensive in themselves, and getting the end users to re-learn the KOS they have been used to.

UKOLN has touched on most of the types of KOS and dealt with various aspects of the major issues on the world's research agenda described above.

2 UKOLN Projects and Activities Related to Subject-based Knowledge Organisation

The projects are described based on the following elements:

> URL:
> Period:
> Funder:
> Who:
> Context:
> Key outputs:

Most projects at UKOLN are related to general metadata and general information and knowledge organisation. These include projects that deal with cross-searching of bibliographic databases, institutional repositories and only touch on subject access points, for example:

1. **LOCAH: Linked Open Copac Archives Hub**

   URL: http://archiveshub.ac.uk/locah/about/
   Period: 2010-2011
   Funder: JISC
   Who: Mimas and UKOLN (Julian Cheal, Adrian Stevenson), in partnership with Eduserv, Talis and OCLC
   Context: Making UK Archives Hub and Copac data available as Linked Data, for the benefit of education and research, enabling new links to be made between diverse content sources and enabling the free and flexible exploration of data so that researchers can make new connections between subjects, people, organisations and places to reveal more about our history and society.
   Key outputs: Archives Hub Linked Data made available at http://data.archiveshub.ac.uk/. Continued in the Linking Lives project (http://archiveshub.ac.uk/linkinglives/). Related publications are available at http://archiveshub.ac.uk/locah/talks/.

2. **RepUK: an aggregation of UK Open Access Institutional Repository metadata**

   URL: http://www.ukoln.ac.uk/projects/repuk/
   Period:
   Funder: JISC
   Who: UKOLN (Mark Dewey, Paul Walk, Monica Duke, Julian Cheal)
   Context: To create an aggregation of metadata from repositories which would serve to support research and development generally in the fields of metadata and/or repositories, to support ongoing services and others. Creating a simple RESTful API to serve third-party retrieval of harvested records.
   Key outputs: Demonstrator available at http://repuk.ukoln.ac.uk/

3. **Tap into Bath**

URL: http://www.ukoln.ac.uk/tapintobath
Period: 2003-2010
Funder: MLA, JISC, SWMLAC
Who: UKOLN (Ann Chapman), the University of Bath Library, in cooperation with libraries, archives, and museums in Bath
Context: Tap into Bath is an example of how to enable users to find out about archives, libraries and museums; in this case all the collections are located in the city of Bath.
Key outputs: The software (a MySQL database and Web app), which was developed as an open source application for free re-use with accreditation (http://www.ukoln.ac.uk/tapintobath/software-and-documentation/).

4. **ROADS (Resource Organisation And Discovery in Subject-based services)**

URL: http://roads.opensource.ac.uk/; http://www.ukoln.ac.uk/metadata/roads/; more information at web.archive.org/web/19970128035811/http://www.ukoln.ac.uk/roads/
Period: 1995-1997
Funder: JISC
Who: UKOLN (Lorcan Dempsey, Rachel Heery); The Institute for Learning and Research Technology, University of Bristol; The Department of Computer Studies at Loughborough University
Context: Aims are to make a significant contribution to the development of a sharable, distributed systems platform for resource discovery services such as subject gateways and to work with subject based services to involve information providers in the description of their own resources, in order to make them as useful and accessible as possible.
Key outputs: ROADS tools, a set of guidelines on metadata, interoperability, cataloguing etc.

5. **RDN: Subject Portals Project**

URL: http://www.portal.ac.uk/spp/
Period: 1999-2004
Funder: JISC
Who: UKOLN (Rosemary Russell); ILRT (Institute for Learning and Research Technology), University of Bristol; contributing data hubs.
Context: Creating an aggregated cross search tool searching across both JISC supported and non-JISC information resources specially selected by the hubs themselves.
Key outputs: Demonstrator (http://www.portal.ac.uk/spp/demo/), project deliverables

There have also been projects dealing with automated extraction of non-subject information and metadata, such as FixREP (http://www.ukoln.ac.uk/projects/fixrep/) and Writeslike.us.

6. **Writeslike.us**

URL: http://www.ukoln.ac.uk/projects/writeslike.us/
Period: 2009
Funder: JISC
Who: UKOLN (Emma Tonkin, Andrew Hewson, Alexey Strelnikov); University of Minho
Context: This project links formal and informal metadata for the purpose of helping individuals discover others with shared topics and interests, which has been described as an important step in linking individuals together without formal social networking information. A prototype repository enhancement service which automatically extracts candidate community participation information from analysis of existing documents and metadata is developed.
Key outputs: A prototype service available at http://writeslike.us/

A complete list of projects, activities and publications on metadata in general is available at http://www.ukoln.ac.uk/metadata/. These are outside the scope of this report.

Subject-based knowledge organisation projects are listed below. Some are also included which had dealt with subject access in more detail as part of establishing larger information retrieval systems, such as Renardus. They are listed from latest to earliest.

1. **EASTER (Evaluating Automated Subject Tools for Enhancing Retrieval): testing and evaluating existing tools for automated subject metadata generation**

   URL: http://www.ukoln.ac.uk/projects/easter/
   Period: 2009-2011.
   Funder: JISC
   Who: UKOLN (lead) (Koraljka Golub, Michael Day); Hypermedia Research Unit, University of Glamorgan; Intute, MIMAS – University of Manchester; Centre for HCI Design, City University London; Dagobert Soergel (consulting expert)
   Context: Subject metadata are most important in resource discovery, yet most expensive to produce manually. Subject metadata are much more difficult to generate automatically especially in comparison to formal metadata such as file type, title, etc. Due to the high cost of evaluation, automated subject metadata tools are rarely tested in live environments of use. Enhancing Retrieval). The purpose of the project was to test and evaluate existing tools for automated subject metadata generation by developing a detailed, comprehensive methodology.
   Key outputs: Two papers submitted to JASIST: D. Soergel & K. Golub: Consistency and performance measures for indexing and for retrieval; K. Golub et al.: A framework for evaluation of tools for automated subject assignment in the context of retrieval.

2. **Enhanced Tagging for Discovery (EnTag): investigating the combination and comparison of controlled and folksonomy approaches to support resource discovery in repositories and digital collections.**

   URL: http://www.ukoln.ac.uk/projects/enhanced-tagging/
   Period: 2007-2009
   Funder: JISC
   Who: UKOLN (lead) (Koraljka Golub, Michael Day); University of Glamorgan, Hypermedia Research Unit; Intute (MIMAS at The University of Manchester); Science and Technology Facilities Council, The Rutherford Appleton Laboratory
   Context: The Enhanced Tagging for Discovery project investigated the combination and comparison of controlled and folksonomy approaches to support resource discovery in repositories and digital collections. The project evaluated the combination of both approaches in the context of repositories and digital collections, attempting to get the best of both worlds. The specific aim was to investigate whether vocabulary control and the use of an established KOS can assist in moving free social tagging beyond personal bookmarking to aid resource discovery in context of the JISC Information Environment and eFramework. The project demonstrated use of tagging in different environments and provided an interface that enables use of a traditional classification scheme to enhance free form tags. The project considered issues such as whether prompting with controlled terminology is beneficial.
   Key outputs: Reports and publications available at project web site. A major paper so far is (JCDL 2009 best paper award nomination): Golub, K. et al. (2009) EnTag: Enhancing Social Tagging for Discovery. Joint Conference on Digital Libraries (JCDL), Austin, TX, June 15-19. Demonstrators are available at http://www.ukoln.ac.uk/projects/enhanced-tagging/demonstrators/.

3. **Terminology Registry Scoping Study (TRSS): analysing issues related to the potential delivery of a Terminology Registry as a shared infrastructure service within the JISC Information Environment.**

   URL: http://www.ukoln.ac.uk/projects/trss/
   Period: 2007-2008
   Funder: JISC
   Who: UKOLN (Koraljka Golub); University of Glamorgan, Hypermedia Research Unit
   Context: A terminology registry lists, describes, identifies and points to sets of vocabularies available for use in information systems and services. It can cover free and publicly available, fee-based and restricted, or organisation-internal vocabularies. The study will analyse issues related to the potential delivery of a Terminology Registry as a shared infrastructure service within the JISC Information Environment. The study considered how a Registry might support development of terminology and other services within the context of a services oriented environment. It described usage scenarios and use cases, investigated requirements and sustainability, study costs and benefits. It provides recommendations to JISC, including those on metadata.

Key outputs: Golub, K; Tudhope, D (2009) Terminology Registry Scoping Study (TRSS): Final report. Available at: http://www.ukoln.ac.uk/projects/trss/dissemination/trss-report-final.pdf

4. **HILT: High-Level Thesaurus project aimed to research and develop solutions for problems pertaining to cross-searching multi-subject scheme information environments.**

URL: Archived at http://web.archive.org/web/20120120072647/http://hilt.cdlr.strath.ac.uk/
Period: 2001-2009.
Funder: JISC
Who: Centre for Digital Library Research (CDLR), University of Strathclyde; EDINA at the University of Edinburgh; UKOLN (Rachael Heery, Rosemary Russel, Michael Day)
Context: Problems relating to disparate terminology use have been an impediment to information retrieval for many years, but the growth of Web, associated heterogeneous digital repositories, and the need for distributed cross-searching within multi-scheme information environments has recently drawn the issue into sharp focus. The HILT project aims to research, investigate and develop pilot solutions for problems pertaining to cross-searching multi-subject scheme information environments, as well as providing a variety of other terminological searching aids.
Key outputs: Demonstrators, also archived, but available at http://web.archive.org/web/20120130074635/hilt.cdlr.strath.ac.uk/hilt4/demonstrators.html

5. **Renardus: Academic Subject Gateway Service Europe.**

URL: archived at http://web.archive.org/web/20070709062035/http://www.renardus.org/
Period: 2000-2002
Funder: EU's Information Society Technologies 5th framework programme
Who: Koninklijke Bibliotheek (National Library of the Netherlands) - (KB); Bibliothèque Nationale de France (National Library of France) - (BNF); Center for Scientific Computing, Finland - (CSC); Die Deutsche Bibliothek (National Library of Germany) - (DDB); Finnish Virtual Library Project, Jyväskylä University Library, Finland - (JyU); Institute for Learning and Research Technology, University of Bristol, UK - (ILRT); NetLab, Lund University, Sweden - (NetLab); Niedersächsische Staats- und Universitätsbibliothek, Göttingen, Germany - (SUB); Technical Knowledge Centre and Library of Denmark - (DTV); Viikki Science Library, University of Helsinki, Finland - (ALUH); Zentralstelle für Agrardokumentation Und - information, Germany - (ZADI); UKOLN (Michael Day).
Context: The service aims to provide a trusted source of selected, high quality Internet resources for those teaching, learning and researching in higher education in Europe. Renardus provides integrated search and browse access to records from individual participating subject gateway services across Europe.
Key outputs: Actual implementation of the service; numerous papers and reports available at the archived site (About Us, Project Archive)

6. **DELOS Network of Excellence on Digital Libraries. Workpackage 5: Knowledge Extraction & Semantic Interoperability**

URL: http://delos-wp5.ukoln.ac.uk/
Period: 2004-2007
Funder: EU 6th Framework Programme, Information Society Technologies Programme
Who: Centre for the History of Science and Department of Computing, Imperial College, UK; Database Research Group, ETH Zurich (Swiss Federal Institute of Technology Zurich), Switzerland; Department for Information and Software Engineering, UMIT University for Health Sciences, Medical Informatics and Technology, Austria; Department of Information Science, University of Milan, Italy; Hypermedia Research Unit, University of Glamorgan, UK; Institute of Computer Science, Foundation for Research and Technology - Hellas (ICS-FORTH), Greece; Netlab Knowledge Technologies Group, Lund University, Sweden; School of Electronics and Computer Science, University of Southampton, UK; School of Informatics, University of Edinburgh, UK; Technical University of Crete, Greece; UKOLN, University of Bath, UK (Elizabeth Lyon, Manjula Patel)
Context: The thematic area of semantic interoperability refers to the application of different vocabularies and terminology used in descriptions of digital objects for both learning and research, collections of those objects, collections of datasets and resources used in the wider cultural heritage sector and in e-research. Indeed, cross-sectoral and cross-domain shared understanding of semantic descriptions is one of the goals of the Semantic Web as envisaged

by Tim Berners-Lee in his "roadmap" published in 1998. The aim of the cluster was to address some of the issues and challenges in this complex area.

Key outputs: Patel, M.; Koch, T.; Doerr, M.; Tsinaraki, C. 2005. Semantic interoperability in digital library systems. In DELOS Network of Excellence on Digital Libraries, European Union, Sixth Framework Programme. Deliverable D5.3.1., available at http://delos-wp5.ukoln.ac.uk/project-outcomes/SI-in-DLs/

7. **DESIRE: Development of a European Service for Information on Research and Education. Phase 2. Work Package 3 (WP3): Indexing and Cataloguing.**

URL: http://www.ukoln.ac.uk/metadata/desire/phase-1/
Period: 1998-2000
Funder: The European Commission, Telematics for Research Sector of the Fourth Framework Programme
Who: UKOLN (Rachel Heery, Tracy Gardner , Michael Day, Manjula Patel, Andy Powell); Institute for Learning and Research Technology (ILRT), University of Bristol;   Department of Library Research, Koninklijke Bibliotheek (Netherlands); Delivery of Advanced Network Technology to Europe Ltd. (DANTE); SURFnet bv [Netherlands]; NetLab, Lund University Library Development Department, University of Lund [Sweden]; Trans-European Research and Education Networking Association (TERENA) [Netherlands]; Computing Centre, Brunel University [UK].
Context: As continuation of Phase 1 (see the following project description, or http://www.ukoln.ac.uk/metadata/desire/phase-1/), Phase 2 had three main areas of activity: caching, resource discovery and directory services.
Key outputs: DESIRE Information Gateways Handbook available at http://web.archive.org/web/20110725232202/http://www.desire.org/handbook/.   Reports on: quality ratings and RDF, metadata registry framework, available at the project URL.

8. **DESIRE: Development of a European Service for Information on Research and Education. Phase 1. Work Package 3 (WP3): Resource discovery and indexing.**

URL: http://www.ukoln.ac.uk/metadata/desire/phase-1/
Period: 1996-1998
Funder: The European Commission, Telematics for Research Sector of the Fourth Framework Programme
Who: UKOLN (Lorcan Dempsey, Rachel Heery, Andy Powell, Michael Day); Institute of Learning and Research Technology (ILRT), University of Bristol; The Department of Computer Studies, Loughborough University; NetLab: Lund University Library Development Department, University of Lund; Koninklijke Bibliotheek: the National Library of the Netherlands.
Context: DESIRE I investigated a series of issues relating to Web technologies and the implementation of pilot information services on behalf of European researchers. Areas of research included cataloguing and indexing, caching, security issues and training. UKOLN's involvement in DESIRE I was focussed on Work Package 3 (WP3) which was concerned with Resource discovery and indexing.
Key outputs: Specification for resource description methods comprising a review of metadata, selection criteria for subject gateways, and the role of classification schemes on the Internet, available at the project's URL.

9. **Cooperative Hierarchical Indexing Coordination**

URL: http://www.ukoln.ac.uk/metadata/tf-chic/
Period: 1997-1999
Funder: TERENA
Who: TERENA Task Force, including UKOLN (Andy Powell)
Context: The task force was concerned with the coordination of harvesting and indexing network resources. Work in the task force built upon existing standards and technologies, such as those employed in Harvest and the DESIRE and ROADS projects.
Key outputs: The task force spawned the pilot project that set up a pilot distributed indexing service based on Whois++, Harvest, ROADS and Z39.50 technology. The demo search interface was developed but is no longer available.

3   UKOLN Publications Related to Subject-based Knowledge Organisation

Taken from http://www.ukoln.ac.uk/metadata/kos/publications.html

Golub, K., Muller, H., Tonkin, E. 2013. Technologies for metadata extraction. In Miguel-Angel Sicilia (Ed.), Handbook of Metadata, Semantics and Ontologies. World Scientific, forthcoming.

Golub, K. 2011. Automated subject classification of textual documents in the context of Web-based hierarchical browsing. Knowledge Organization, 3(38), p. 230-244.

Matthews, B; Jones, C.; Puzoń, B; Moon, J; Tudhope, D.; Golub, K.; Lykke Nielsen, M. 2010. An evaluation of enhancing social tagging with a knowledge organization system. Aslib Proceedings: New Information Perspectives, 4/5(62), p. 447-465. DOI: 10.1108/00012531011074690.

Tonkin, E., Pfeiffer, H. and Hewson, A. 2010. An evidence-based approach to collaborative ontology development. In Workshop on Matching and Meaning 2010, 31 March - 1 April 2010, Leicester, UK.

Tonkin, E. and Pfeiffer, H. D., 2010. Data-driven or background knowledge ontology development. In: International Conference on Knowledge Management (ICKM), 22-23 October 2010, Pittsburgh, USA. Paper No. 134.

Golub, K., Lykke Nielsen, M. 2009. Automated classification of Web pages in hierarchical browsing. Journal of Documentation, 65(6). P. 901–925. DOI: 10.1108/00220410910998915.

Golub, K., Jones, C., Lykke Nielsen, M., Matthews, B., Moon, J., Puzon, B., Tudhope, D. 2009. EnTag: Enhancing social tagging for discovery. Joint Conference on Digital Libraries (JCDL), Austin, TX, June 15-19. P. 163-172. DOI: 10.1145/1555400.1555427

Golub, K., Tudhope, D. 2009. Terminology Registry Scoping Study (TRSS): final report.

Golub, K., Jones, C.; Lykke Nielsen, M.; Matthews, B.; Moon, J.; Tudhope, D. 2009. Enhanced Tagging for Discovery (EnTag): final report.

Pfeiffer, H. and Tonkin, E., 2009. Tagging in Context: Information Management across Community Networks. In: Bouras, C., Poulopoulos, V. and Tsogkas, V., eds. Handbook of Research on Social Interaction Technologies and Collaboration Software: Concepts and Trends. IGI Global.

Pfeiffer, H. D., Tonkin, E., Lindner, M. R., Kipp, M. and Millen, D. R., 2008. Tagging As A Communication Device: The Impact of Communities on Transforming Tag Information. In: ASIS&T'08, People Transforming Information - Information Transforming People, October 2008, Columbus, OH.

Golub, K.; Jones, C; Lykke Nielsen, M; Matthews, B; Moon, J; Tudhope, D. 2008. Enhancing social tagging with a knowledge organization system. ALISS, Vol 3, No 4, July 2008, pp. 13-16.

Tonkin, E., Tourte, G. and Zollers, A., 2008. Performance Tags- Who's running the show? In: Proceedings 19th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research, Columbus, Ohio.

Tonkin, E. and Pfeiffer, H. D., 2008. Case study of software-assisted collaborative ontology building. In: ICKM 2008, October 2008.

Tonkin, E. and Muller, H., 2008. Keyword and metadata extraction from pre-prints. In: ElPub, June 2008.

Tonkin, E., Corrado, E. M., Moulaison, H. L., Kipp, M. E., Resmini, A., Pfeiffer, H. and Zhang, Q., 2008. Collaborative and Social Tagging Networks. Ariadne, 54.

Tonkin, E., 2008. Orthography, Structure and lexical choice as identity markers in social tagging environments. In: IADIS International Conference on Web Based Communities.

Tonkin, E., 2006. Folksonomies: The Fall and Rise of Plain-text Tagging. Ariadne, 1 (47). [html]

Tonkin, E., 2006. Searching the long tail: hidden structure in social tagging. In: Proceedings of the 17th ASIS SIG/CR Classification Research Workshop.

Tudhope, D.; Koch, T.; Heery, R. 2006. Terminology Services and Technology. JISC state of the art review. [html]

Guy, M. and Tonkin, E., 2006. Folksonomies: Tidying up Tags? D-Lib Magazine, 12 (1).

Day, M., Koch, T. and Neuroth, H. 2005. Searching and browsing multiple subject gateways in the Renardus service. In: van Dijkum, C., Blasius, J. and Durand, C., eds. Recent developments and applications in social research methodology: Proceedings of the RC33 Sixth International Conference on Social Science Methodology, Amsterdam 2004. Opladen: Budrich Verlag.

Patel, M.; Koch, T.; Doerr, M.; Tsinaraki, C. 2005. Semantic interoperability in digital library systems. In DELOS Network of Excellence on Digital Libraries, European Union, Sixth Framework Programme. Deliverable D5.3.1. [

Heery, R. 2004. Metadata futures: steps towards semantic interoperability. In: Diane Hillmann and Elaine Westbrooks, (eds.), Metadata in Practice: a work in progress, American Library Association, Chicago 2004, pp. 257-271.

Heery, R. 2003. Delivering HILT as a JISC IE shared service: HILT Project deliverable.

Koch, T., Neuroth, H. and Day, M. 2001. Renardus: Cross-browsing European subject gateways via a common classification system (DDC). In: "Subject Retrieval in a Networked Environment". Proceedings of the IFLA Satellite Meeting sponsored by the IFLA Section on Classification and Indexing and the IFLA Section on Information Technology, 14-16 August 2001, Dublin, OH, USA. UBCIM Publications - New Series Vol. 25, Muenchen 2003. pp25-33.

Koch, T., Neuroth, H. and Day, M. 2001. DDC Mapping Report. Renardus D7.4, Internal report.

Koch, T., Neuroth, H. and Day, M. 2001. DDC Mapping Guidelines. Renardus D7.4 internal deliverable.

Russell, R. and Day, M. 2001. Automated and manual approaches to the provision of thesauri and subject vocabularies. HILT project report.

UKOLN contributor to: Martin Belcher, Virginia Knight and Emma Place, (eds.). 1999. DESIRE Information Gateways Handbook: 2.5 Subject classification, browsing and searching. http://www.desire.org/handbook/

Koch, T. and Day, M. 1997. The role of classification schemes in Internet resource description and discovery. (EU Project DESIRE, Deliverable D3.2.3).

4 UKOLN Presentations Related to Subject-based Knowledge Organisation

Taken from http://www.ukoln.ac.uk/metadata/kos/presentations.html

Golub, K.; Lykke Nielsen, M; Moon, J; Tudhope, D. 2009. Enhancing social tagging with a knowledge organization system. IFLA 2009 Satellite Meeting "Emerging trends in technology: libraries between Web 2.0, semantic web and search technology", Florence, 19-20 August 2009.

Matthews, B; Jones, C.; Puzon, B; Moon, J; Tudhope, D.; Golub, K.; Lykke Nielsen, M. 2009. An evaluation of enhancing social tagging with a knowledge organization system. ISKO UK Conference, London, 22-23 June. [

Golub, K., Tudhope, D. 2008. TRSS: Terminology registry scoping study. ECDL NKOS workshop, 19 September 2008.

Golub, K., Tudhope, D., Lykke Nielsen, M., Moon, J. 2008. EnTag: Enhanced Tagging for Discovery. Dublin Core Special NKOS session, 24 September 2008.

Matthews, B; Golub, K.; Jones, C; Moon, J; Lykke Nielsen, M; Tudhope, D. 2008. Enhancing social tagging with a knowledge organization system. ALISS Summer Conference 2008.

Golub, K., Jones, C., Lykke Nielsen, M., Matthews, B., Moon, J., Tudhope, D. 2008. EnTag. Presentation at JISC MDR SIG, 12 February 2008, Birkbeck.

Golub, K., Tudhope, D. 2008. Delivering a terminology registry. LIDA conference, Dubrovnik and Mljet, Croatia, 2-7 June 2008.

Tonkin, E., Baptista, A. A., Pinheiro, S., Hooland, S. v., Resmini, A., Mendiz, E. and Neville, L., 2007. Kinds of Tags: a collaborative research study on tag usage and structure. In: The 6th European

Networked Knowledge Organization Systems (NKOS) Workshop, at the 11th ECDL Conference, Budapest, Hungary.

Pfeiffer, H. D., Zhang, Q., Tonkin, E., Corrado, E. M. and Resmini, A. 2007. Tagging and Social Networks: The Impact of Communities on User Centered Tagging. In: Panel presentation to ASIS&T 2007 Annual Meeting, Joining Research and Practice: Social Computing and Information Science.