Renato Rocha Souza (Federal University of Minas Gerais, Brazil)
Douglas Tudhope (University of Glamorgan, UK)
Maurício Barcellos Almeida (Federal University of Minas Gerais, Brazil)

The KOS spectra:
a tentative typology of Knowledge Organization Systems

Abstract: This work tries to propose a set of evaluation dimensions for the analysis of the Knowledge Organization Systems (KOS), building over previous research and the available literature on the subject. It presents a compiled taxonomy of KOS, a set of tentative characteristics proposed in the literature and the authors' spectra proposal. The full details of the typology are not covered in the scope of the article, but will be available as an ontology in the near future.

1: Aim and scope of the study

The process of representation of knowledge is in the core of many scientific fields, but it seems to be paramount in Library and Information Science (LIS). In fact, LIS takes upon itself the task of organizing and facilitating the retrieval of the registries of information that arise from the knowledge produced in all the other fields. Much of the LIS theories, processes and instruments are dependent on successive abstractions over the relevant characteristics of a chosen world, or the information gathered and processed about this world, registered in information systems and documents.

From Knowledge Organization Systems (KOS) to the inverted indexes of Information Retrieval Systems (IRS); from the surrogate files to the documents' abstracts, representations aim at promoting retrievability of information, or knowledge elicitation. KOS may be considered special cases of representations – or meta-representations – as they are used also for aiding knowledge organization and information registries organization which, in their turn, are also representations. The goal of this paper is to discuss why and how the KOS should be tentatively classified on a new basis, aiming to shed some light to the discussion.

2: Representation and Knowledge Organization Systems

According to Hodge (2000), KOS are at the heart of the library environment. She defines the term KOS as something that:

> …encompass all types of schemes for organizing information and promoting knowledge management. Knowledge organization systems include classification and categorization schemes that organize materials at a general level, subject headings that provide more detailed access, and authority files that control variant versions of key information such as geographic names and personal names. Knowledge organization systems also include highly structured vocabularies, such as thesauri, and less traditional schemes, such as semantic networks and ontologies.

KOS are also important in the Semantic Web, given the need of vocabulary disambiguation and the highly formalized structures needed to allow machine "understanding". Besides the well known web oriented languages for knowledge representation, as RDF and OWL, there are specifications being discussed for a new standard called SKOS, aiming to express LIS KOS in the Semantic Web environment (W3C, 2009). Knowledge of the KOS and its characteristics is necessary for the

responsible design of any IRS or Knowledge Base System, especially in the digital environment, allowing the architects to make grounded decisions on the project.

Despite its importance and however often discussed, there is no agreement on how to evaluate KOS and what would be the parameters or dimensions with which they would be classified. In fact, there is no consensus about the whole range of KOS available, though many candidate lists have been attempted (HODGE, 2000, BERGMAN, 2005, TUDHOPE et al., 2006; WRIGHT, 2006 & 2008; BSI, 2007) There is a reasonably large amount of work in the L&IS and Computer Science literature proposing some kind of classification spectrum, using dimensions such as "semantic level" (OBRST, 2004; BERGMAN, 2005; DACONTA et al, 2005), "ontology level" (LASSILA & McGUINESS, 2001), "ontological precision" (GUARINO, 2006) , "complexity" (SMITH & WELTY, 2001; NISO, 2005, p.17), "structure level and language control" (ZENG & SALABA, 2005) and even "time/money" (BERGMAN, 2005). Those and the related works generally tend to consider small subsets of the structures accepted as KOS, and they almost always take "semantic", "complexity" or "degree of formalization" as their main parameter for comparison, leaving some important aspects behind. Some other works have tried to be more comprehensible with KOS description and evaluation dimensions (SOERGEL, 2001 & 2001b; TUDHOPE, 2004; WRIGHT, 2006 & 2008; ALMEIDA & SOUZA, 2009), but, still, leading to rather different classification structures, according to the characteristics taken in consideration. Based on the available literature and their own previous works on the subject, the authors intend to i) present a compiled taxonomy of KOS; ii), propose a set of useful dimensions to classify them.

3: A taxonomy of KOS

No matter how extensive, all lists attempting to enumerate all the KOS will fail under other eyes' perspective, as long as the different interpretations about what may be called a KOS will lead to different results. Wright (2006, 2008) distinguishes KOS from KRRs (Knowledge Representation Resources) and alternatively names them as "Knowledge Organization Schemes". In fact, the "terminology relating to terminology" is often confusing (HODGE, 2000), as also is the "concept of concept" (KLEIN e SMITH, 2005), what makes more difficult the task of knowledge representation. In most of the previously cited KOS taxonomies attempts, it can be noticed that there is no distinction between types of KOS and languages that can be used for representing them. Regarding this issue, we are adopting a comprehensive and *lato sensu* approach, considering KOS "all types of schemes for organizing information and promoting knowledge management" (HODGE, 2000), and building over all previously cited works. The current discussion excludes any specific instance of a KOS type, although that would form part of future work. The FIG. 1 shows the summary of the KOS collected by the authors.

As seen in the map, the main criterion for division was the KOS structure types, ranging from the *Unstructured Texts* to those that regard *Concepts, Relationship and Layout* as part of the structure. The ones classified under *Term and/or Concept Lists* present simple structures (mainly alphabetical displays, but usually no hierarchies), and the *Concept and Relationship Structures* comprises a large range of structures that present some different degree of relationship expressiveness. The simpler ones present

hierarchies with loose hyponym/hyperonym relationships, but Thesauri include meronomy/holonomy along with some non-specified associative relationship and the Formal Ontologies allow the representation of all sorts of relationship types, depending on the expressiveness of the language used for representing them.

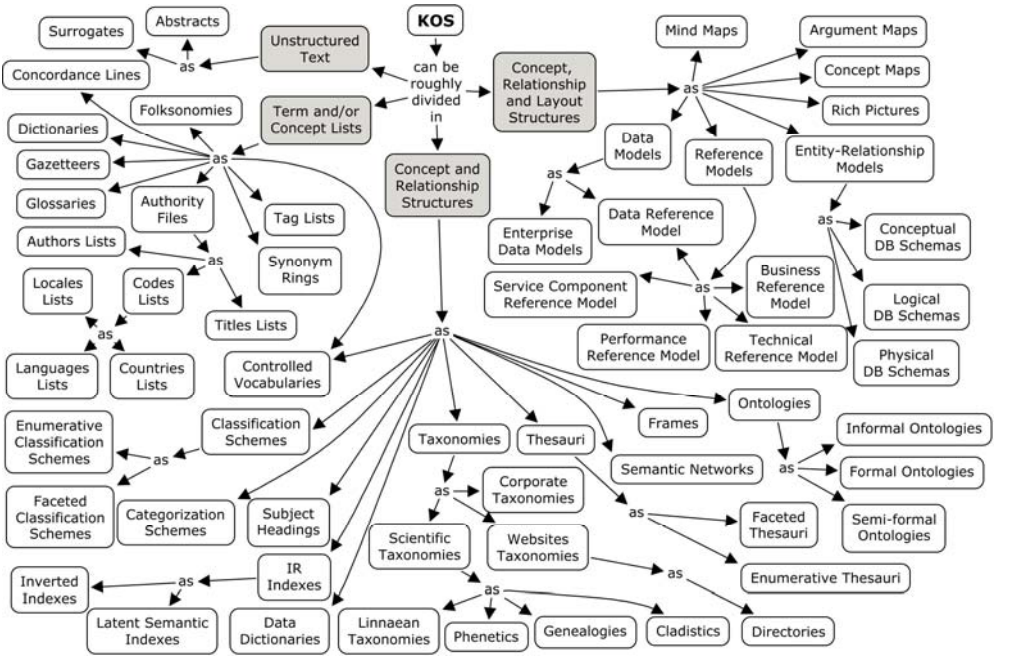

FIG 1: KOS by type

It is evident, at this point, that one cannot escape arbitrariness on trying to classify things in general, and this KOS classification is not an exception. Hence, it is useful to comment the main differences to the previous KOS taxonomies proposed by Hodge (2000) and Wright (2008).

Firstly, in spite of the heterogeneity, we decided to be inclusive, reminding Hodge's definition quoted above (2000). Therefore, Abstracts, Concordance Lines and IR indexes, among others, were included as KOS because: i) they are used for knowledge organization and information retrieval; ii) they promote knowledge management; iii) they are knowledge representation structures. Using the same rationale, we have excluded the standard formats (as HTML, SGML, etc.) and products (as the Wordnet) as specified by Wright (2008), because the former are tools to represent KOS and the latter is a specific instance of a semantic network/lexical database.

Lastly, it is important to highlight that many structures (like *controlled vocabularies*) could have been classified otherwise, according to the sense chosen and example taken. Also, the representation in the FIG. 1 does not allow the breakdown of all concepts, like semantic networks, due to the space limits of the concept map.

4: Dimensions to evaluate

A more difficult task than to enumerate all types of KOS is deciding which would

be the dimensions to evaluate them, along with the scales for it. In this regard, we have compared characteristics adopted from the above cited authors, in order to choose our own set of them. The TABLE 1 shows the results:

Hodge, 2000	Structure and complexity, Relationship between terms, Historical function
Soergel, 2001a and 2001b	Purpose, Coverage of concepts and terms, Sources, Quality of usage analysis, Conceptual analysis and conceptual structure, Terminological analysis, Use of precombination in the index language, Access and display, Format of presentation of the vocabulary, Updating
Tudhope, 2004	Entities (types, coordination, size, depth), Relationships (types, expressiveness, formality), Typical application to objects in domain of interest (purpose), Relationship applying concepts to objects in domain
Wright, 2006 and 2008	Communities of Practice, Systematic resources, Non-systematic resources, Technology orientation, Degrees of indeterminacy, Language & knowledge-oriented standards, Standards bodies
Almeida and Souza, 2009	Representational power, Semantic Expressiveness, Intelligibility (for Humans), Formalization (machine oriented)

Table 1: Dimensions proposed to evaluate KOS in the literature.

Besides covering many characteristics, the cited authors have emphasized different aspects of the types of KOS. Some of them are only meaningful if applied to KOS instances, and some are related to the decisions on the implementation, while the same system can be represented in different ways. They would hardly be used as general template for KOS evaluation and classification, unless some specific framework were explicitly adopted.

5: The KOS proposed spectra

In our model we have tried to capture those aspects in a coherent and integrated way, dealing with some slight differences in the meaning of the characteristics, without letting any important dimension behind. We have added explicitly some dimensions that were only implicitly, or never, covered before such as media or sensitivity. We have also proposed a scale of values for some of the dimensions, in order to allow comparison and evaluation of KOS over certain predefined parameters.

The first attempt to represent the typology was made through building a thesaurus on the subject, but as a knowledge base, it lacked the possibility to express some of the relationships between KOS expressiveness and the other dimensions, as well as dealing with polyhierarchies appropriately. It is also difficult to deal with many different kinds of “is a” relationships between concepts, as they are always treated in a thesaurus as the common hyponym/hyperonym. The authors are now working in an ontology to express the full set of dimensions, instances and evaluation scale and this paper presents the basic dimensions so far incorporated.

These are divided into *intrinsic* and *extrinsic*, and the former are subsequently divided into *essentia*l and *accidental*. The *intrinsic* dimensions are related to the KOS

per se, taken as an isolated entity of the users and the environment. Among the *intrinsic*, the *essential* are closely related to the type (or the broad "class" to which the KOS belong). The *accidental*, by their turn, encompass characteristics of a given instance of the KOS, and can be different for each implementation of the same intellectual work, with different levels of information carried. The *extrinsic* dimensions are related to the environment in which the KOS is used. The dimensions can be examined in the FIG. 2 and are discussed in the following paragraphs.

The *intrinsic/essential* are divided in *structural* characteristics and *standardization*. The *structural* comprises characteristics of the *entities* and *relationships* that are presented in the KOS. For the *entities*, we have *entities types* (i.e. words, strings, numbers, concepts, etc.) and *entities systematisation* (i.e. random, alphabetic, systematic, enumerative, etc.). For the *relationships*, we have the *relationship types* that are present in the KOS (i.e. loose hierarchies, synonyms, hyponym/hyperonym, meronomy/holonomy, etc.). A KOS needs to be evaluated over the types of entities represented and the set of relationships present.

The *standardization* is related to the existence of a body of standards that might set rules or guidelines on the specific KOS structure (i.e. ANSI/NISO Z39.19, BS 5723:1987, etc.).

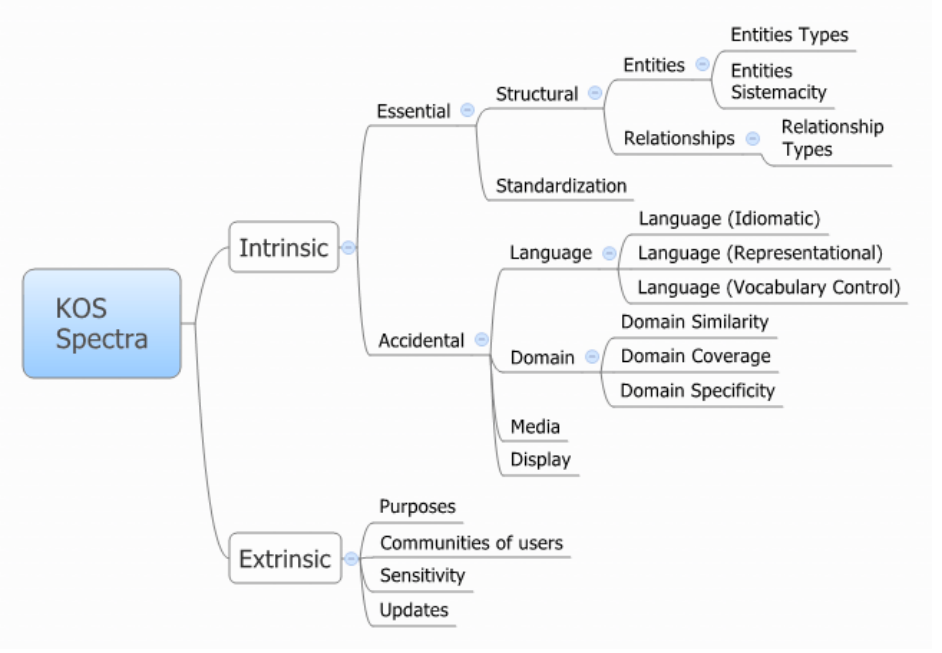

FIG 2: KOS evaluation dimensions, or spectra

The *intrinsic/essential* are *language*, *domain*, *media* and *display*. *Language*, in its turn, is divided in *idiomatic* (monolingual, multilingual); *representational* (i.e. plain text, markup languages, diagrammatic, formal languages, etc.) and *vocabulary control*

(i.e. natural language, controlled language, artificial language). The *domain* dimension is divided in *similarity*, *coverage* and *specificity*. *Similarity* reflects the relationship of the entities to the domain (i.e. instances of the domain, strongly/weakly related domain concepts, etc.). *Coverage* states the depth to which the domain is covered (i.e. shallow, deep) and *specificity* relates to the ontology level (i.e. foundational/upper; task based, cross-domain, domain specific). *Media* is related to the informational substratum (i.e. printed, digital, etc.) and *display* informs the way information is presented (i.e. graphical, simbolic, plain textual, sistematic textual).

The *extrinsic* dimensions are *purposes* (i.e. terminology control, indexing, classification, knowledge representation); *communities of users* (i.e. librarians, webdesigners, translators, etc.); *sensitivity* (i.e. classified, unclassified, public available, etc.) and *updates* (i.e. no update, user updatable, etc.).

Some of the most common dimensions covered by the previous works are emergent properties of those we are covering. “Complexity” or “Semantic”, for example, are related to the structure (entities, relationships), the representational language chosen and also to the domain coverage aspects. Ontology “level” and “precision” are related to the domain similarity, specificity and coverage. The “time/money” aspect is a product of the aspects in many of the current dimensions, as decisions taken in a deployment process.

6: Discussion and future work

The literature reviewed for this paper reveals that we are far from having a consensus on KOS taxonomies and the related terminology. This paper aimed to build on the previous works, in order to contribute to the evolving discussion, presenting some of the most important aspects to be taken in account when evaluating and choosing a specific KOS.

There are many aspects by which KOS can be evaluated, and many of them cannot be taken isolated, as they are cross-dependent. The full spectra should address not only the KOS as an information structure archetype, but also a specific product and all its possible derived instances, displays and codifications, in a similar way that the FRBR treats works, expressions, manifestations and items for bibliographic records (IFLA, 2009).

Due to size limitations, we could not show the full details of the proposed spectra, but the current plans are to develop and publish the full model as a high level ontology in the near future.

References

Almeida, M. B. and Souza, R.R., 2009, The semantic in the Semantic Web: a critical evaluation. *Submitted to the Encontros Biblio Journal, January, 2009.*

Bergman, M. K., 2007, *An Intrepid Guide to Ontologies.* <www.mkbergman.com//?p=374>.

BSI – British Standards Institution BSI 8723: *Structured vocabularies for information retrieval — Guide — Part 3: Vocabularies other than thesauri*. Draft. <www.standardsuk.com/shop/products_view.php?prod=44141>.

Daconta, M. C.; Obrst, L. J.; Smith, K. T., 2005, *The Semantic Web; A guide to the future of XML, Web Services and Knowledge Management.* Indianapolis: Willey. 312p.

Geser, G., 2008, *STERNA Technology Watch Report. Full Report.* STERNA Del.6.5, 10 December 2008. Salzburg Research. (161 pages). <www.sterna-

net.eu/index.php/en/downloads>.
Guarino, N., 2006, Ontology and Terminology: how can formal ontology help concept modeling and terminology? in *EAFT-NordTerm ws on Terminology, Concept Modeling and Ontology*, Vaasa, February 10th, 2006
Hlava, M. M. K., 2007, *Insuring Compatibility and Crosswalks*. <www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2007/presentations/NKOS%202007-HLava.ppt>.
Hodge, G., 2000, *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. <www.clir.org/pubs/abstract/pub91abst.html>.
IFLA, *Functional Requirements for Bibliographic Records: Final Report*. 2009. <www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf>
Klein, G. and Smith, B., 2005, Concept Systems and Ontologies. <ontology.buffalo.edu/concepts/ConceptsandOntologies.pdf>
Lassila, O. & McGuiness, D. L., 2001, *The Role of Frame-Based Representation on the Semantic Web*. Knowledge Systems Laboratory Report KSL-01-02, Stanford University
McComb, Dave., 2003, *Semantics in Business Systems: The Savvy Manager's Guide*. Morgan Kaufmann; 1 edition
NISO, 2005, *ANSI/NISO Z39.19-2003: guidelines for the construction, format, and management of monolingual thesauri*. 2005.
Obrst, L., 2004, Ontologies *& the Semantic Web for Semantic Interoperability*. <www.web-services.gov/OntologiesSemanticWebSemInteropSICOP909-Obrst.ppt>.
Smith, B. and Welty, C., 2001, Ontology: Towards a new synthesis. in, *Formal Ontology in Information Systems*. Eds. Chris Welty and Barry Smith. Pp. iii-x. Ongunquit, Maine: ACM Press. <portal.acm.org/citation.cfm?doid=505168.505201>.
Soergel D., 2001a, *The representation of Knowledge Organization Structure (KOS) data.: a multiplicity of standards*. JCDL 2001 NKOS Workshop, Roanoke. <www.clis.umd.edu/faculty/soergel/SoergelNKOS2001KOSStandards.PDF>
Soergel, D., 2001b, *Evaluation of Knowledge Organization Systems (KOS): Characteristics for describing and evaluating KOS*. JCDL 2001 NKOS Workshop, Roanoke. <nkos.slis.kent.edu/2001/SoergelCharacteristicsOfKOS.doc>.
Tudhope, D., Alani H., Jones C., 2001, Augmenting thesaurus relationships: possibilities for retrieval. in *Journal of Digital Information*, 1(8). <jodi.ecs.soton.ac.uk/Articles/v01/i08/Tudhope/>
Tudhope, D., Koch T., Heery, R., 2006, Terminology Services and Technology. in *JISC state of the art review*. <www.jisc.ac.uk/media/documents/programmes/capital/terminology_services_and_technology_review_sep_06.pdf>.
Tudhope, D., 2006, A tentative typology of KOS: towards a KOS of KOS? in *NKOS Workshop, ECDL*. <www.ukoln.ac.uk/nkos/nkos2006/presentations/tudhope.ppt>.
W3C - World Wide Web Consortium., 2009, SKOS - Simple Knowledge Organization System. <www.w3.org/2004/02/skos/>.
Wright, S. E., 2005, ISO TC 37 Standards: Basic Principles of Terminology. in *NKOS JCDL 2005 Workshop*, Denver. <nkos.slis.kent.edu/2005workshop/TC37.ppt>
Wright, S. E., 2008, Typology for Knowledge Representation Resources. in *NKOS-CENDI* September, 2008. <nkos.slis.kent.edu/2008workshop/SueEllenWright.pdf>.
Zeng, M. L. and Salaba, A., 2005, Toward an International Sharing and Use of Subject Authority Data. in *FRBR Workshop*, OCLC. < www.oclc.org/research/events/frbr-workshop/presentations/zeng/Zeng_Salaba.ppt>.

Web documents have been accessed June 01, 2009.