# Specifying Intersystem Relations: Requirements, Strategies, and Issues

Felix Boteram* and Jessica Hubrich**

* *RESEDA project* – Institute of Information Management, Cologne University of Applied Sciences, Cologne, Germany, <felix.boteram@fh-koeln.de>
** *CrissCross project* – Institute of Information Management, Cologne University of Applied Sciences, Cologne, Germany, <jessica.hubrich@fh-koeln.de>

Felix Boteram is a librarian and research assistant working for the RESEDA project at the Institute of Information Management at the Cologne University of Applied Sciences, Germany. RESEDA aims to develop models for the representation of semantic data in systems for knowledge organization from an interdisciplinary perspective.

Jessica Hubrich works at the Institute of Information Management at the Cologne University of Applied Sciences, Germany. She is team leader of CrissCross, a research project funded by the German Research Foundation (DFG) focused on mapping subject headings of the German Subject Headings Authority File (SWD) to classes of the Dewey Decimal Classification (DDC).

Boteram, Felix, and Hubrich, Jessica. **Specifying Intersystem Relations: Requirements, Strategies, and Issues.** *Knowledge Organization*, *37(3)*, 216-222. 15 references.

**ABSTRACT:** Ideally, intersystem relations complement highly expressive and thoroughly structured relational indexing languages. The relational structures of the participating systems contribute to the meaning of the individual terms or classes. When conceptualizing mapping relations the structural and functional design of the respective systems must be fully taken into account. As intersystem relations may differ considerably from familiar interconcept relations, the creation of an adequate inventory that is general in coverage and specific in depth demands a deep understanding of the requirements and properties of mapping relations. The characteristics of specific mapping relations largely rely on the characteristics of the systems they are intended to connect. The detailed declaration of differences and peculiarities of specific mapping relations is an important prerequisite for modelling these relations. First approaches towards specifying intersystem relations are presented with special respect to linkages between universal decimal classifications and thesauri.

## 1.0 Introduction

Integrating concept schemes into information retrieval systems can considerably contribute to the effectiveness of subject search provided that they are adjusted to the requirements of the specific information space. A controlled vocabulary improves recall and precision whereas (ideally expressive) semantic relations facilitate enhanced retrieval processes like search expansion, search modification, and concept exploration. In an environment of a fast-growing number of distributed information resources indexed with different concept schemes, conventional knowledge organization systems quickly reach the limit of their functionality. Creating an adequate new knowledge organization system from scratch is laborious and expensive whereas existing languages remain appropriate in respect of local contexts. Therefore linkages between indexing languages are established to profit from the existing indexing data and additionally to expand the usage of individual concept schemes with respect to exploratory searching. The functionalities of mappings within the frame of knowledge exploration, however, are seldom the object of research.

We outline a long-term perspective that is directed at differentiated explorative searching processes based on specified relations. The ideas presented in this context are the result of the methodological support of the CrissCross project and the research done by the RESEDA project that develops representational models for semantic data. First, an international comprehensive knowledge organization system that constitutes the overall future perspective of the presented approach is sketched. Then, strategies for defining relation types and for developing adequate inventories that can be applied to both interconcept and intersystem relations are described. After a brief glance at the expressivity and functionality of mapping relations established so far, issues that arise with the effort of specifying mapping relations in particular are discussed and exemplified on the basis of the linkages constructed within the CrissCross project.

## 2.0 Model of an international comprehensive knowledge organization system

The considerations in this paper are based on an interoperability model that constitutes an enhancement of the "spine approach" pursued within Renardus (cf. Day, Koch, and Neuroth 2005; Koch, Neuroth, and Day 2003) or HILT (cf. Nicholson, Dawson, and Shiri 2006; Macgregor, McCulloch, and Nicholson 2007). Compared with other interoperability models like cross-concordances, the advantage is that the number of intermediates and the inevitable shifts of meaning encompassed in crosswalks from one knowledge organization system to another can be reduced to a minimum. Indexing languages can be added easily without the need for modifying the spine or the schemes which are already included.

In our model, the spine consists of a widespread knowledge organization system, most likely a decimal classification, which has been evolved into an internationally acknowledged ontology by means of internationalization (Figure 1). The multiple concept schemes connected to it are indexing languages modified into semantic networks with typed interconcept relations of logical validity. In addition, expressive intersystem relations function as bridges to the spine (Gödert 2008; Hubrich et al. 2008). In retrieval environments, the proposed overall knowledge organization system constitutes an efficient tool for supporting all levels of subject retrieval and exploration processes (cf. Boteram 2009). Information seekers need not necessarily be familiar with the characteristics of the individual indexing languages – all the information they need to know is included in this knowledge system.

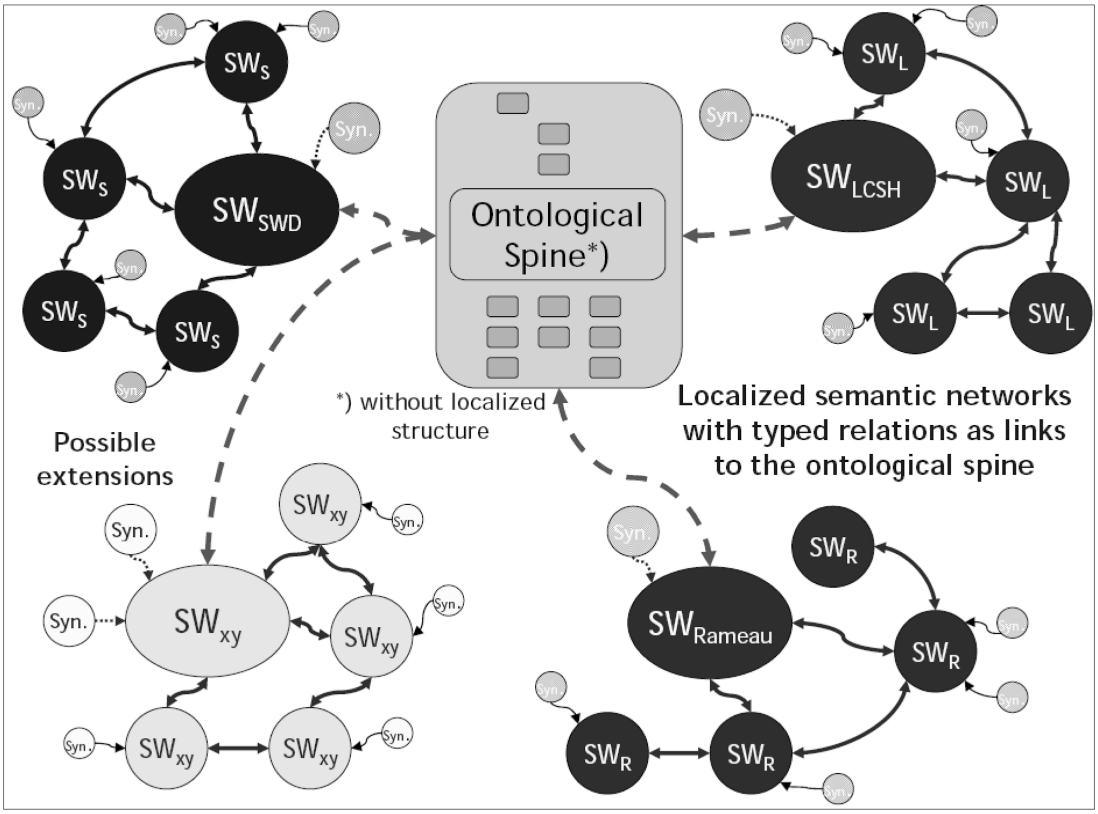


*Figure 1:* Model of a comprehensive knowledge organization system

https://doi.org/10.5771/0943-7444-2010-3-216 https://www.inlibra.com/de/agb - Open Access -

## 3.0 Relational information for exploratory searching

For most information seekers, a query or an exploratory approach into a subject field begins with an individual, sometimes isolated, concept. These individual concepts and the information provided by an intentional definition or synonyms serve as a starting point. If such definitions and additional information are scarce or do not exist at all, the user will take a look at the immediate lexico-semantic neighbourhood and the semantic relations become a helpful and indispensable instrument for orientation and exploration. When it comes to exploring a subject field independently from searching for documents, the representation of the subject field by knowledge organization system structures is equally important. Especially the ordered display of the many subtle and complex interconnections between large numbers of individual conceptual entities provides the structural and conceptual basis for the user's orientation and efficient navigation.

Automated support for such search and exploration strategies—both in a retrieval situation and as a pre-search mechanism—is based on the processing of structural and logical information contained in the indexing language. Such processes can be used for example for preselecting related terms or for suggesting a probable query expansion which can be either discarded or further modified by the end-user. In principle, machine-assisted processes always have to be complemented by the cognitive interpretation of semantic information. Irrespective of whether they are based on automated processing or cognitive interpretation, all strategies rely on thoroughly researched and skillfully designed information contained in the conceptual entities as well as the relational structure of the individual indexing languages and the overall system. The individual terms or conceptual entities and all semantic interconcept relations always have to be understood and analysed together, with a clear focus on their function of supporting and facilitating retrieval strategies. As many modern retrieval techniques and exploration strategies heavily depend on these relational structures, they are the primary focus of our proposal.

For the applications and user scenarios outlined in the preceding paragraphs, the relational information is of particular importance, as it has one considerable advantage: unlike most semantic information on individual concepts, some characteristics of semantic relations can be formalized and therefore made accessible to automated processing. These semantic relations, which integrate all concepts into a meaningful context, enable programs to automatically process and interpret information on a subject area represented by a specific indexing language or a knowledge organization system.

There are several other kinds of information that need to be integrated into an overall system when specifying and defining the characteristics of semantic relations types designed to support advanced retrieval strategies. The primary information is about the semantic content proper, denoting to what kind of relation the specified type belongs. This kind of information is also contained in rather unspecific interconcept relations of conventional thesauri, indicating that a term is more or less general or merely related. Although the semantic content proper can only be accessed and interpreted cognitively, some semantic propositions imply certain logical characteristics which can be interpreted or processed automatically and therefore have to be integrated into the definition of the respective types. The generic hierarchical relation which is referred to as "broader/narrower term" in conventional thesauri implies the logical characteristic of transitivity, which like all other logical characteristics is crucial for a number of important search mechanisms.

## 4.0 Strategies for specifying types and creating inventories of semantic relations

When comparing different strategies for defining and specifying types of semantic relations, the initial situation and the objectives of the specification process as well as the requirements of the anticipated strategies of retrieval and exploration have to be taken into account. We will focus on two different strategies: an inductive approach using strategies of abstraction and starting from highly specific initial relations and a deductive strategy specifying existing semantic relations of little or no specificity at all. The inductive approach, which could also be referred to as a bottom-up strategy, first establishes a rather large basic inventory of highly specific relations. This basic inventory serves as the unprocessed feedstock for a number of consecutive processes of gradual refinement. By aggregating relations representing semantically similar or identical propositions and subsuming them under a new, more general type, the large amount of highly specific relations can be arranged and compacted. By repeating this process of abstraction and generalization several times, various levels of abstraction can be achieved, resulting in a concise and functional set of semantic relations of custom-made specificity whose semantic expressivity is based on and derived from the entire body of conceptual entities of an indexing language.

A method seemingly opposite to this inductive approach is a deductive one, sometimes also associated with a top-down strategy of planning and defining hierarchical inventories. This approach of specifying semantic relations starts from an existing or given set of already defined and established semantic relations as it can be found in conventional thesauri or standard classificatory systems. By means of isolating and defining possible and useful subtypes using a strategy of differentiation and specification, such a small set of initial relation types can gradually be expanded and differentiated until the desired degree of specification is reached. By arranging the acquired set of specified relations in a hierarchically ordered inventory, specific relations can be defined as subtypes of other relations. Regardless of the strategy used to reach a higher degree of specificity and semantic differentiation, the result will always be a larger inventory displaying a higher complexity than what the end-user might expect from previous experiences with conventional vocabularies or thesauri.

Even the most advanced and specified set of semantic relations has to be designed with the end-user in mind. When exploring a subject area the end-user must be able to access the multitude of different semantic relations in a structured inventory allowing for comfortable and intuitive selection of the relevant relations. Therefore, relational inventories have to be arranged in a well-structured, comprehensible array which can be handled intuitively. This can be achieved by arranging the newly created inventories of specified semantic relations in a hierarchically structured taxonomy. In this context, taxonomy is not to be understood in the strict philosophical or epistemological sense, but simply as a thoroughly structured, hierarchical arrangement of elements.

Such a hierarchically organised inventory has several advantages. Firstly, the complexity of excessively differentiated inventories as mentioned above is reduced. Secondly, all semantic relations when arranged in a taxonomy, in which the more specific types of relations can be defined as subtypes of other more general types, form a network of interdependencies and connections. These connections and relations between different types can render even lofty and complex systems more accessible and manageable. Elements of the semantic content as well as some of the logical characteristics can be inherited due to the hierarchical structure. In such a taxonomy, the multitude of highly complex semantic relation types can be traced back and, if necessary, reduced to a small number of basic types.

When designing a similar taxonomy as an experiment for an American Library Association workshop, Dee (1997) was able to arrange a total of 165 thoroughly specified semantic relations in a taxonomy which allowed him to relate every single relation type to one of the three fundamental types of hierarchical, associative and equivalence relations. Although not designed for actual implementation in an existing system, this inventory illustrates the expressive potential and structural advantages of a taxonomy for the definition and arrangement of specified semantic relations.

## 5.0 Inventories of intersystem relations

If we take a glance at the wide range of activities aiming at establishing semantic interoperability, we notice that some relation types used are either similar to or can even be traced back to the basic relations found in almost every structured knowledge organization system, i.e. equivalent, hierarchical, associative relationship (cf. W3C 2009a and b). Besides, a variety of additional relation types exist that are exclusively directed at describing mapping relations. Some of these indirectly reflect mapping strategies and can be derived from the number of concepts of an indexing language linked with a singular concept of another indexing language. One such is, for instance, the "incomplete equivalence" mapping relation, resulting from a one-to-many mapping strategy (cf. IFLA 2009, 17). An overview of mapping types is given by McCulloch and Macgregor (2008, 4-7).

In the course of developing SKOS (Simple Knowledge Organization System) which aims at providing a standard framework for presenting existing knowledge organization systems as part of the Semantic Web, an additional type for intersystem relations is defined that clearly indicates their role as interconnectors: mappingRelation. The sub-properties of this relation type possess machine-readable logical characteristics which enable inferences (cf. W3C 2009a and b). The mapping relations are, however, mainly limited to the basic semantic relations: equivalent, hierarchical, and associative. Although SKOS has become a standard, the definition and specification of mapping relations requires further differentiation, as "doubts remain over the current adequacy of the [...] Mapping

Vocabulary Specification (MVS) for inter-terminology mapping" (McCulloch and Macgregor 2008, 70). The relatively small inventory of relations combined with the constriction of their logical characteristics compromises its usage, especially for modelling relations which connect typologically different indexing languages. Thus using the suggested SKOS relations as top-level relations within a hierarchically structured inventory will require slight modifications in order to facilitate the system's extension with further specifications.

A differentiated inventory of mapping relations based on intellectual mappings between the *Laborline Thesaurus* and the *Library of Congress Subject Headings* was developed by Chaplan (1995). It consists of 19 match types most of which refer to the linguistic form of the headings (e.g. spelling variation, word order variation) rather than to the semantic content proper. These may be valuable for describing linkages that were created automatically but they are not adequate for the characterization of conceptual relations. This may also be the reason why in a test conducted within the HILT project only a minor number of these types turned out to be of practical value (cf. McCulloch and Macgregor 2008). An inventory of specific intersystem relations with semantic content has not yet been developed.

In the following section, we present our own method for specifying conceptual intersystem relations and a detailed analysis of these relations. These are based on a *bottom-up* strategy starting with observations concerning the characteristics of mapping relations and their usage in a comprehensive knowledge organization system. The mappings created within the CrissCross project were closely examined in order to find criteria for specifying relations between typologically different knowledge organization systems.

## 6.0 Specifying intersystem relations

Intersystem relations are specific subtypes of semantic relations which share characteristics with regular interconcept relations. At the same time they differ considerably from them. They represent a special kind of equivalent relation modelled for the purpose of enabling crosswalks between indexing languages. However, the logical characteristics of equivalence, reflexivity, symmetry and transitivity, cannot apply generally to intersystem relations due to the structural, typological and linguistic differences between indexing languages and the varying degrees of complexity of the concepts. Modelling of and specifying intersystem relations requires a new understanding of equivalence and the function of logical characteristics such as symmetry or transitivity for specific mapping relations.

In this section, issues that may arise when describing the semantic content of intersystem relations between typologically different concepts schemes are outlined. They are illustrated with examples of the mappings created within the CrissCross project between the German Subject Headings Authority File (SWD) and the Dewey Decimal Classification (DDC) (cf. Hubrich 2008). The SWD is an indexing language designed for post-coordinated use whose concepts are genuinely context-independent; the specific contextual meaning is constructed only in the act of indexing. On the contrary, the DDC is to a large extent an enumerative scheme with precombined context-sensitive concepts in which several subjects or topics are explicitly united within classes.

When specifying intersystem relations, an initial decision has to be made with respect to the reference points. The mapping is explicitly established between a concept (SWD heading) and a class of concepts (DDC class). There may be rare cases in which the fact that a concept is mapped to a class of concepts appears irrelevant for determining the type of intersystem relations. The experience with the mappings created within the CrissCross project shows, however, that in most cases it does make a difference. Taking a DDC class as a whole as the reference point limits the leeway in specifying the relation type: concepts can either be instances of the class or not. SWD headings may alternatively be described as hierarchically narrower. Subject headings that cannot be unrestrictedly described as part of the corresponding class are difficult to handle. The mapping relation between the German subject heading "Zoologie" (zoology) and the DDC class "590", for example, cannot be easily specified, as the DDC class includes both, the discipline itself and animals as its objects of study. Thus, for our purpose it is more practical to take the concept mentioned in the class heading as reference point in order to achieve a higher degree of expressivity of intersystem relations. The intersystem relation established between the SWD heading "Zoologie" (zoology) and the DDC class "590" which is denoted "Animals" and refers to animals from a biological point of view, i.e. animals in the wild, can be specified as a discipline-object relation. This also supports information seekers who will rather deal with the class headings than with the topics further associated with the class when conducting processes of conceptual exploration.

The "zoology" example belongs to the few cases in which an SWD heading and a DDC class have the same knowledge field context. In most cases, the SWD heading refers or pertains to more than one knowledge field or discipline. This leads to another issue that has to be taken into account when specifying the established mappings: two linked concepts may be semantically different in more than just one aspect which ideally has to be made explicit within the description of the intersystem relation. For example, the SWD heading "Papageien" (parrots) is mapped to two DDC classes: "636.6865" and "598.71." The class "636.6865" deals with parrots as pets whereas the class "598.71" deals with psittaciformes, including parrots from a biological point of view (i.e., parrots in the wild). The relation between the German subject heading and the DDC class "598.71" is characterized by two aspects: it is a kind of hierarchical relation in reference to the main topic of the class and at the same time it denotes a specific context that is different from that in "636.6865." Both are valuable information for exploratory processes and should be integrated in the semantic description of this intersystem relation. At this point, a relation type that is rarely used comes in helpful: the perspective hierarchies. This type explicitly indicates that a certain point of view is provided (cf. Svenonius 2001, 151-52). We suggest a deliberately general relation type indicating a change in perspective. Such a perspective relation holds between the general term "Papageien" and the more specific class "636.6865" whose perspective is focused on a single aspect (parrots as pets). The same applies to the mapping between "Papageien" and "598.71" which can be specified even further as a perspective-hierarchical relation. The specifications of these types are still being developed.

Ideally, intersystem relations should not only support cognitive processes of conceptual exploration but should also facilitate focused queries in heterogeneous information spaces. Providing additional pragmatic information complementing the described semantic content of a mapping relation supports strategies granting access to specific information resources and structuring sets of information resources. Thus, when specifying intersystem relations further aspects may also be considered: in a precombined classification there may be special notations which will be applied when a topic is treated interdisciplinarily or comprehensively. For example, the DDC class "598.71" deals not only with parrots from a biological point of view but also with parrots from an interdisciplinary point of view as this is explicitly mentioned in the classification. Should this information be integrated in the description of the intersystem relation and if so, how should this be resolved? The same question arises in respect of all kinds of application-oriented aspects of intersystem relations. For instance, in the CrissCross project, the established mappings are further specified by "Degrees of Determinacy" that are defined in accordance with the topic-class relations which are inherent to the DDC and coupled with an option for number building in the process of document indexing. This allows for a specific structuring of information resources indexed with DDC notations (Hubrich 2008). The benefit of this should also be taken into account when thinking about strategies for specification of intersystem relations and developing a general theory of mapping relations.

## 7.0 Conclusion

Specifying mapping relations improves the effectiveness of search processes and supports differentiated processes of knowledge exploration in heterogeneous information spaces. It needs to be placed in the broader context of the development of a comprehensive international knowledge organization system including a spine-focused interoperability model that integrates various enhanced indexing languages with specified interconcept relations.

As intersystem relations share many characteristics with regular interconcept relations, the ideas, approaches and strategies for the definition and specification of interconcept relations can equally well be used equally well for the design and application of intersystem relations. However, mapping relations have special characteristics that also have to be taken into account when modelling semantically and logically precise mapping relations. First approaches and considerations in respect to specifying intersystem relations have been presented. Still, further research is essential to guarantee the functionality of mapping relations.

## References

Boteram, Felix. 2009. *Semantic interoperability in an international comprehensive knowledge organisation system*. Pre-print version of the paper presented at the ISKO UK 2009 conference, 22-23

June. Available: http://www.iskouk.org/conf2009/papers/boteram_ISKOUK2009.pdf

Chaplan, Margaret A. 1995. Mapping *Laborline Thesaurus* terms to *Library of Congress subject headings*: implications for vocabulary switching. *Library quarterly* 56: 39-61.

Day, Michael, Koch, Traugott, and Neuroth, Heike. 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 Methology, Amsterdam 2004.* Opladen: Budrich. Available: http://opus.bath.ac.uk/14366/1/day-renardus-paper-v2.pdf.

Dee, Michel. 1997. Appendix B: Taxonomy of subject relationships. In American Library Association, Subcommittee on Subject Relationships, Reference Structure ed., *Final Report to the ALCTS/CSS Subject Analysis Committee*. Available: http://www.ala.org/ala//mgrps/divs/alcts/mgrps/ccs/cmtes/subjectanalysis/subjectrelations/msrscu2.pdf.

Gödert, Winfried. 2008. Ontological, spine, localization and multilingual access: some reflections and a proposal. In *New Perspectives on Subject Indexing and Classification. Essays in Honour of Magda Heiner-Freiling*. Leipzig et al: Deutsche Nationalbibliothek, pp. 233-40.

Hubrich, Jessica. 2008. CrissCross: SWD-DDC-Mapping. In *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen and Bibliothekare* 61 (1): 50-58.

Hubrich, Jessica et al. 2008. Improving subject access in global information spaces: reflections upon internationalization and localization of knowledge organization systems (KOS). In *New Perspectives on Subject Indexing and Classification. Essays in Honour of Magda Heiner-Freiling*. Leipzig et al: Deutsche Nationalbibliothek, pp. 261-67.

IFLA Classification and Indexing Section, Working Group on Guidelines for Multilingual Thesauri. 2009. *Guidelines for multilingual thesauri*. IFLA professional reports, 115. The Hague: IFLA Headquarters. Available : http://archive.ifla.org/VII/s29/pubs/Profrep115.pdf.

Koch, Traugott, Neuroth, Heike and Day, Michael. 2003. Renardus: Cross-browsing European subject gateways via a common classification system (DDC). In McIlwaine, Ia C. ed., *Subject retrieval in a networked environment: proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14 - 16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC*. UBCIM Publications, New Series, 25. München: Saur, pp. 25-33. Available: http://www.ukoln.ac.uk/metadata/renardus/papers/ifla-satellite/ifla-satellite.pdf.

McCulloch, Emma and Macgregor, George. 2008. Analysis of equivalence mapping for terminology services. *Journal of information science* 34: 70-92.

Macgregor, George, McCulloch, Emma, and Nicholson, Dennis. 2007. Terminology server for improved resource discovery: analysis of model and functions. *Second International Conference on Metadata and Semantics Research, 11-12 Oct. 2007*, Corfu, Greece. Available: http://strathprints.strath.ac.uk/3435/1/strathprints003435.pdf.

Nicholson, Dennis, Dawson, Alan, and Shiri, Ali. 2006. HILT: a terminology mapping service with a *DDC* spine. *Cataloging & classification quarterly* 42 (3/4): 187-200.

Svenonius, Elaine. 2001. *The intellectual foundations of information organization*. Cambridge, Mass.: MIT Press.

W3C. 2009a. *SKOS simple knowledge organization system primer.* W3C Working Group Note 18 August 2009. Available: http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/.

W3C. 2009b. *SKOS simple knowledge organization system reference.* W3C Proposed Recommendation 15 June 2009. Available: http://www.w3.org/TR/skos-reference/.

URLs last checked on 5th February 2010