This paper is a pre-print version presented at the ISKO UK 2009 conference, 22-23 June, prior to peer review and editing. For published proceedings see special issue of Aslib Proceedings journal.

ISKO UK

"Content Architecture"

Semantic interoperability in an international comprehensive knowledge organisation system

Felix Boteram

Institute of Information Management (IIM)
Cologne University of Applied Sciences,
Cologne, Germany

Abstract

In this paper, the functional and relational characteristics and requirements for various types of semantic interoperability in a comprehensive international knowledge organisation system are being discussed with regard to an analysis of the underlying retrieval paradigms. Furthermore, this paper analyses the potential benefits and perspectives of the selective transfer of modelling strategies from the field of semantic technologies for the refinement of relational structures of inter-system and inter-concept relations as a requirement for expressive and functional indexing languages supporting advanced types of semantic interoperability.

1. Constructing an international comprehensive knowledge organisation system

In a world where the sheer abundance of information and the heterogeneity of the systems used to access and to document this information requires the unified access to a multitude of different interconnected information retrieval systems and knowledge organisation systems, the interoperability is essential for efficiently managing this information. This calls for the integration of all relevant systems into one comprehensive international knowledge organisation system providing unified and standardised access and advanced retrieval functionalities. To provide all these functionalities across all systems involved, the system must guarantee interoperability at all levels. Achieving the technical, functional and semantic interoperability for such integrated systems is an essential prerequisite of such systems and requires efforts on conceptual, structural and functional levels.

In recent years, the increasing number of linguistically, structurally and even typologically different systems that require integration has lead to the development of a model of a comprehensive international knowledge organisation system involving a "core system" as the central element with the potential of interconnecting a multitude of various systems. This "core system", as proposed by (Gödert 2008), which functions as an ontological spine as well as a switching language (McCullogh and Mac Gregor 2008), offers a hub interconnecting all systems involved: By linking all the individual systems to the "spine" like satellites, they can all be accessed via the core system as the initial starting point. Alternatively, when beginning the retrieval process in a specific system, the spine facilitates switching systems and languages by providing a link to all the other systems. When doing so, the spine and the interconnecting

elements attached to it have to bridge the individual systems' heterogeneity, thereby offering access to the information organised with a system different from the system in which the initial query was performed. Unlike other approaches that interconnect several highly heterogeneous systems in a randomly arranged web of cross-concordances and mappings, the utilisation of a core system minimises the number of intermediaries needed to interconnect any two systems. Attaching all involved systems to only one central system reduces the accumulation of inevitable minimal shifts of meaning that are likely to take place when too many intermediaries are needed to indirectly bridge two different systems via several one-to-one mappings.

2. Characteristics of the "core system"

As the structure and characteristics of the core system are crucial for the functionality of the entire system, designing and constructing it from scratch or selecting and modifying an existent system is of considerable importance. In this section, the necessary characteristics of the "core system" will be analysed justifying the choice of a decimal classification system to act as such a functional core. By designing the core-system as a simple hierarchical classificatory structure, many advantages of this type of knowledge representation can be used: The hierarchical backbone offers a simple, adaptable and functional basic structure. The representation of the concepts as classes, the possibility of a language-independent representation of complex concepts by means of notations and the universality of the most basic semantic inter-concept relations are likely to guarantee maximum compatibility with whatever system is attached to this ontological spine. Although sufficiently specified inter-concept relations are most commonly of particularly great relevance for the semantic coherence of verbal systems, different types of semantic relations can also be found in complex classificatory systems consisting of more than just generic hierarchies. The most important characteristic of the envisaged "core classification" is its differentiated relational structure with its mono-hierarchical core. Therefore, the classification envisaged can function not only as a switching language, but also as a structural backbone for various advanced retrieval strategies.The differentiated relational structure and the specified linking strategies used in this comprehensive system and its "core-classification" facilitates the implementation of models of advanced conceptual interoperability or even a thematic interoperability as described in the following section.

3. Retrieval paradigms and models of semantic interoperability

The degree and quality of semantically interoperable systems, and the models that provide the theoretical and structural fundament to their construction and implementation, always have to be evaluated with regard to the extent of their suitability to support retrieval strategies. However, the association of a particular retrieval paradigm with the respective models of semantic interoperability is not to be understood as a rigid one-to-one attribution, but rather as a loose allocation. This is intended to illustrate the interdependence of retrieval paradigms and models of semantic interoperability, as well as the correlation between retrieval paradigms of varying complexity and efficiency and the level of semantic interoperability they require for an efficient integration of several heterogeneous systems.

3.1. Simple pattern-matching and basic models of semantic interoperability

The most basic method of search is based on matching individual words. Such an approach of a rather unspecific full-text search based on a simple pattern-matching strategy is easy to design and to implement. On the one hand, it neither requires any further processing or enrichment of the documents, nor does it imply the use of a particular indexing language: On the other hand, the lack of a controlled and structured indexing language makes it much more difficult or even impossible for the user to predict the elements of a particular vocabulary used in indexing – a basic precondition in information retrieval. Even when using a highly developed and complex indexing language, the limitations of a search strategy using a simple word match are evident; restricting the search strategies to word-matches is not very efficient due to the high specificity of the individual concepts used. Fuzziness, or a deliberately unspecific expansion of the initial query, can only be defined by a simple truncation of the search string. The semantic interoperability required to support such a strategy can be provided by basic models of semantic interoperability establishing equivalence relations between individual semantically-identical concepts in a one-to-one relation, without further specifying any types of relations.

3.2. Conceptual queries and conceptual interoperability

If the retrieval process is to focus on the conceptual level instead of individual words as the concrete realisations of such concepts, synonyms and other useful access points have to be linked to the representatives of the respective concepts by means of equivalence relations. These conceptual queries integrate all the relevant synonyms into the query formulation thereby enhancing recall. These relations have to be defined and explained to the end-user as the structural elements of larger conceptual systems represented by means of highly-developed indexing languages. Models of semantic interoperability designed to support these paradigms have to account for this structure by providing simple forms of conceptual interoperability.

3.3. Semantic exploration and enhanced conceptual interoperability

If the retrieval strategies are also to offer the end-user features of a conceptual navigation of an exploratory nature accessing the entire semantic environment, and allowing for a progressive and interactive variation of his queries (query-modification), a detailed relational structure is needed. Such an expressive, detailed and multi-dimensional relational structure also has to be taken into account and treated appropriately by the respective model of an enhanced conceptual interoperability. Continually expanding or focussing a search by selecting specific relation types with individual aspects, roles or attributes specified in the definition of the semantic content of the relations is the primary strategy for knowledge exploration within individual systems.

3.4. Thematic exploration and advanced thematic interoperability

An advanced information retrieval system is not only meant to support the end-user with exploring thematic fields, but also to help him clarify, define and verbalise his information need (query-clarification). In order to broaden his own knowledge about the subject field prior to formulating and submitting a query, the system should enable the user to discover and to access new concepts and connections between aspects or details of concepts that were previously unrelated in his personal conceptual representation of the subject field. This requires a system to represent all conceptual and structural information about a subject in a way that can be easily

accessed and explored and that allows the end-user to intuitively integrate new concepts and coherences into his own cognitive representation of the subject field.

4. Relational structures and retrieval functionalities

Modern retrieval strategies, as applied in information retrieval systems, require a comprehensive, differentiated and complex conceptual basis in order to facilitate an efficient information retrieval and exploratory search based on detailed and formalised information on the subject area. The information on the subject area is modelled and represented by an indexing language. Although it primarily serves as a tool for indexing documents or other sources of information, the indexing language itself can also contain information about the thematic context that needs to not be expressed by an indexer using the language, but by the indexing language itself as it is already an integral part of this language.

Specifying both inter-system and inter-subject relations is essential as most advanced retrieval strategies depend on the interpretation of the relational structure between individual concepts, whereas inter-system relations are needed to allow the transit from one system to another. The more the information about the properties of the different relations can be integrated into the individual systems, the more it is possible to use these structures for sophisticated and refined strategies of information retrieval. As different characteristics support different strategies for retrieval, navigation and exploration strategies, it is important to specify these differences and to represent them appropriately using poly-dimensional definitions. Whereas the realisation of a directed and controllable query-expansion relies on the proper definition of the logical characteristics of transitivity of a semantic relation, the exploration of a subject field by the end-user requires the representation of the semantic content.

Both mapping relations and inter-concept relations are an integral part of the designed systems as they contribute to the semantic expressivity of the individual languages and the functionality of the overall system as a whole. The layout, design and definition of a universal, comprehensive, yet concise and functional inventory of relationship types is a crucial precondition for the quality of enhanced semantic interoperability and efficient information retrieval in networked systems on an international level.

Continually expanding or focussing a search by selecting specific relation types with individual aspects, roles or attributes specified in the definition of the semantic content of the relations is the primary strategy for knowledge exploration within individual systems. As these individual systems are only the starting point for the end-users thematic exploration, these strategies, and the information contained in the relations these strategies are based on, have to extend the individual systems. Therefore it is necessary/desirable that inter-system relations also have detailed semantic information, even if the formal specification is more important for the overall function of the mappings. For inter-system relations, the information provided by the relation itself has to alert and inform the user about the nature of linguistic, structural and typological differences between the systems, the possible consequences for the representation of information and adequate search strategies and also about the characteristics and functionality he can expect when entering specific systems. Although users might not be interested in the characteristics of the individual systems or the underlying relational structure but only in the conceptual surface, it should be emphasized that a general knowledge, or even a deeper understanding of the relations and all their specifications, will enhance the understanding of the subject area and the end-users' ability to conduct efficient searches.

4.1. Mapping strategies and inter-system relations

The expressive and functional modelling of semantic information by means of a relational structure as proposed by this paper can improve and enhance both the semantic richness of the individual systems, including the core-system, and the mapping relations interconnecting the various systems with the ontological spine. Using differentiated and highly-expressive mapping relations to interconnect different systems with the core-structure is necessary to ensure semantic interoperability between structurally and typologically different systems like classifications, thesauri or more complex ontologies.

When planning semantic interoperability, the characteristics and requirements of the underlying retrieval paradigms have to be taken into account which in turn demand that the functional, logical and semantic characteristics of the relationship types used to build the underlying relational structure have to be defined in order to express, display and use this information for the orientation and navigation through the semantic network of the conceptual structures.

The strategies used, and the degree of specificity with which these characteristics can be defined, varies considerably depending on the type and therefore also the functions of such relations. The definition of inter-concept relations within individual indexing languages focuses on the semantic content proper. In contrast, the information that has to be defined in order to specify mapping relations interconnecting two different systems has to include formal specifications. Not all types of relations need to be specified logically, but as a basis for basic retrieval strategies, a minimum of logical specification and logical validity allowing simple inferences is necessary. Therefore, the standardised and formalised definition of all logical characteristics is desirable, although the claim of strict overall logical validity is problematic as it renders the entire model contestable, as logical validity can hardly be fully implemented in a system based on natural language.

Whereas the specification of typological differences is particularly important for the design and definition of inter-system relations, the precise definition of logical characteristics is essential for the definition of intra-system, inter-concept relations. As the logical validity is fundamental for inferences that can help discover new information on the subject field. In the context of the construction of inventories of inter-system relations, the questions of the possibilities and preconditions of a partial structural alignment are vital: How can different relational structures of conceptual networks be aligned and harmonised and how can the remaining differences and the resulting incongruencies be handled efficiently?

Particularly when attempting to develop and improve existing systems by augmenting their degree of semantic interoperability, the consideration and careful analysis of the functional and structural differences and peculiarities have to be specifically tailored for the retrieval strategies they are designed to support. The project CrissCross (Hubrich 2008) features a detailed specification of various types of inter-system relationships, that can be used for retrieval purposes and can be the basis for further specification.

4.2. Inter-concept relations within individual satellite systems

Inter-concept relations within individual and independently developed indexing languages and knowledge organisation are vital for the expressivity of these systems and their efficiency in information retrieval. By expanding a search, e.g. by including all narrower terms or related terms into the original query-formulation and afterwards narrowing down the expanded set of results by focussing on a particular type of hierarchical or associative relation characterised by a specific aspect or role defined in the semantic content, the end-user gains a flexible, controlled and differentiated tool for an advanced conceptual search and knowledge exploration. To increase the semantic richness and formal expressivity, the selective transfer of representational techniques and modeling strategies from the field of semantic networks can offer a promising perspective. Integrating ways of representing information in representational frameworks or fully fledged modeling-languages increases the expressivity and functionality of of both, verbal and classificatory languages. Particularly the more detailed representation of semantic relations and the definition of their semantic, logical and funcrional characteristics, can be improved by applying modeling strategies like defining all types of relations and organizing them in well structured inventories.

5. Defining and stratifying international inventories for semantic relations

In order to facilitate strategies of knowledge exploration, the relational structure and the expressiveness of the entire system have to meet certain requirements: The formal specification and the logical characteristics of the semantic relations within these structural frameworks have to be defined in a standardised way. The semantic content has to be expressed in a way that allows the end-user a quick, intuitive and well-structured approach to the information provided by the system. When relevant information about a topic is formalised and modelled according to internationally established agreements, all its relevant aspects and related topics are represented in a way that allows the user to intuitively integrate this background information into his pre-existing knowledge about the subject; relating this newly gained information to his information need can considerably enhance the cognitive basis for a search or explorative access.

The adequate, well-structured and neatly-arranged presentation of the semantic relations, and all the information contained within the relational structure and the bridging system, is vital for the user's awareness, understanding and efficient use in retrieval situations. For each relation, all three types of properties have to be defined and presented in a way that allows the end-user to focus on a specific perspective without losing sight of the other aspects. Offering a selection of semantic relations could be complemented by automatic alerting functions when the formal specifications indicate the crossing of system boundaries and the passage to a different system. Also, the change in logical properties, like the sudden discontinuation of a transitive relation in a long and otherwise fully transitive hierarchy, has to be documented for the user.

To what extent a mutual interdependence and conditionality between formal specifications, logical properties and the semantic content of individual semantic relations may exist will be the subject of further research. If such interdependencies exist, they have to be documented accordingly. This will be an integral part of the process of finding, assessing and conceptualizing different relations and arranging them in inventories. In general, all three types of information on semantic relations have to be taken into account when designing, selecting or accumulating various types of relations in the process of drafting and designing respective typological inventories to accommodate all types of inter-subject and inter-system relations. The

following short descriptions of the different types of specifications and definitions represent the three layers of a multi-dimensional definition for all types of relations.

5.1. Formal specification

A formal specification, as proposed by (Green 2008), can express language differences between systems and different types or structural characteristics of the individual systems. For the end-user, the information - whether he is searching with an underlying classificatory structure or a verbal system - is important to correctly interpret the structural and semantic information contained in the individual concepts and the overall coherent structure.

5.2. Logical characteristics

The second level definition and representation of the relations' logical properties provides the formalised logical substructure. This is the functional fundament for any machine-assisted reasoning designed to access and use information not explicitly modelled in the system.

5.3. Semantic content

The semantic content of the relation has to be specified by a detailed, concise and formalised semantic type-definition. Represented in a neatly ordered inventory, the various relation types are the central device to build the relational structure representing relevant information transcending the focus on individual concepts.If an end-user navigates within a single system along the relational structure provided by the inter-concept relations, he might primarily be interested in the semantic content or the logical properties of the relations. The information about the additional systems or the interconnecting "core system" will become relevant for him when switching between various systems in order to adjust his search strategies to the functional and structural characteristics of the comprehensive system's different indexing languages. It is therefore necessary for all three properties to be defined for all types of relations, even if all the information is not always equally relevant to the user in different situations and for all retrieval and exploration strategies.

6. Prospects

The structural characteristics of the underlying system and the corresponding model of advanced thematic interoperability have to be designed accordingly: The informational content and semantic structure of the entire field have to be made accessible to an exploratory approach supported by a controlled and focused navigation. Models for such a specific and highly-efficient semantic interoperability can improve access to various heterogeneous systems by establishing a link between equivalent concepts documenting the thematic context of the various systems. Designing, adjusting and aligning indexing languages is particularly important when combining and integrating several heterogeneous types of such systems into one connected scheme. A comprehensive international knowledge organisation system consisting of a network of heterogeneous interconnected systems requires a highly-functional relational structure in order to provide an efficient substructure for successful information retrieval. As the principles and strategies of advanced information retrieval systems largely depend on semantic information, new concepts and strategies to achieve semantic interoperability have to be developed.

Dedicated to the development of models for the represenatation of semantic data, the RESEDA project initiated at the Institute of Information Management at the Cologne University

of Applied Sciences promotes these developments by providing the theoretical foundations and developing exemplary inventories and testing their applicability for various subject fields as well as for the entire comprehensive international knowledge organisation systems. The developed modles will serve as functional substructures for information retrieval and new approaches to knowledge exploration.

Keywords:

Interoperability, semantic relations, information retrieval, knowledge exploration, relational inventories

Refrences

Gödert, W. (2008) "Ontological Spine, *Localization and Multilingual Access" In New Perspectives on subject Indexing and Classification* Deutsche Nationalbibliohek, Leipzig, Frankfurt a.M. Berlin, pp.233-240.

Green, R. (2008). Making visible hidden relationships in the Dewey Decimal Classification. How Relative Index Terms relate to DDC classes. In: Culture and Identity in Knowledge Organization. Proceedings of the Tenth International ISKO Conference 5–8 August 2008. p. 8–14.

Hubrich, J. (2008). CrissCross. SWD-DDC-Mapping. In: VÖB-Mitteilungen 61 (3). S. 50 – 58 [Electronic version]. Retrieved October 30, 2008 from: *http://www.univie.ac.at/voeb/php/downloads/vm6120083.pdf*

Macgregor, G. / McCulloch, E. / Nicholson, D. (2007). *Terminology server for improved resource discovery. Analysis of model and functions.* In: Second International Conference on Metadata and Semantics Research, 11-12 Oct 2007, Corfu, Greece. [Electronic version]. Retrieved October 30, 2008 from *http://eprints.cdlr.strath.ac.uk/3435/01/mcculloch.pdf*

McCulloch, E. / Macgregor, G. (2008). Analysis of equivalence mapping for terminology services. In: Journal of Information Science 34 (1) 2008. p. 70–92