t9th NKOS Workshop at
ECDL 2010
Glasgow / UK

Proposal for a presentation

Stratified semantic relations:

Information retrieval and knowledge exploration in distributed knowledge organization systems

Felix Boteram

Efficient strategies of information retrieval and knowledge exploration in heterogeneous information spaces require that all knowledge organization systems are carefully interconnected fully integrated by functional, expressive and formalized intersystem relations. Most strategies for mapping vocabularies focus on linking individual concepts providing simple and unspecified semantic relations. To fully benefit from all the systems involved a more refined and differentiated approach is recommended:

Particularly when mapping knowledge organization systems or indexing languages which differ typologically like classificatory and verbal systems of varying complexity and granularity, providing statements of the semantic similarity or even identity of individual concepts is not enough as it does not reflect the conceptual and relational context in which each individual concept is integrated. Therefore more expressive and differentiated propositions on the characteristics of the relation between the two connected concepts represented by individual terms or classes should be provided.

Intersystem relations can support various retrieval and exploration functionalities, ranging from automatic query-expansions, the modifications of initial queries and widening the range and focus of general subject retrieval, to more explorative strategies.

As these functions vary considerably, the semantic potential invested in the semantic relations constituting has to be expressive and multidimensional, containing various types of information: The technical functionalities require a precise and logically valid formalization of the logical properties of all intersystem relations, while the other functionalities crucial to

the end-user's orientation in the multidimensional relational structure might require semantic information which can only be processed intellectually.

When using the differentiated and expressive relational structure of several mapped indexing languages of interconnected knowledge systems independently from actual documents in the respective repositories for means of knowledge exploration, it is vital that the end-user is informed about the structural and conceptual characteristics and the semantic granularity, and other features and peculiarities of each specific indexing language / knowledge organization system. This is particularly important when supporting an exploratory approach: when exploring interconnected information spaces, following intersystem relations provides access to new and unfamiliar subject areas. When combined with highly specified interconcept relations within a system they can provide invaluable orientation and navigational support in the course of an explorative search conducted by an end-user with little or no previous knowledge of the newly discovered field.

As the idea of complete semantic equivalence is problematic, it is essential that mapping relations provide relevant and precise information on the similarities, commonalities and also the differences between two concepts thereby also accounting for their varying levels of complexity and the granularity of the overall system. This is particularly important when linking general concept schemes like universal classifications with highly specified thesauri or subject heading authority lists of individual fields.

The presentation will assess and discuss strategies of integrating these various types of information, suggesting the use of a stratified or multi-tiered definition of intersystem relations. Information invested in sophisticated mapping relations can be designed to support various functionalities. Formalized information can be used to support machine-based functionalities. For example, investing mapping relations with degrees of determinacy indicating the precision and extension of the semantic congruency of two terms can be used for ranking and selection purposes.

This approach suggests using a stratified definition of intersystem relations. In this context the concept of stratification is to be understood as the detailed differentiation and modular multi-layer type-definition of intersystem and interconcept relations distinguishing various types of information. The three layers include (A) a typological specification containing system-

specific information, (B) the logical properties and (C) the semantic content of the semantic relations.

(A) At first, a typological specification is necessary to alert the end-user when switching from one type of system to another, like when going from a verbal system to a universal classification or vice versa. As the context of an individual concept and the structural characteristics of a particular indexing language are vital to the concept's proper interpretation these differences have to be taken into account when achieving interoperability as a precondition for switching systems.

(B) The second level, defining and representing the relations' logical properties, provides the formalized logical substructure. This is the functional fundament for any machine-assisted reasoning, designed to access and use information not explicitly modelled in the system.

(C) On a third level, the semantic content of the relation has to be specified by a detailed, concise and formalized 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. The presentation will discuss to what extent and in what way information about the semantic environment of the interconnected terms as well as the relational structure of the entire systems can be integrated into the mapping relation and how these can be used to support information retrieval and knowledge exploration.

The integration of these three levels of formal, logical, and semantic specification into a stratified relational structure provides the expressive and functional substructure for knowledge representation, information retrieval and conceptual and semantic interoperability in a comprehensive international knowledge organization system. The principles, strategies and proposed inventories will be illustrated and discussed with examples.