

Documentation thesauri and classification schemes as macrostructures of knowledge vs. terminology as microstructure of knowledge

– How to combine knowledge micro- and macrostructures for knowledge management –

Introduction

Future knowledge processing and management in all environments – here in the meaning of vertical dimensions, such as subject fields, as well as of horizontal dimensions, such as language communities – needs a unified theory of knowledge organisation and knowledge structuring based on science theory and harmonised methods. Terminology science as developed over the years from the first findings of Eugen Wüster via the standardising activities of ISO/TC 37 “Terminology and other language resources” to recent research and development carried out at many research institutions over the world, provides a theoretically sound basis for managing the ‘terminologies’ (viz. concepts and concept systems represented by symbolic representations, such as terms, graphical symbols, formulae etc., and descriptive representations such as definitions, explanations, complex formulae etc.) as the ‘microstructures of specialised knowledge’.

Terminology data modelling

The scientific ‘objects’ of terminology science are first of all ‘concepts’ corresponding to (concept) classes of (real – i.e. material and immaterial) objects. They are conceived according to the classificatory and categorising human intellect and represented for communication purposes by means of linguistic and/or non-linguistic symbols (or better more general: concept representations). Concepts are the constituting elements of any scientific theory. They are used to formulate propositions and sets of propositions into theories and theory systems. Seen under this perspective concepts are units of scientific-technical thinking (when they conceived), units of knowledge (if recognised by a certain community of experts) and units of communication (if communicated by means of spoken or written communication). No specialised information and knowledge can do without terminology in this broad meaning. As a consequence of this, knowledge management cannot function efficiently and effectively without taking modern terminology science and its methodology in a variety of applications as well as the tools developed for these applications into account.

In a terminology database most of the data are centring around or are needed to process and manage data on different kinds of concept representations (s. ANNEX 1). These can be seen as the core elements of every terminology record, containing as a rule several or many such representations representing one concept. All of them form the contents structure of the terminology record with – depending on the data model chosen can comprise many sub-records according to the repeatability within language and the repeatability by language (s. ANNEX 2). Most of the other data – from a methodological point of view – are additional data to qualify the representation in question – like the layers of an onion (s. ANNEX 3). For language engineering purposes it might be interesting, that the same model can – slightly adapted – be used for processing and managing also linguistically more complex SPL (special purpose language) units, such as terminological phrasemes and SPL boiler plate texts (s. ANNEX 2).

Indexing and retrieval languages (documentation languages)

Under the theory of ordering classification schemes and documentation thesauri (originally developed as ‘documentation languages’ for information indexing and retrieval purposes) are also based on concepts and concept systems: namely theme concepts and their systems representing the ‘macrostructure of specialised knowledge’. On the one hand such documentation – or indexing and retrieval (I&R) languages – languages are needed to subdivide large quantities of data (on all kinds of real objects in the above meaning) into manageable portions under ‘pragmatic’ considerations – including the scientific objects called concepts. No terminology database – containing terminological records, which can be considered as and are often called ‘documents’ – can be efficiently maintained without a classification scheme and/or a thesaurus (or a ‘classaurus’ as a combination of both). On the

other hand virtually all objects can be 'managed' by means of a combination of a concept classification (viz. terminology) and one or more documentation languages.

In I&R language databases the main elements also represent 'concepts' – however representing 'themes'. The respective data collections often look very much like terminology data collections, but are different in terms of conceptual structure, relation to SPL communication and first of all function. In order to clarify the situation, one must use different terms in order to clearly identify the different types of terminology and I&R languages (s. ANNEX 4).

New types of meta-structures

Given the quantitative increase of subject fields, their ever increasing terminologies and the existence of many thousand documentation languages for different purposes, a new tool is needed in order not to get lost in the 'mindspace' (intellispace) of knowledge. For this purpose many scientists conduct research on new 'metastructures' of knowledge, which allow 'panorama views', 'birds eyes' views' etc. on 'landscapes of knowledge' in order to keep our scientific and technical knowledge somehow in a coherent system (notwithstanding some 'black wholes' and theoretical contradictions, which however do not threaten this metasystem of knowledge). In addition 'metadata' systems are conceived in order to make all these systems for different concept representations compatible and interoperable with each other. For more object-oriented data, 'ontologies' are a focus of research.