

## **A Back Bone Thesaurus for Digital Humanities**

by Maria Daskalaki and Lida Charami (FORTH)

***In order to integrate the new digital technologies and the traditional research in the humanities and enable collaboration across the various scientific fields, we have developed a coherent overarching thesaurus with a small number of highly expressive, consistent upper level concepts which can be used as the starting point in order to harmonize the numerous discipline and even project specific terminologies into a coherent and effective thesaurus federation.***

Digital humanities is a relatively new interdisciplinary field which involves the integration of the emerging new digital technologies with traditional research in the humanities in order to ensure the long term preservation of knowledge and enable collaboration across the various humanities fields. This integration, however, is not as easy and straight-forward as it sounds as it entails the creation and use of a “common language” in the form of a classification scheme that would enable the communication between different disciplines. The actual state-of-play, however, is somewhat different, with different groups of scholars usually developing their own jargon in order to build thematic vocabularies that are discipline or even application specific. As Barry Smith [1] observes “different databases may use identical labels but with different meanings; alternately the same meaning may be expressed via different names”. This inevitably introduces an unnecessary fragmentation of knowledge that inhibits research and collaboration. Given this situation, it becomes obvious that there is an urgent need to create a common scheme that would enable interoperability between the different scholarly fields and thus support researchers by giving them access to uniformly marked up datasets for query and by providing a guide for the production of systematic terminologies which would avoid methodological errors that typically lead to inconsistencies and incompatibilities between classification systems.

Despite the clear challenges to the construction of such a unifying framework, we argue that “a global knowledge network” [2] is feasible. Building on concentrated research on classification methodology to address these challenges, we have developed a system, the Back Bone Thesaurus (BBT), that aims to allow access, compatibility and comparison across heterogeneous [3] classification systems.

This system, elaborated after the research of a multi-disciplinary team of experts, is based on a consistent methodology designed to enable intersubjective and interdisciplinary classification development and integration without forcing specialists and experts to abandon their own terminology. The methodology relies on the principle of faceted classification and the idea that a limited number of top-level concepts can become a substantial tool to harmonize the numerous discipline and even project specific terminologies into a coherent and effective federation [4] in which consistency can progressively be carried from the upper layers to the lower ones.

In order to define the BBT facets, we started by examining existing vocabularies from the fields of History, Archaeology, Ethnology, Philosophy of Sciences, Anthropology, Linguistics, Theatre Studies, Musicology and History of Art, we analyzed these data using a bottom up strategy in order to discover appropriate upper level concepts. The research consciously avoided the projection of any preconceived formulations of knowledge onto the material, precisely in order

to identify the broader, fundamental categories that would be applicable across the Humanities. The top level concepts thus derived, despite their generality, can be easily specialized in order to express the particular meaning of the different domains without leading to inconsistencies. This is achieved through the detection of the intensional properties of these concepts and the rigorous and proper application of the IsA relationship.

In order to express the exact meaning of the top level terms/concepts defined in the BBT, we provide explicit definitions on the basis of their intensional properties which cannot be replaced without loss of meaning since they are the sum of the properties, state of affairs, qualities that constitute the necessary and sufficient conditions for identifying a term/concept.

The BBT facets are further subdivided into a number of hierarchies using the IsA relation which dictates that the scope of each narrower term subsumed under a broader term must fall completely within the scope of the broader term. In other words, every subsumed term must belong to the same inherent category as its broader concept. Using the IsA relation as the criterion for building the BBT hierarchies ensures that consistency is maintained since all narrower terms must possess all the fundamental properties attributed to the broader concepts of the hierarchy into which they are subsumed. In other words, by using the IsA relation we avoid categorical errors that may result from the subsumption of terms under facets or hierarchies, which have properties different than those of the higher level terms. The strict, proper application of the IsA relation thus serves as a logical control to avoid contradictions and achieve objectivity and interdisciplinarity.

The BBT is an ongoing work and we are currently in the process of reviewing the material we have at our disposal in order to identify additional facets and hierarchies.

Further information about the BBT can be found at <http://www.backbonethesaurus.eu/>

#### References:

- [1] B. Smith, "Ontology", in *The Blackwell Guide to the Philosophy of Computing and Information*, L. Floridi, ed. Oxford: Blackwell, 2004, p. 158.
- [2] M. Doerr, D. Iorizzo, D. "The dream of a global knowledge network – A new approach", in *Journal on Computing and Cultural Heritage*, 2008, 1(1), p. 1.
- [3] M. Daskalaki, M. Doerr, "Philosophical background assumptions in digitized knowledge representation systems", in *Dia-noesis: A Journal of Philosophy*, 2017, (3), 17-28.
- [4] R. Kramer, R. Nikolai, C. Habeck, "Thesaurus federations: loosely integrated thesauri for document retrieval in networks based on Internet technologies" in *International Journal on Digital Libraries*, 1997, (1), 122-131.

Please Contact:

Dr. Maria Daskalaki, ICS - FORTH, Greece

[daskalak@ics.forth.gr](mailto:daskalak@ics.forth.gr)