
Approaches to Collaborative Multilingual Thesauri Development in Educational Context

Goran Zlodi^{*†1} and Tomislav Ivanjko

¹University of Zagreb, Faculty of Humanities and Social Sciences – Croatia

Abstract

The paper discusses approaches, outcomes and experiences of an ongoing project of collaborative Croatian terminology contribution to the Art and Architecture Thesaurus (AAT) through the cooperation of university teachers, master level students and museum professionals. The project is conducted on the University of Zagreb Faculty of Humanities and Social Sciences (FHSS). A model proposed in the project can accelerate the otherwise time-consuming process of developing multilingual thesauri through greater student engagement while achieving multiple educational goals on real-world tasks. The methods of quality control and student engagement from crowdsourcing projects methodologies are applied, but they were revised and improved with respect to the specific needs and requirements of the educational context.

The process is segmented into the following stages:

1. Generating a corpus of the relevant terms.

Two thousand most frequent terms from four top facet (materials, techniques, objects and periods) were selected from the databases of partner heritage institutions.

2. Translation and linking.

Translation of terms into English and linking them to relevant AAT concepts (accompanied with references to relevant reference literature and lexicographical sources) is performed by students. Mapping of Croatian and English terms for specific concepts is further implemented through semantic technologies.

3. Quality control.

Methods include peer-checking among students and teacher supervision. Concepts that are particularly demanding (e.g. complex art techniques) are examined by a separate group of students, where examination also includes detailed literature research and comparison. The final stage of quality control is provided by scholars and museum professionals from the corresponding field.

4. Including the terms into AAT.

After the quality check, the terms would be included in the AAT and made openly available as Linked Open Data (LOD), provided by the Getty Research Institute. Further open formats for thesauri exchange (Zthes, RDF, JSON) would be provided. This will allow the reuse of project results by other national and international vocabulary projects and create a basis for enriching metadata and enabling multilingualism. That would increase the visibility, accessibility and interoperability of diverse national heritage in an international context,

*Speaker

†Corresponding author: gzlodi@ffzg.hr

thereby supporting multicultural communication at human and machine level.

During the mapping and validation of different concepts from monolingual and multilingual thesauri, via different methods and tools (SPARQL queries, LOD matching, using APIs), participants are becoming aware of linguistic and cultural differences in meaning and vocabulary construction (classifications, selection of preferred terms, fluid spatio-temporal boundaries etc.).

In this paper, we would like to discuss how participants (students, heritage professionals, scholars) perceive different aspects of vocabulary development and control and how they practice scholarly primitives (especially annotating, comparing, referring and representing) in the context of the ongoing developments in the methods and technologies related to controlled vocabularies and ontologies in the field of Digital Humanities.

Keywords: vocabularies, thesauri, crowdsourcing, education, scholarly primitives, linked open data, digital humanities