

# The Model of Collaborative Terminology Contribution to the *Art & Architecture Thesaurus*: Application and Improvement of Crowdsourcing Methods in an Educational Context

Goran Zlodi\*, Tomislav Ivanjko\*, Petra Štefičar\* and Marta Marochini\*

\* University of Zagreb, Faculty of Humanities and Social Sciences / Department of  
Information and Communication, Zagreb, Croatia  
gzlodi@ffzg.hr

**Abstract** - The paper elaborates the model of collaborative terminology contribution to the *Art & Architecture Thesaurus* (AAT) through the cooperation of university teachers, students and museum professionals. The elaborated model can accelerate the otherwise time-consuming process of developing multilingual thesauri through greater student engagement while achieving multiple educational goals on real-world tasks. This model includes a statistical selection of the most frequent Croatian art and heritage domain terms and their links to relevant AAT concepts (accompanied with references to relevant literature and lexicographical sources). This process is conducted by students. Subsequent quality control methods include peer-checking and teacher supervision. Mapping of Croatian and English terms for specific concepts was further implemented through semantic technologies (SPARQL queries). Concepts that are particularly demanding are examined by scholars and museum professionals in the corresponding field. After the quality check, the terms would be included in the AAT and made available to open access as Linked Open Data (LOD), provided by the Getty Research Institute. The methods of quality control and student engagement are applied from crowdsourcing projects methodologies, but they were revised and improved concerning the specific needs and requirements of the educational context.

**Keywords** - *thesaurus, linked open data, crowdsourcing, citizen science, education, learning outcomes, Bloom's taxonomy*

## I. INTRODUCTION

The Getty Vocabulary Program is a program of The J. Paul Getty Trust, a California charitable trust, a cultural and philanthropic institution dedicated to the preservation, research, presentation and interpretation of the world's artistic heritage. That program has developed a structured vocabulary called the *Art & Architecture Thesaurus* (AAT), containing concepts and terms, hierarchical, equivalent and associative relations, sources, and scope notes on objects, materials, techniques, styles, periods, and other concepts related to art, architecture and the broader field of cultural heritage. The AAT is particularly important in the contemporary information and communication environments (e.g. Semantic Web), as an important resource that enhances access to information about art,

architecture, and other material cultural heritage. AAT is available to the public under the Open Data Commons Attribution License (ODC-By) 1.0. as LOD (Linked Open Data) on the SPARQL endpoint at vocab.getty.edu.

In November 2018, the University of Zagreb, Faculty of Humanities and Social Sciences (FHSS) and The J. Paul Getty Trust signed the *Agreement For Vocabulary Contributions*, according to which the FHSS, as the contributing institution would contribute approximately two thousand (2000) terms translated in the Croatian language to the AAT.

## II. THE PROJECT OF TRANSLATING OF AAT TERMS TO CROATIAN

Thus, according to the methodology described by the AAT editorial guidelines for contributing large translations [1], the project (hereinafter referred to as Project) of translating terms from source (AAT) to the target language (Croatian) was established. According to the same methodology, this type of mapping represents selective mapping or selective translation as a type of partial mapping, as AAT source terms are chosen as most frequent terms from the collection management databases of partner heritage institutions.

The method of translation and mapping applied in this Project is a direct mapping. This type of mapping refers to the matching of concepts one-to-one in the source and target languages and assumes that there will be the same meaning and level of specificity between the two terms representing the concept in each language [1]. In the process of mapping or translating involved are only descriptors in the source language and descriptors in the target language. The Project uses a hierarchical structure from the AAT.

### A. Participants

The importance of the controlled terminology and the need for its alignment with the AAT Thesaurus has long been discussed in the Croatian museum community. The last systematic approach to creating terminology in a museum context was in 1987 through the publication of the

*Classification of museum and gallery objects by type of object* [2]. As this is a very time-consuming process and which requires the involvement of a significant number of experts, no systematic project has been established so far.

Thus, the initiative to launch the project came from the FHSS, which involved the collaboration of academics, researchers, museum professionals and students.

As stated in the AAT editorial guidelines, “It is critical that translation work is carried out by authorized experts in content, language, and thesaurus construction. Identifying multilingual equivalents is not a simple matter of providing literal translations or looking up words in a dictionary.” [1] Table I shows expertise required and how they are distributed to participants.

The expert group includes university professors and museum professionals from partner institutions (for example, the Ethnographic Museum Zagreb, contributed with previously translated terms from past projects).

The project is managed and coordinated by the FHSS Information and Communication Sciences Department. Students from two courses participated in the project: *Museum Documentation I* (Undergraduate program (bachelor's level)) and *Museum Documentation II* (Graduate program (master's level)). Two student groups are involved in the project: Student Group 1 (SG1), which includes all students in courses, and Student Group 2 (SG2), which includes specially motivated students who have agreed to invest more of their free time than they were required by regular study program.

TABLE I. DISTRIBUTION OF REQUIRED EXPERTISE TO THE PARTICIPANTS

Expertise required	Participants	
	Experts (Museum professionals, academics, researchers)	Student Group 2 (SG2) – advanced students)
Content experts	Museum professionals, academics, researchers.	Students in humanities study groups (e.g., art history, archaeology, ethnology, cultural anthropology).
Language experts	Translators, museum professionals, academics, researchers.	Students of language study groups (e.g. English, Italian).
Thesaurus experts	Academics (information sciences and museology professors).	Students of information sciences (e.g. museology, librarianship, archivists).
Technical experts	Academics (information sciences and museology professors).	Students of information sciences.
Project coordinators	Academics (information sciences and museology professors).	Students of information sciences.

In the approach presented, most of the work can be done by students (mapping, finding references), while final quality control and solving complex conceptual and terminological problems are performed by experts. As the applied approach entails greater student engagement

(especially in the first stages of the project), and the term translation projects require expertise, additional methods of quality checking were provided.

### III. WORKFLOW OF TRANSLATION AND MAPPING

#### A. Stage 1. – Selection of terms

Firstly, a criterion was established to select terms from the following domains: materials, techniques, styles and periods, object names and roles of agents. This criterion was applied because the terms in these categories are present in the structure of catalogue entries used in printed or online catalogues. Subsequently, the most frequent terms from the above categories were selected from the partner institutions' collection management databases. The selected categories of terms, examples of terms and AAT top facets and hierarchy names are specified in Table II.

TABLE II. SELECTED CATEGORIES OF TERMS AND CORRESPONDING AAT TOP FACETS

Selected categories of terms	Example	AAT top facet and hierarchy name
Materials	marble (rock)	Materials Facet
Techniques	direct carving	Activities Facet - Processes and Techniques
Object Names (types of objects)	sculpture (visual works)	Objects Facet
Roles of Agents	sculptor	Agents Facet
Styles and Periods	Early Renaissance	Styles and Periods Facet

#### B. Stage 2. – the selection of reference sources

In order for a Croatian term to enter the AAT and join other terms in other languages that refer to the same AAT concept, it is necessary to find references to that term in at least three relevant reference sources in Croatian. This phase of the project consisted of collecting these sources, based on an analysis of the required and optional literature of a series of courses in the departments of art history, archaeology and ethnology and cultural anthropology and further consultations with experts.

#### C. Stage 3. – terms translation and mapping

This stage of translation and mapping was carried out by Student group 1 (SG1).

- Each student is assigned several terms in Croatian for the translation and mapping of which he or she is responsible.
- Each individual Croatian term (e.g. *litografija*) is searched in the following online encyclopaedias and lexicographic sources:
  - The Croatian encyclopaedia [3]
  - The Proleksis encyclopaedia online [4]
  - STRUNA - Database of Croatian Special Field Terminology [5]
  - Croatian Wikipedia

- A reference (record URL) is recorded for each term found.
- STRUNA often includes information about equivalents in other languages (e.g. *lithography*), which can serve as a starting point for finding terms in AAT. Also, using the term in Croatian Wikipedia, one simply follows the link to the record in other languages (students also note a reference to the term in English Wikipedia).
- Then, students search for that term on the online version of the Arts & Architecture Thesaurus. If they find it, they add the following data from the AAT:
  - The Subject\_ID for the master AAT concept.
  - The preferred term from the AAT (usually in the plural form, often with a qualifier in brackets).
  - The alternative descriptor (AD) – the term most often found in the singular form in the AAT.
  - The term which would be displayed in catalogue entries – in the AAT it is most often the alternative descriptor (without the qualifier) – as the equivalent to the Croatian term which would be found in reference sources.
- Mapping of Croatian terms for specific concepts was further implemented through semantic technologies by SPARQL queries.
- Notes are recorded in case of problems with the translation of a particular term.

#### D. Stage 4. – proof-checking and quality control

This stage of translation and mapping was carried out by Student group 2 (SG2).

- Students first proof-checking the terms that SG1 students had worked on. During the proof-checking, the English translation of the term provided by an SG1 student was compared to the translations found in bilingual dictionaries and other reference literature. In case of incorrect translation, the error is recorded and categorized in order to further improve the system of assigning terms to students of the particular study groups. The correct term is recorded and mapped according to the appropriate AAT concept.
- Students conduct selected literature research (while SG1 students consult only online sources, SG2 students consult also extensive printed sources).
- If students find references, they note the following:
  - Reference to do the specific source.
  - The number of the page on which reference was found.

- Definitions of concepts from each source are also recorded.
- Notes are recorded in case of problems with the translation of a particular term.

#### E. Stage 5. – Review and final verification

This stage of quality check and translation is carried by experts.

- Concepts and terms that are categorized by students as demanding are given for review and translation to experts of different specialities, depending on the domain to which the concepts belong.
- After review and verification terms are designated as suitable for export to the AAT by the required XML format.

### IV. EDUSOURCING AND BLOOM'S TAXONOMY

Since the proposed model relies heavily on the inclusion of students in the process one of the main concerns (apart from ensuring the quality of translation and mapping), is ensuring student engagement while at the same time achieving multiple educational goals on real-world tasks. Similar efforts involving the assistance of a large number of volunteers to address various research problems in the field of science have often been brought together under the term *citizen science* and have become a reliable and scalable solution that supports scientific endeavours [6]. A large number of domains are used by them, making the most of the work of volunteers to assist professionals and scientists in processing a large amount of research data that they enhance their work. The combination of human computing and social computing has proven to be very effective, not only in achieving original scientific goals but also in reaching unforeseen discoveries made by community members [7]. This approach has become extremely popular in recent years, and a large number of papers have explored the various applications of systems and initiatives gathered under the term crowdsourcing [8]. Accordingly, an increasing number of implemented projects base their business models on the work of the many, including science and education.

For applying such approaches in the educational context, we propose a term *edusourcing*. Edusourcing is an approach that applies methods and practices derived from crowdsourcing projects to student engagement activities and is tailored for addressing the specific needs and requirements of the educational context.

Since one of the main goals of the educational process is reaching the desired learning outcomes at the highest possible cognitive levels, there is a need to examine the model in that aspect. One of the most widely applied taxonomies of learning outcomes is Bloom's taxonomy that identifies "six levels within the cognitive domain, from the simple recall or recognition of facts as the lowest level, through increasingly more complex and abstract mental levels, to the highest order which is classified as evaluation" [9]. Table III lists the desired outcomes

according to the aforementioned taxonomy from the lowest to the highest level.

TABLE III. COGNITIVE DOMAINS OF EDUSOURCING ACCORDING TO BLOOM'S TAXONOMY

Level	Actions	Outcome
Knowledge	define, list, name, order, recognize, relate, recall, repeat	Gathering term definitions enables students to define, list and recognize basic field terminology
Comprehension	classify, discuss, explain, identify, indicate, report, review, select	Translation of the terms enable students to classify, discuss and select appropriate terminology within the field
Application	apply, choose, demonstrate, sketch, solve, use, write	Linking the terms to the AAT enables students to apply and demonstrate their knowledge in a structured environment
Analysis	analyse, calculate, compare, contrast, discriminate, examine, experiment	Comparing the work process and results between different students enables them to analyse and compare their own knowledge
Synthesis	assemble, construct, create, design, develop, formulate, prepare, propose, write	Incorporating the results within a coherent AAT structure enables students to gain valuable insight into the construction and maintenance of controlled vocabularies
Evaluation	assess, attach, choose, compare, predict, rate, select, evaluate	Performing quality control enables a student to compare and evaluate different approaches through peer-checking

As we can see from the Table III, the proposed model of edusourcing can successfully reconcile the contribution through the quality translation of multiple multilingual thesaurus terms, while at the same time reach desired learning outcomes at the highest possible cognitive levels through greater student engagement.

## V. PERCEIVED BENEFITS

In this chapter, we would like to discuss how the participants (students, heritage professionals, scholars) perceive different aspects of vocabulary development and benefits from the project.

### A. Benefits for vocabulary

Among other contributions vocabulary we can single out the following:

- Enriching the Art & Architecture Thesaurus with the first contributions of Croatian terms.
- Contribution to world knowledge in the context of Linked Open Data (LOD) and Semantic Web.

### B. Benefits for students

Student benefits can be viewed from several complementary perspectives:

- The Project provided information science students with an insight into some of the tasks of heritage professionals via practical work, and an experience of participating and contributing the vocabulary from different perspectives (as a user/as a professional). For example, students of museology have gained various experiences and knowledge such as getting to know AAT in more detail, getting acquainted with problems and limitations in translation, a deeper understanding of the need to standardize museum terminology and documentation, resolving doubts and problems in a group with an interdisciplinary approach etc.
- It enabled interdisciplinary work for double major students who study both information sciences combined with languages or humanities (art history, archaeology, ethnology and cultural anthropology).
- To the students of humanities studies, this project has brought various experiences and knowledge such as an expansion of their own vocabulary of English terms, developing the skill of critically reviewing the literature used for references, experience in interdisciplinary problem-solving, etc.
- Language students have gained experience in translation of professional terminology thus connecting their two different majors.
- For both groups the Project enabled the practical application of theoretical concepts, principles and methods to solve a problem or task in a real-life situation, creating new and integrating existing knowledge into a functional whole.

### C. Benefits for university professors and study program

Benefits for the educational process can be categorized into the following:

- Developing a scalable solution for different courses and educational levels.
- Supervising the educational process by controlling the quality of work.
- Possibility of better insight into the applied knowledge of students.
- More active collaboration between students and teachers and increasing student motivation.
- Connecting different study groups and perspectives through the prism of digital humanities.

### D. Benefits for Croatian cultural heritage community

Among the other perceived benefits are the following:

- Identification of most frequent terms used for indexing of cultural heritage in Croatian museums.

- Translation and mapping Croatian terminology to AAT.
- Improvement of multilingualism which enables greater visibility and accessibility of Croatian cultural heritage.

## VI. CONCLUSION

The proposed model of collaboration between academics, museum professionals, and students can successfully facilitate the contribution of terms to a multilingual thesaurus. By applying edusourcing approach, greater student engagement can accelerate translation and mapping of terms and ensure the quality of the process, while at the same time reach desired learning outcomes at the highest possible cognitive levels.

## REFERENCES

- [1] P. Harpring, "Contributing large translations : guidelines for multilingual equivalency work". J. Paul Getty Trust, 2015. [Online]. Available: [http://www.getty.edu/research/tools/vocabularies/guidelines/aat\\_4\\_4\\_appendix\\_d\\_contributors.html#4\\_4\\_2](http://www.getty.edu/research/tools/vocabularies/guidelines/aat_4_4_appendix_d_contributors.html#4_4_2). [Accessed: 27-Jan-2020].
- [2] "Dokumentacija i klasifikacija muzejskih i galerijskih predmeta", *Muzeologija*, vol., br. 25, str. 3-117, 1987. [Online]. Available: <https://hrcak.srce.hr/95072>. [Accessed: 27-Jan-2020].
- [3] „Hrvatska enciklopedija“. [Online]. Available: <http://enciklopedija.hr/>. [Accessed: 27-Jan-2020].
- [4] „Proleksis enciklopedija“. [Online]. Available: <https://proleksis.lzmk.hr/>. [Accessed: 27-Jan-2020].
- [5] „Struna : Database of Croatian Special Field Terminology“. [Online]. Available: <http://struna.ihjj.hr/>. [Accessed: 27-Jan-2020].
- [6] Rick Bonney i dr., „Next steps for citizen science“, *Science* 343, 6178 (2014): 1436-1437.
- [7] Ramine Tinati i dr. „An investigation of player motivations in Eyewire, a gamified citizen science project“, *Computers in Human Behavior*, 73, (2017): 527-540.
- [8] Daren C. Brabham, *Crowdsourcing*. (Cambridge: The MIT press, 2013)
- [9] J. Rutkowski, K. Moscinska, and P. Jantos, "Application of Bloom"s taxonomy for increasing teaching efficiency—case study," unpublished, International Conference on Engineering Education ICEE-2010, July 2010. Available: [http://www.ineer.org/Events/ICEE2010/papers/W13A/Paper\\_1292\\_1417.pdf](http://www.ineer.org/Events/ICEE2010/papers/W13A/Paper_1292_1417.pdf). [Accessed: 27-Jan-2020].