The EuroScholar Ontology: Arqus Termbase Enhancement (Short Paper)

Esther Castillo-Pérez[*,†], and Silvia Montero-Martínez[†]

University of Granada, C/Puentezuelas 55, 18002, Granada, Spain

Abstract

Higher education is now a globally interconnected setting, bringing together diverse languages and cultures. In these multilingual environments, effective terminology management is crucial for facilitating seamless and high-quality communication. This study specifically addresses the enhancement of the Arqus Termbase prototype, the terminological database of the Arqus European University Alliance, encompassing seven languages. The primary focus is on refining the conceptual structure through the creation of an ontology. This advancement aims to optimise the end-user experience and enhance terminological information within the realm of European higher education.

Keywords

Terminological database, ontology, conceptual structure, Arqus Termbase

1. Introduction

The academic world, a hub of diverse cultures and languages, is highly internationalised. Universities frequently launch programmes and activities to fortify internationalisation—a process integrating an international, intercultural, or global dimension into higher education aims and functions [1]. While this process signifies quality and excellence [2], its intercultural and multilingual nature introduces challenges. This holds true for entities like the European Union (EU), where standardising terminology is crucial due to the multilingual nature and volume of documents. The ultimate aim is to ensure consistency and correctness of terms, facilitating easier, faster, and more reliable communication processes [3]. In such complex contexts, the development of centralised terminology resources, exemplified by IATE, is indispensable.

The design and management of these tools, falling under the scope of terminography or terminology management [4], encompass tasks such as extraction, compilation, description, documentation, and dissemination of terms [5]. In normative contexts, eliminating inconsistencies and controlling new and obsolete terms are integral to the workflow. These systematic tasks contribute to the accurate use of terminology and enhance profitability, productivity, and institutional image [6,7,8]. However, creating terminology resources for systematic terminology management also requires constructing the underlying conceptual

3rd International Conference on "Multilingual digital terminology today. Design, representation formats and management systems" (MDTT) 2024, June 27-28, 2024, Granada, Spain

[*] Corresponding author.

[†] These authors contributed equally.

esthercaspe@ugr.es (E. Castillo-Pérez); smontero@ugr.es (S. Montero-Martínez)

0000-0002-1030-7693 (E. Castillo-Pérez); 0000-0002-2064- 401X (S. Montero-Martínez)

structure for a coherent organisation of specialised knowledge tailored to the tool's end-users. In the academic realm, there are limited terminological databases dedicated to handling institutional and higher education terminology. A couple of exceptions include the bilingual terminological database UGRTerm from the University of Granada (https://ugrterm.ugr.es/en/) and the multilingual Arqus Termbase (ATB) prototype, from the Arqus European University Alliance (https://arqusterm.ugr.es).

This preliminary work seeks to improve the conceptual structure of ATB. The primary objective is to formulate a methodology for crafting a new ontology featuring both hierarchical and non-hierarchical relations, specifically designed for the context of European higher education. This endeavour involves mapping term entries to this knowledge structure, enabling end-users to visualise entities and relations inherent in the teaching, learning, and research processes.

2. Ontology development

Ontology development and terminology management are essential in structuring knowledge within specialised domains. Organisations utilise taxonomies and ontologies for organising both knowledge and terms [9]. Taxonomies provide a hierarchical structure based on a single criterion, while ontologies enable more complex structures, incorporating both hierarchical and non-hierarchical conceptual relations.

In knowledge engineering, ontologies are defined as explicit specifications of a conceptualisation [10]. Studer et al. [11] emphasise that ontologies must not only be explicit but also formal and interoperable, possessing the capacity to be shared. Roche's [12] definition, aligned closely with the field of terminology, describes ontologies as "a conceptualisation of a domain," providing a formal definition of concepts and their relations to describe a reality shared by a community of practice. Similarly, Moreno Ortiz [13] defines ontologies as "conceptual and terminological descriptions of a shared understanding of a specific domain."

Therefore, ontologies are formal and explicit representations of concepts and relations, fostering a shared understanding of a specialised domain and promoting interoperability and semantic clarity. They serve as a foundational framework for information systems, facilitating data integration and advanced reasoning. Ontologies stand out for their comprehensive inclusion of attributes and relations, offering more detailed descriptions for domain entities compared to terminologies [14].

Numerous methodologies exist for ontology construction, including METHONTOLOGY [15] and On-To-Knowledge Methodology [16], among others. While these methodologies share common features like conceptualisation and formalisation, distinctions emerge based on the specific scope or field of application. In the realm of terminology, Temmerman and Kerremans' Termontography [17] integrates a sociocognitive approach with a semasiological perspective for ontology development and analysis. Roche's onomasiological model, Ontoterminology [18], is employed to construct notional systems and operationalise terms.

In this study, a semasiological proposal was implemented to construct a preliminary EuroScholar Ontology (EuSO), which takes into account the features of Arqus Termbase (ATB) and end-user requirements. Building upon the methodology employed for the NavSAFE ontology [19], the proposed approach introduces novel stages.

3. Materials and methods

Since ontologies establish conceptual structures fostering more coherent, systematic, and interoperable terminological resources [20, 21], the primary aim of this study is to enhance the conceptual framework of the ATB, operating under the premise that the creation and utilisation of ontologies in the realm of terminology yield significant benefits.

Thanks to the incorporation of an ontological model (EuSO), more intricate relations will be introduced, enhancing the richness of the resource. To achieve this, the subsequent sections delineate the principal characteristics of the multilingual termbase, the European Higher Education Area (EHEA) corpus, the TermoStat tool, and the WebProtégé software, employed in ontology development. Ultimately, the proposed methodology aligns with the ontology's scope, the intricacy of the domain, and the resources at hand.

3.1. Arqus Termbase

The Arqus Termbase, a prototype multilingual termbase developed by the Arqus European University Alliance, consolidates terminology in the context of the Alliance to foster communication coherence and fluidity by centralising term management and eliminating inconsistencies. ATB currently includes seven languages of partner universities and incorporates a simple and advanced search system. Users can also access terms through various access points, including an alphabetical and a thematic list [22]. Such efforts are anticipated to contribute to improving the overall image of the Alliance, as outlined by Montero-Martínez and Castillo-Pérez [23].

The ATB is structured into four terminology collections: Arqus ad hoc Glossaries, Arqus Alliance Proprietary Multilingual Terminology, Arqus Partner's Institutional Terminology, and Arqus Multilingual European Higher Education Terminology. This paper focuses on European higher education terminology, particularly on the current conceptual structure depicted in Figure 1. The existing structure is a taxonomic model adapted from the UGRTerm database [24]. However, to enhance the resource with more specific information and improve user experience, enriching this structure with more complex relations that interlink entries is deemed essential.

Figure 1: Hierarchy for the Arqus Multilingual European Higher Education Collection

3.2. The European Higher Education Area corpus

The EHEA corpus was created ad hoc for this study. It currently contains 1,639,445 words and comprises 85 documents in English. The documents were chosen based on Buendía Castro and Ureña's criteria [25], which emphasise authority, content, and design. This selection process prioritised factors such as authorship, topic relevance, availability of articles or full texts, and accessibility. They include official and specialised texts within the European higher education domain, such as books, reports, communiqués, declarations, statements, and guidelines from official bodies like the European University Association, ministers of the European Higher Education Area, and the European Students' Union.

To compile the corpus and conduct linguistic analysis, the Sketch Engine tool (www.sketchengine.eu) was used. This software provides numerous functionalities, including Word Sketch for summarising grammatical and collocational behaviour, Wordlist for generating frequency lists, and Concordance for displaying instances of words or phrases in context.

3.3. WebProtégé

Protégé (http://protege.stanford.edu) is a free and open-source software developed by the Stanford Center for Biomedical Informatics Research at Stanford University School of Medicine. It is designed for constructing domain models and managing terminologies, ontologies, and knowledge bases across various application domains.

This work utilises the web-based version known as WebProtégé. This tool enables the editing, development, uploading, and sharing of ontologies in a collaborative and user-friendly environment. It features change tracking, revision history display, and facilitates discussions among collaborators. WebProtégé is fully compatible with the latest OWL 2 Web Ontology Language, and other formats like RDF/XML, Turtle, XML, or OBO. It is noteworthy that this web version is cross-compatible with Protégé Desktop.

3.4. Methodological framework for ontology development

To develop the EuSO a preliminary semasiological approach with six phases has been implemented. However, it should be noted that ontology development is an iterative process, allowing for revisiting and refining previous steps. Therefore, the order proposed in this methodology may vary throughout the ontology's development. The stages include the following processes:

1. Domain delimitation: Prior to ontology development, it is crucial to define the specific domain under consideration.
2. Terminology selection: Following domain delineation, the relevant terminology from ATB is chosen for the selected field.
3. Conceptualisation: The ontology's classes, hierarchical and non-hierarchical organisation, and conceptual relations are defined.
4. Formalisation and implementation: The ontology must be formalised and implemented using the chosen ontology development tool.
5. Expert validation: Ensuring the ontology's validity and accuracy involves validation by domain experts, who verify its quality and specificity.

6. Update and maintenance: This final stage involves updating the ontology based on expert feedback and recommendations, as well as ongoing maintenance and improvement.

This termino-ontological proposal draws inspiration from the methodology employed in the development of the NavSAFE ontology by Losey León and Corpas Pastor [26]. Notably, it introduces variations, including domain delimitation, expert validation, and a concluding stage for updating and maintenance.

4. The EuroScholar Ontology

Following the proposed methodology, tailored to the needs and scope of the Arqus Termbase, a tentative model for the EuSO has been developed[1]. For this preliminary work, domain delimitation was restricted to the teaching, learning, and research processes. This choice was made based on the significant impact these processes have on the development of higher education. Moreover, this initial study only contemplates the first four stages for ontology development.

Once the domain and the scope for ontology development were determined, terminological selection encompassed all terms within the ATB associated with the teaching, learning, and research processes. These terms were drawn from official sources such as the EuroVoc Thesaurus and the TESE (Thesaurus for Education Systems in Europe). However, term selection was enhanced with the EHEA corpus and, to validate the relevance of terms in the ATB, they were cross-checked against terms from the corpus using the TermoStat tool (http://termostat.ling.umontreal.ca).

Subsequently, in the conceptualisation stage, gathered classes were organised based on hierarchical relations already documented in the taxonomic model of ATB. The classes were drawn from the ATB's existing entries, covering a spectrum of general and specific concepts pertaining to the teaching, learning, and research processes. These classes provided insights into the who, where, when, and how of these processes, delineating the involved actors, settings, methodologies, and temporal aspects. The EHEA corpus also plays a vital role in such conceptualisation. The extraction and analysis of lexical constructions and contexts, related to selected terms, provides enriching conceptual information, in the form of hierarchical and non-hierarchical relations that enhance conceptual structuring.

Examples 1 (a, b, c), identified in the EHEA corpus, contain lexical units and constructions that formalise entities such as the people involved in the teaching, learning and research process (*doctoral candidates*, *undergraduate students, students, student body*) (AGENT/PATIENT), educational programmes (*Bachelor programmes, Master programmes, doctoral programmes*) (INSTRUMENT), and higher education institutions (HEIs) (*EUA member institutions, institutions of the European Higher Education Area*) (LOCATION: HIGHER-EDUCATION-INSTITUTION). Furthermore, relevant conceptual relations in the teaching and learning process can be abstracted: 1(b) STUDENT *enrolls in* BACHELOR/MASTER/DOCTORAL PROGRAMME, 1(c) STUDENT *studies at* HEI, STUDENT *enrolls in* EHEI.

[1] The EuSO is currently under development and not yet available online.

a) "(...) research teams of doctoral candidates and undergraduate students to build teamwork."
b) "(...) only 17.5 % of all students are enrolled in Bachelor programmes, compared to 75 % in Master programmes and 7,5 % in doctoral programmes."
c) "(...) which is more than half of the 17 million students studying at EUA member institutions, or about a quarter of the student body enrolled in the institutions of the European Higher Education Area."

Some of the terms collected for analysis in this work pertained to the entity UNDERGRADUATE, given its significance as a principal participant in the teaching, learning, and research processes. Conceptual analysis for the UNDERGRADUATE class preceded manual formalisation and implementation in WebProtégé. Figure 2 depicts class hierarchy, where UNDERGRADUATE is a subordinate of the entity STUDENT, itself a subordinate of AGENTS and PATIENTS involved in the process of teaching, learning, and research (Figure 2). Moreover, non-hierarchical relations such as *enrolled_in*, *studies*, and *studies_at* are also shown, inherent to the entity UNDERGRADUATE.

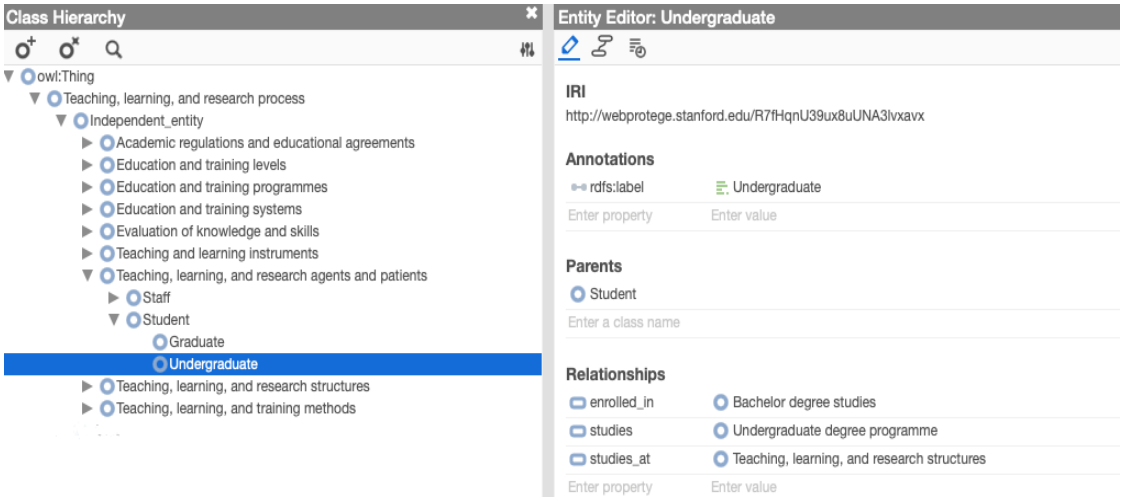

Figure 2: The entity UNDERGRADUATE (STUDENT)

The ontological properties for UNDERGRADUATE are also visualised in Figure 3, with a partial representation of the relations and attributes. Thus, visual representations are employed to illustrate the preliminary conceptual information that links the UNDERGRADUATE entity with other classes, already conceptualised and formalised within the WebProtégé software.

The implementation of the EuSO in the ATB will introduce relations in terminological entries, forming conceptual networks and structures for visual representation within the termbase. Integrating knowledge representation structures, as illustrated in Figure 3, will streamline end-user queries. This will ensure that information gathered in the ontologised conceptual structure is presented visually, enhancing knowledge transfer and user experience with the ATB.

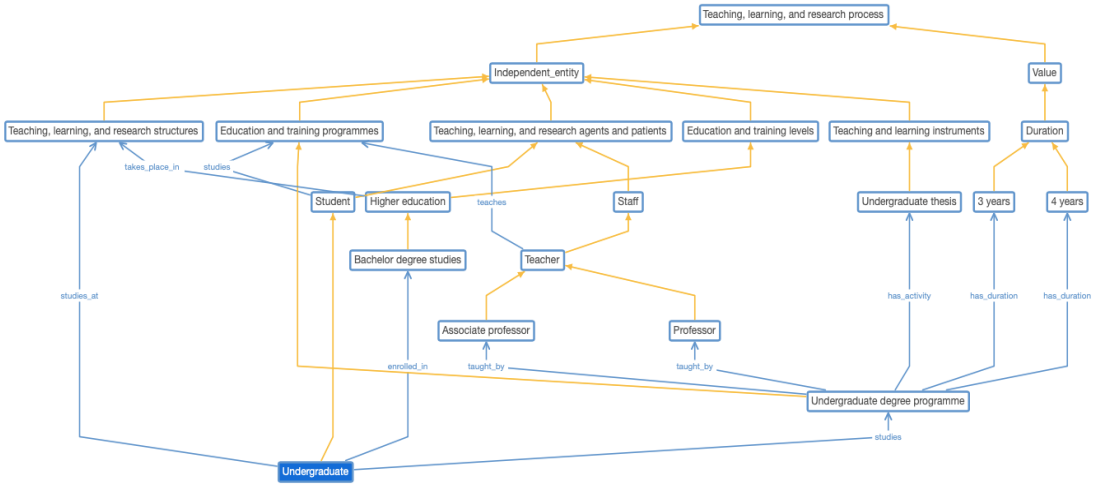

Figure 3: Properties of UNDERGRADUATE (STUDENT)

5. Conclusions

Arqus Termbase has been conceived as a tool for enhancing communication within the Arqus European University Alliance. In such a context, ensuring comprehensive information and user-friendly access for non-expert users is essential.

This preliminary work focuses on proposing a methodological approach for the development of the EuroScholar Ontology—an ontology for the teaching, learning, and research process in European higher education. The goal is to enhance and validate the conceptual structure of Arqus Termbase. This ontology will implement relations in terminological entries, creating conceptual networks and structures for visual representation within the termbase.

Such enhancement will ultimately contribute to improving end-user experience with ATB. In other words, this research aims to make ATB a more comprehensive resource for various user profiles, including the Arqus university community, as well as other stakeholders. The development of this ontology could also benefit other specialised resources in European higher education, by providing a conceptual framework for improvement and validation.

Acknowledgements

This work was supported by the Spanish Ministry of Universities [*Predoctoral Grant for the Training of University Lectures (FPU), FPU20/03089*], the European Commission [*Arqus European University Alliance, 612247-EPP-1-2019-1-ES-EPPKA2-EUR-UNIV*], the European Commission Erasmus + European Universities [*Arqus II, ERASMUS-EDU-2022-EURUNIV-1*], and the Spanish Ministry of Science and Innovation [*Integración transversal de la cultura en una base de conocimiento terminológico medioambiental, PID2020-118369GBI00*]. In addition, this work has been carried out using the WebProtégé tool.

References

[1] J. Knight, Higher Education in Turmoil: The Changing World of Internationalization, Sense Publishers, 2008. URL: https://brill.com/view/book/9789087905224/BP000007.xml.

[2] F. Maringe, N. Foskett, Globalization and internationalization in higher education: Theoretical, strategic and management perspectives, Continuum International Publishing Group, 2010.

[3] S. Rirdance, A. Vasiljevs, Towards Consolidation of European Terminology Resources, EuroTermBank Consortium, 2006. URL: https://www.eurotermbank.com/assets/data/EuroTermBank_Towards_Consolidation_of_European_Terminology_Resources.pdf.

[4] S. Montero-Martínez, P. Faber, M. Buendía-Castro, Terminología para traductores e intérpretes, 2nd. ed., Ediciones Tragacanto, 2011.

[5] TerminOrgs, Terminology Starter Guide, K. Warburton (Ed.), 2016. URL: https://www.terminorgs.net/downloads/TerminOrgs_StarterGuide_V2.pdf.

[6] TerminOrgs, Terminology Starter Guide, K. Warburton (Ed.), 2016. URL: https://www.terminorgs.net/downloads/TerminOrgs_StarterGuide_V2.pdf.

[7] K. -D. Schmitz, D. Straub, Successful Terminology Management in Companies, tcworld GmbH, 2018.

[8] R. Granda, K. Warburton, Terminology Management as Data Management, in: Proceedings of the 2001 Conference of the Centre for Advanced Studies on Collaborative Research, CASCON '01, IBM Press, 2001, pp. 4.

[9] TerminOrgs, Terminology Starter Guide, K. Warburton (Ed.), 2016. URL: https://www.terminorgs.net/downloads/TerminOrgs_StarterGuide_V2.pdf.

[10] T. R. Gruber, A translation approach to portable ontology specifications, in: Knowledge Acquisition, vol. 5, London, UK: Academic Press, 1993.

[11] R. Studer, V. R. Benjamins, D. Fensel, Knowledge engineering: Principles and methods, IEEE Transactions on Data and Knowledge Engineering 25 (1-2) (1998) 161-197.

[12] C. Roche, Le terme et le concept: fondements d'une ontoterminologie, in: Terminologie & Ontologie : Théories et Applications (TOTh), Annecy, France, 2007, pp. 1-22. URL: https://www.researchgate.net/publication/220801842.

[13] A. Moreno Ortiz, Ontologías para la Terminología: Por Qué, Cuándo, Cómo, Tradumàtica 6 (2008).

[14] D. L. Rubin, N. F. Noy, M. A. Musen, Protégé: A tool for managing and using terminology in radiology applications, Journal of Digital Imaging 20(1) (2007) 34-46. doi: 10.1007/s10278-007-9065-0.

[15] A. Gómez-Pérez, M. Fernández-López, O. Corcho, Ontological Engineering, Springer Verlag, 2004.

[16] Y. Sure, S. Staab, R. Studer, On-To-Knowledge Methodology (OTKM), in: S. Staab, R. Studer (Eds.) Handbook on Ontologies. International Handbooks on Information Systems, Springer, 2004, pp. 117–132. doi: 10.1007/978-3-540-24750-0_6.

[17] R. Temmerman, K. Kerremans, Termontography: Ontology Building and the Sociocognitive Approach to Terminology Description, in: E. Hajicová, A. Kotesovcová, & J. Mírovský (Eds.), Proceedings of CIL 17, Matfyzpress, 2003. URL: http://cvc.ehb.be.

[18] C. Roche, Le terme et le concept: fondements d'une ontoterminologie, in: Terminologie & Ontologie: Théories et Applications (TOTh), Annecy, France, 2007, pp. 1-22. URL: https://www.researchgate.net/publication/220801842.
[19] M. A. Losey León, G. Corpas Pastor, La terminología del dominio de la seguridad de la navegación marítima en inglés y en español: Formalización ontológica a partir de recursos digitales y corpus especializados, Peter Lang, 2023. doi: 10.3726/b20209.
[20] C. Vargas Sierra, El léxico especializado y las ontologías, in: E. Alcaraz Varó, J. Mateo Martínez, F. Yus Ramos (Eds.), Las lenguas profesionales y académicas, Ariel, 2007, pp. 41–52.
[21] I. Durán-Muñoz, M. R. Bautista-Zambrana, Applying Ontologies to Terminology: Advantages and Some Disadvantages, Hermes-Journal of Language and Communication in Business 51 (2013) 65-77.
[22] S. Montero-Martínez, E. Castillo-Pérez, Arqus Termbase: Multilingual Terminology for the Arqus European University Alliance, Languages in European University Alliances: challenges and issues, Conseil Européen pour les Langues/ European Language Council, 2022.
[23] S. Montero-Martínez, E. Castillo-Pérez, Arqus Termbase: Terminology for the Arqus Community, 3rd Wrocław Terminological Meeting, TERMOS 2023 "Terms and Their Users", 2023.
[24] S. Montero Martínez, UGRTerm®: Gestión terminológica para el ámbito académico, in: P. Sorbet & V. del Valle Cacela (Eds.), Traducción e Interpretación Especializadas en ámbito Panhispánico, Peter Lang GmbH, 2023, pp. 177-190. URL: http://ebookcentral.proquest.com/lib/ugr/detail.action?docI.
[25] M. Buendía Castro, J. M.Ureña Gómez-Moreno, ¿Cómo diseñar un corpus de calidad? Parámetros de evaluación, Sendebar 21 (2010) 165-180. URL: http://revistaseug.ugr.es/index.php/sendebar/article/view/372.
[26] M. A. Losey León, G. Corpas Pastor, La terminología del dominio de la seguridad de la navegación marítima en inglés y en español: Formalización ontológica a partir de recursos digitales y corpus especializados, Peter Lang, 2023. doi: 10.3726/b20209.