Evaluating Large Language Models for Terminology Extraction from Multilingual Technical Standards [*]

Ewelina Kwiatek[1,*,†] and Małgorzata Kodura[1,†]

[1] *University of the National Educational Commission, Kraków, Poland*

Abstract

This paper examines whether contemporary LLMs can independently provide reliable terminology extraction and glossary compilation from specialized multilingual standards, or whether hybrid workflows with expert validation are still required. Three AI tools, ChatGPT 5.2, Google Gemini 3 Pro and Claude Sonnet 4.5, were evaluated using the technical standard *EN 13318: 2000 Screed material and floor screeds – Definitions* and its Polish version as input. Each tool was tested with two prompting strategies: a user designed prompt and a tool-generated prompt. The results highlight differences in multilingual document processing, output usability, and system behavior. While all tools demonstrated potential to accelerate terminology extraction, certain limitations and inconsistent performance were also observed.

Keywords

terminology extraction, technical standards, AI tools

1. Introduction

The rapid advancement of artificial intelligence and natural language processing technologies has opened new opportunities for automatic knowledge extraction from technical documents. Among these opportunities, terminology extraction seems to occupy a central role, enabling the identification of domain-specific concepts and definitions that constitute the foundation of expert communication among manufacturers, engineers, inspectors, and standardization bodies. Accurate and systematic extraction of terminology is particularly important in safety-critical areas, where misinterpretation of specialized vocabulary can lead to regulatory non-compliance, safety hazards and, ultimately, legal actions. Technical standards, especially those issued under the ISO and EN frameworks and their national implementations, contain highly specialized vocabulary that must be interpreted accurately. However, the density, length, and the multilingual nature of these documents pose significant challenges for the manual terminology extraction and maintenance.

Although technical standards are highly structured documents that include validated terminology approved by expert committees, their direct usability for terminology management remains limited in practice. In particular, in languages where comprehensive, publicly available national termbases have not been developed, such as Polish, standards often provide one of the most reliable and systematically curated sources of domain-specific terminology. However, the terminology they contain is typically embedded in document structures and not readily available in formats suitable for reuse in professional terminology management systems. This creates a need for methods that can support the transformation of document-bound terminology into structured, reusable resources. In this context, the present study explores the potential of large language models (LLMs) to facilitate this process. Rather than replacing expert-approved terminology, the aim is to investigate whether LLMs can assist in the operationalization of existing terminological knowledge by extracting and structuring it into termbase-ready formats.

[*] *5th International Conference on "Multilingual digital terminology today. Design, representation formats and management systems" (MDTT) 2026, June 25-26, 2026, Zadar, Croatia.*

[1*] Corresponding author.
[†] These authors contributed equally.
ewelina.kwiatek@uken.krakow.pl (E. Kwiatek); malgorzata.kodura@uken.krakow.pl (M. Kodura)
0000-0002-3383-0217 (E. Kwiatek); 0000-0002-7307-409X (M. Kodura)

Despite recent breakthroughs in large language models (LLMs) and their capabilities, the application of AI-driven terminology extraction from regulatory and standardization documents remains relatively unexplored. Existing research has focused primarily on general domain or specific linguistic challenges, leaving a gap in the understanding of how effectively AI tools perform in highly specialized, structured and regulated technical domains.

The aim of this paper is to evaluate the effectiveness of AI-based methods in extracting and aligning terminology from Polish and English versions of technical standards on definitions of screed material and floor screeds, at the same time identifying the strengths and limitations of these tools in technical, multilingual contexts. This study investigates whether AI models can reliably support terminology management processes and enhance technical documentation workflows from the perspective of translators and terminologists. The research advances current discussion on integrating artificial intelligence in industrial documentation, regulatory compliance and technical translation.

2. Literature review

Research on automatic terminology extraction (ATE) has evolved substantially over the past decades, moving from early linguistic and statistical approaches toward hybrid and AI-driven methods. Linguistic approaches relying on grammar-based rules and part-of-speech patterns [1] offer precision in structured texts but require significant expert input, especially for morphologically rich languages like Polish, as demonstrated by the TermoPL tool [2]. Statistical approaches, such as TF-IDF, PMI, Log-Likelihood, C-value and NC value, have been shown to perform quite effectively across various domains by identifying terms through frequency patterns and co-occurrence statistics [3, 4]. However, as demonstrated in comparative analyses [5], purely statistical approaches often struggle to capture the conceptual meaning of terms and either over-generate or miss important low-frequency domain terms, limiting their usefulness for specialized, terminology-dense documents. More reliable term lists across various domains could be obtained using hybrid methods combining linguistic filters and statistical ranking [6].

More recent research focuses on machine learning and transformer-based sequence labeling models, which achieve strong performance in identifying domain-specific terms across languages [7]. However, these models require annotated training data, and their performance varies across languages due to training imbalances. Studies referring to multilingual extraction tools [5, 6] confirm that bilingual term alignment remains limited in practice.

Large language models (LLMs) represent the newest development in ATE. Experiments conducted by Giguere and Iankovskaia [8] show that GPT-4 outperforms traditional statistical extractors by reducing false positives and capturing low-frequency domain terms more effectively. However, prompting-based evaluations [9, 10] reveal that LLMs show high recall but inconsistent precision, and can hallucinate terms, making human validation essential. Conceptual analyses [11, 12, 9] argue that LLMs lack stable concept systems, reinforcing the continued need for terminologists to control meaning and ensure coherence, particularly in multilingual and technical domains.

The reviewed studies support a hybrid methodology that combines rule-based extraction, statistical ranking, transformer-based processing and LLM-assisted candidate generation, anchored by expert validation to ensure terminological accuracy in specialized, multilingual settings [12]. This study seeks to examine whether, at the current stage of LLM development, such a hybrid workflow remains necessary or whether an advanced AI model can, on its own, provide quickly and efficiently reliable terminology extraction in the form of records used as input for a terminological base to be immediately used and updated without any additional manual operations. This solution would enable faster and more efficient terminology-building activities in a multilingual environment based on international standards, which is especially important for languages where a national terminological base is not available, as in the case of Poland, where, apart from limited internal resources, no comprehensive state-maintained terminological database currently exists.

3. Standardization

Translators are influenced by standardization in their work just like in everyday life. Standards ensure that common systems and products, such as electricity, transport, communication networks, and household appliances work safely, reliably, and consistently. In the translator's profession, standards play a similar role. They guide the translation process and ensure interoperable software and clear quality requirements. Standards define the "technical state of the art" and are authoritative, though subordinate to legal regulations.

However, it is important to note that there are different degrees of authoritativeness of the standards. International standards, which are issued by global bodies, e.g. International Organization for Standardization (ISO) are highly authoritative and widely accepted. Regional standards are applied and authoritative within an area. In the European Union, they are developed by such bodies as European Committee for Standardization (CEN), which deals with non-electrotechnical standards, European Committee for Electrotechnical Standardization (CENELEC), which works on electrical and electronic standards, and European Telecommunications Standards Institute (ETSI), which develops telecommunications, broadcasting, and ICT (information and communication technology) standards. National standards are valid within a single country, whereas industry or company standards created by industry or organizations are authoritative only within specific sectors or companies.

As far as the content of the standards is concerned, two types of standards can be distinguished: basic standards (or horizontal standards) and subject standards. Basic standards include methodology standards, such as those for information processing, information and documentation, and —most importantly— terminological principles and methods, which fall under the scope of ISO/TC 37. Subject standards cover a wide range, including product standards, testing standards, process standards, service standards, interface standards, and data standards.

Terminology standards that include standardized terminology form another separate category, and they can be seen as basic standards for a specific subject area (or scope) within a technical committee [13].

On the one hand, translators should not over-estimate the authoritative nature of standards (including standardized terminology) and should be aware of different degree of binding power depending on the type of the standards. Warburton and Martin [14] demonstrate that ISO standards often contain terminological inconsistencies and missing definitions, complicating automated extraction. On the other hand, standards are useful as parallel texts, as terminological standards provide terminology to be used in the target text of a technical translation.

The potential of using standards for creating a terminological database was recognised by the International Organization for Standardization, which in 2009 established a concept database (ISO/CDB). The database allows to search the content of ISO standards. It operates as an on-line browsing platform (https://www.iso.org/obp/ui). The access to the ISO/CDB content is in a read-only mode. It is open to all and free of charge. Users can either register for accounts or login as guests (without username and password).

ISO/CDB covers three important categories of concepts: terms and definitions, graphical symbols, and codes (country, currency, language and script). Users can either look for a specific term or symbol or they can browse standards, collections and publications by putting the search word that should appear in them or in their titles. It offers data in English, French, Spanish and Russian [15].

The organizations responsible for standardization in Poland are the Polish Military Centre for Standardization, Quality and Codification, which deals with the standardization of military terminology and has developed the MILTerm termbase; the Commission on Standardization of Geographical Names Outside the Republic of Poland, which issues guidelines on the translation of geographical names; and the Technical Committee 256 (TC 256) on Language, Translation and Terminology, which is the Polish equivalent of ISO/TC 37. TC 256 participates in the elaboration of methods, standards, and guidelines for terminology, but there is no evidence in the literature that it develops any widely available terminological resources itself.

4. Materials

The text sample used for the evaluation of effectiveness of various LLMs for the generation of terminological database is the technical standard *EN 13318: 2000 Screed material and floor screeds – Definitions* and its Polish version *PN-EN 13318:2002P Podkłady podłogowe oraz materiały do ich wykonania – Terminologia.*

EN 13318:2000 includes 100 terms presented in a table. Terms are divided into categories and numbered within those categories. Three columns include terms in three languages: English, German and French, as demonstrated in Figure 1 below.

2.12	**Lightweight screed** Screed where the hardened dry density after 28 days is less than 1400 kg/m^3.	**Leichtestrich** Estrich mit einer Trockenrohdichte nach 28 Tagen von unter 1400 kg/m^3.	**Chape légère** Chape dont la masse volumique apparente, après 28 jours de séchage, est inférieure à 1400 kg/m^3.
2.13	**Flooring** Uppermost layer of a floor that is designed to provide a wearing surface.	**Bodenbelag** Oberste Schicht eines Bodens, die als Nutzschicht dient.	**Revêtement de sol** Couche supérieure d'un sol, utilisée comme couche d'usure et de finition.
2.14	**Synthetic resin** A reactive organic polymer binder for a flooring system comprising one or more components which react at ambient temperature	**Synthetisches Reaktionsharz** Ein reaktives organisches Polymerbindemittel aus einer oder mehreren Komponente(n), die bei üblicher Umgebungstemperatur reagieren	**Résine synthétique** Liant à base de polymères organiques réactifs, à un ou plusieurs composants, qui réagit à la température ambiante.

Figure 1: Presentation of terms in EN 13318: 2000

PN-EN 13318:2002P includes 100 terms and definitions. Terms are provided in Polish, English, French and German. The definitions are available only in Polish, as demonstrated in Figure 2 below.

2.12 podkład lekki

Podkład o gęstości po 28 dniach twardnienia mniejszej niż 1400 kg/m^3.

lightweight screed
chape légère
Leichtestrich

2.13 posadzka

Wierzchnia użytkowa warstwa podłogi.

flooring
revêtement de sol
Bodenbelag

2.14 żywica syntetyczna

Jednoskładnikowe lub wieloskładnikowe spoiwo na bazie reaktywnych polimerów organicznych, reagujących w temperaturze otoczenia.

synthetic resin
résine synthétique
Synthetisches Reactionsharz

Figure 2: Presentation of terminology in PN-EN 13318: 2002P

5. Prompting

The experiment involved the application of large language models for automated terminology extraction and glossary compilation from technical standards prepared in two language versions. Three AI tools were tested: ChatGPT v5.2, Google Gemini 3 Pro and Claude Sonnet 4.5, all operating in the thinking mode. To ensure data confidentiality, the option to share user data for model improvement was disabled in all tools.

The input material consisted of original Polish and English technical standards in PDF format, uploaded directly to each system. Identical documents were used across all tools to ensure comparability.

For each AI tool, two prompting conditions were applied. The first used a user-designed prompt that clearly specified the required output format, including table headings to facilitate direct import into an existing term management system, e.g. in a CAT tool. The second conditions involved the use of prompts generated by the AI systems themselves. However, preliminary testing revealed that only the prompt generated by Claude Sonnet reliably produced outputs conforming to the desired structure of the glossary. Prompts suggested by ChatGPT and Google Gemini did not achieve comparable results and were therefore excluded from the final experimental setup. Consequently, the prompt generated by Claude Sonnet, which described the extraction and formatting process in a greater procedural detail while targeting the same glossary structure, was used as the reference prompt across the experiment.

Each model produced two glossary outputs based on the same input data. These outputs were subsequently analyzed with regard to structural conformity, completeness, and suitability for direct use as terminological records without further manual processing.

6. Results

The results revealed notable differences in how the three AI tools processed the input documents and generated the glossary. ChatGPT was unable to extract content from the technical standard in the Polish language and treated it as non-readable input, suggesting the use of external online resources instead. As a consequence, the glossary generated in response to both prompts was based exclusively on English source terms, with the Polish terms and definitions produced as translation of the English entries rather than extracted directly from the Polish standard. Moreover, ChatGPT initially compiled only a sample glossary containing six terms, while a complete glossary of 100 terms was produced only after additional prompting and extended processing time. The glossary was generated both as a text file with separators and as an xlsx file.

Gemini successfully extracted terminology from both the Polish and English PDF documents. In both prompting conditions, the tool produced a correctly structured glossary; however, the output was provided in the form of a code to be copied to a notebook environment and saved as a CSV file, to be consequently opened in Excel. While the column structure and formatting were correct, this intermediate step reduced the immediacy of the output for direct use as a termbase.

Claude was able to read and process both PDF documents and generate the glossary in the expected tabular format. When using the user-designed prompt, the tool initially failed to respond and required multiple retries before producing output. In contrast, when executing the prompt generated by the tool itself, Claude delivered a complete and accurate glossary on the first attempt. An additional advantage was the transparency of the process, as the tool generated, displayed and executed the Python code used to produce the glossary file, allowing the user to follow the extraction and compilation steps in real time. This feature was not intended to promote the direct use of LLM-generated Python scripts for terminology extraction, but rather to demonstrate that comparable results can be achieved without advanced programming knowledge. At the same time, the visibility of the generated code provides users with a basic insight into the underlying process, making the method more accessible and informative for those with little or no experience in Python.

7. Comparison of term extraction with LLMs and traditional NLP approaches

Since technical standards contain explicit term lists, terminology identification may seem largely trivial. So may be extraction of definitions which in standards are validated and approved by expert committees.

However, LLMs proved to be equally efficient for specialized term extraction when non-structured documents with non-established terms and non-explicit definitions are used. Moreover, it turned out to be more accurate than an established corpus-based tool (Sketch Engine).

To evaluate the practical utility of LLMs for specialised term extraction, a comparative experiment was conducted using a monolingual English language corpus of approximately 900 words, compiled from specialised web-based texts in the domain of floor underlays and related building materials.

Term extraction from the corpus was performed using three methods: semi-automatic term extraction with Sketch Engine, manual extraction by the human user (the output of this extraction serves as the reference baseline for evaluating system performance using precision and recall values) and extraction with three LLMs.

Term extraction with Sketch Engine yielded 553 candidate terms, comprising 350 single-word items and 203 multiword units. Following manual validation and the elimination of duplicates, 51 terms were retained for analysis.

Manual extraction conducted independently of computational tools produced comparable, yet marginally more comprehensive results, identifying 8 additional terms.

Sketch Engine achieved high recall (86.4%), which implies that it caught most of the real terms in the reference set. On the other hand, low precision (9.2%) means that it flagged many non-terms in the extraction process. The list of candidate terms delivered by Sketch Engine requires considerable manual filtering to remove irrelevant candidates.

The LLM used in this part of the research were the same as the models used for term extraction from technical standards, described in the previous part of the paper: Claude Sonnet 4.6, Gemini 3 Flash and ChatGPT 5.3. Each model was prompted using an identical instruction set, derived from a structured terminology extraction prompt that had been iteratively refined — including a self-improvement step in which all three LLMs were asked to enhance and adapt the original, user prepared prompt. The prompt specified inclusion and exclusion criteria for term candidates, required base/lemma form normalisation, and requested output in a structured Excel glossary format with designated empty columns for target language equivalents, explicitly marked as pending human translation.

The three models produced notably different results. Claude Sonnet extracted 41 term candidates and adhered to the design principle of the prompt in relation to deferring translation to a human specialist: target language columns were visually flagged in the output spreadsheet but left blank, thereby avoiding unsupported equivalence decisions. In addition, Claude generated definitions for the extracted terms. However, these definitions were not uniform in origin. Some were based directly on contextual information available in the source corpus, whereas others were generated using the model's internal domain knowledge rather than explicit evidence from the text. This introduces and additional layer of uncertainty, as such knowledge-based definitions are more susceptible to hallucination and consequently, they require formal validation prior to professional use. What is important, when prompted to indicate the source of the definitions, the model demonstrated partial meta-awareness by explicitly indicating which definitions were context-derived and which relied on domain knowledge and required additional expert validation.

Gemini returned only nine terms, distributed across two separate tables, one per language, with target language equivalents already populated by the model itself. This behaviour introduces a risk of hallucinated translations and, compounded by the absence of a directly downloadable bilingual spreadsheet, requires additional workarounds that reduce practical usability for translators and terminologists. ChatGPT extracted 20 terms but similarly proposed target language translations autonomously, again presenting a hallucination risk and reducing the terminologist's control over the output.

These differences in model behaviour have direct implications for professional usability. A terminology extraction workflow is most reliable when the role of the LLM is limited to identifying and structuring source language term candidates, with all cross-linguistic decisions reserved for a qualified human specialist. On this criterion, the output of Claude Sonnet was the most fit for purpose, although its definition-generation component still requires critical review and validation before integration into professional terminological resources.

The term candidates extracted by Claude Sonnet showed substantial agreement with the results of manual extraction performed in Sketch Engine. 28 out of 41 terms extracted by Claude Sonnet overlapped with terms identified by both Sketch Engine and human annotation. However, only three items from the list of terms identified by Claude Sonnet were not validated by the human user as terms (*hygienic screed system*, *SA2.5* and *three-part system*).

The software demonstrates high precision (92.7%), indicating that nearly all extracted candidates are valid terms. However, recall remains moderate (47.5%), suggesting that the software fails to capture approximately half of the terms present in the reference baseline (manual extraction set). The 10 additional valid terms extracted by the software but absent from the gold standard may warrant review—they could represent valid terms overlooked during manual extraction.

Claude Sonnet is particularly efficient in identification of multi-word units, e.g. *vacuum assisted shotblaster* or *polymer modified cementitious screed* or in establishing the nominal forms of the terms, e.g. *chemical etching* (from *chemically etched* that appears in the corpus).

8. Conclusion

The result of the experiment demonstrated that AI tools, in particular Claude Sonnet, not only can perform definition and term extraction effectively, but also provide the extraction results in a form that does not require any additional processing, providing a comprehensive input for the creation of an effective termbase, to be used immediately in CAT tools and larger terminological databases. Therefore, the role of the terminologist in this case consists in determining what information should be included in the database (e.g., data categories), identifying the sources from which these data should be obtained, and subsequently ensuring that all entries are correctly populated. Additionally, the terminologist does not need extensive knowledge of prompt engineering or proficiency in the Python programming language, because the AI tool provided with adequate information about the expected input and output data is capable of generating an effective prompt and executing the task correctly without further technical intervention.

In contrast to Claude Sonnet, the outputs of the other two models diverged considerably from the reference set. Claude Sonnet demonstrated sensitivity to contextual term boundaries, as the model correctly identified multi-word terms whose components were separated by intervening words in the source text. This capacity for discontinuous term recognition is an advantage over manual extraction approaches that rely on surface-level string matching.

A relevant asymmetry was observed between the two methodological approaches with respect to corpus scope. When a website URL was supplied to Sketch Engine, the tool searched not only the target page for terms, but also linked subpages, expanding the corpus beyond the material explicitly provided. This resulted in a larger pool of term candidates compared to the LLM-based extraction, strictly bounded by the uploaded document. This distinction is methodologically significant: the broader coverage offered by Sketch Engine may capture a greater number of terms across a domain, whereas LLM-based extraction offers greater precision and reproducibility, relative to a defined, controlled corpus. The two approaches are therefore best seen as complementary, and the choice between them depends on the goal of the terminologist or the translator, namely exhaustive domain coverage versus precise, document-specific extraction.

While the limitations of large language models, such as the risk of hallucination or inconsistency, are well documented, their increasing accessibility and widespread use among translators make it necessary to examine their practical availability in real-world workflows. The findings of this study suggest that, when used in controlled conditions and combined with appropriate prompting strategies, LLMs can support the transformation of structured, document-bound terminology into reusable terminological resources. This is particularly relevant in contexts where no comprehensive national termbase exists and where terminology must be compiled from primary sources, such as technical standards. In such cases, the use of LLMs does not replace expert validation but rather accelerates the initial stages of terminology extraction and structuring. Consequently, exploring their application in terminology work can contribute to more efficient terminology-building processes, provided that their use remains guided by domain expertise and critical evaluation.

Large language models, particularly Claude Sonnet, have also demonstrated effectiveness in extracting terminology from uncontrolled environments such as websites. Although their recall is lower than that achieved by dedicated term-extraction tools like Sketch Engine, their high precision indicates that the vast majority of the candidate terms they identify are valid. In addition, LLMs such as Claude Sonnet can be prompted to generate definitions for extracted terms and to specify whether these definitions draw on contextual information from the source corpus or on the model's internal domain knowledge. While both types of definitions provide a valuable starting point, they nonetheless require verification by subject-matter experts, as LLMs may hallucinate and online content is not always reliable or error-free.

Declaration on Generative AI

The author(s) have employed Generative AI tools to conduct the experiment described in this paper.

References

[1] X. Wang, T. Ahonen, J. Nurmi, Applying CDMA technique to network-on-chip. IEEE transactions on very large scale integration (VLSI) systems 15.10 (2007), 1091-1100. doi:10.1109/TVLSI.2007.903914.

[2] K. Kageura, E. Marshman, Terminology extraction and management, in: M. O'Hagan (Ed.), The Routledge Handbook of Translation and Technology, Routledge, New York, 2020, pp. 61–77. doi:10.4324/9781315311258-4.

[3] M. Marciniak, A. Mykowiecka, P. Rychlik, TermoPL - a Flexible Tool for Terminology Extraction, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation, European Language Resources Association (ELRA), Portorož, Slovenia, 2016, pp. 2278–2284. URL: https://aclanthology.org/L16-1361/.

[4] A. Oliver, A system for terminology extraction and translation equivalent detection in real time. Efficient use of statistical machine translation phrase tables, Machine Translation 31 (2017) 147–161. doi:10.1007/s10590-017-9201-7.

[5] M. T. Pazienza, M. Pennacchiotti, F. M. Zanzotto, Terminology extraction: An analysis of linguistic and statistical approaches, in: S. Sirmakessis (Ed.), Knowledge Mining. Studies in Fuzziness and Soft Computing, vol 185., Springer, Berlin, Heidelberg, 2006, pp. 255–279. doi:10.1007/3-540-32394-5_20.

[6] H. Costa, A. Zaretskaya, G. Corpas Pastor, M. Seghiri, Nine Terminology Extraction Tools: Are they useful for translators?, Multilingual 27 (2016) 14–20.

[7] A. Rigouts Terryn, V. Hoste, E. Lefever, A gold standard for multilingual automatic term extraction from comparable corpora: Term structure and translation equivalents, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), European Language Resources Association (ELRA), Miyazaki, Japan, 2018, pp. 1803-1808.

[8] K. Xu, Y, Feng, Q. Li, Z. Dong, J. Wei, Survey on terminology extraction from texts, Journal of Big Data 12:29 (2025). doi:10.1186/s40537-025-01077-x.
[9] J. Giguere, A. Iankovskaia, Leveraging large language models to extract terminology, in: Proceedings of the First Workshop on NLP Tools and Resources for Translation and Interpreting Applications, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria, 2023, pp. 57–60.
[10] F, Massion, Terminology in the Age of AI: The Transformation of Terminology Theory and Practice, Journal of Translation Studies 4 (2024) 67–94. doi:10.3726/JTS012024.04.
[11] H. T. Hong Tran, C. E. González-Gallardo, J. Delaunay, A. Doucet, S. Pollak, Is Prompting What Term Extraction Needs?, in: E. Nöth, A. Horák, A., P. Sojka (Eds.), Text, Speech, and Dialogue. TSD 2024. Volume 15048 of Lecture Notes in Computer Science, Springer, Cham, 2024, pp. 17-29. doi:10.1007/978-3-031-70563-2_2.
[12] B. Heinisch, Large language models for terminology work: A question of the right prompt?, Journal for Language Technology and Computational Linguistics 38 (2025) 13–30. doi:10.21248/jlcl.38.2025.280.
[13] B. Heinisch, Terminologists as Stewards of Meaning in the Age of LLMs: A Digital Humanism Perspective, in: H. L. Cardoso, R. Sousa-Silva, M. Koponen, A. Pareja-Lora (Eds.), Language Understanding in the Human Machine Era. Proceedings of the 2nd LUHME Workshop, Bologna, 2025, pp. 49–56. doi:10.21747/978-989-9193-73-4/lan2.
[14] C. Galinski, R. Weissinger, Terminology standardization and translation standards, Magyar Terminológia 3 (2010) 8–20. doi:10.1556/MaTerm.3.2010.1.3.
[15] K. Warburton, L. Martin, ISO Standards - Are they terminologically coherent?, Paper presented at the First International Conference on Terminology, Languages, and Content Resources (LaRC), Seoul, South Korea (2011). URL: https://www.researchgate.net/publication/275173767_ISO_Standards_-_Are_they_terminologically_coherent.
[16] International Organization for Standardization, The ISO Concept Database ISO/CDB released New tool for standards development and use of standards. Accreditation and Quality Assurance 14 (2009). doi:10.1007/s00769-009-0615-7.