Pretranslating SKOS Thesauri for Better Matching Performance

Felix Kraus[1,*], Nicolas Blumenröhr[1], Danah Tonne[1] and Achim Streit[1]

[1]*Karlsruhe Institute of Technology, Karlsruhe, Germany*

Abstract

Ontology matching (OM) systems typically perform poorly on non-English SKOS thesauri and terms, which are widespread in domains such as the Digital Humanities. In this work, we propose adding a dedicated English pretranslation step before matching, allowing existing OM systems to process multilingual thesauri more effectively without modification. To evaluate this approach, we apply WOKIE, a SKOS-specific translation pipeline introduced in prior work, as a preprocessing tool for translating thesaurus labels. Using real-world datasets from recent OAEI tracks, we assess the impact of pretranslation on multiple OM systems. Among 72 direct comparisons, 53 (73.6%) showed improved F1-scores after translation, with a mean increase of 0.275 ± 0.21. Our approach can easily be integrated into existing workflows and supports more effective reuse of multilingual thesauri.

Keywords

Ontology Matching, Thesauri, Multilingual, Translation, Digital Humanities, OAEI

1. Introduction

Knowledge organization systems have long been integral to the Digital Humanities (DH) domain, particularly in libraries and archival contexts. Recently, these systems have evolved significantly with the rise of interconnected and decentralized knowledge graphs [1]. Controlled vocabularies and thesauri[1] are essential tools for organizing and structuring knowledge across all domains. They are frequently integrated into research-oriented software applications, for example in tasks such as querying linked data or annotating datasets. Among the available models, the Simple Knowledge Organization System (SKOS) data model has become the dominant model for representing thesauri because it provides a clear structure of terms and their hierarchical relationships [2].

Since research topics are frequently addressed across different communities or institutions, overlaps between thesauri are common [3]. In addition, platform-specific requirements often result in slightly adapted terminologies, further contributing to knowledge fragmentation [4]. As a result, various strategies have been explored to address the challenges of heterogeneity across thesauri.

One solution is to integrate multiple knowledge graphs in real time into a unified search interface [4]. This approach offers advantages such as simplified management and the ability to selectively include or exclude resources. However, it may be impractical for metadata management or analytical purposes due to possible information conflicts between sources. Another strategy aims to link the terms of project-specific thesauri to one or more reference thesauri [5]. While this approach helps reduce heterogeneity, it becomes unsuitable for highly domain-specific terms unless those terms are covered by the reference thesauri. A single, shared thesaurus can be the most suitable option when moving away from the aim of designing a universal thesaurus for all of DH, which is neither realistic

OM-2025: The 20th International Workshop on Ontology Matching collocated with the 24th International Semantic Web Conference (ISWC 2025), November 2nd, Nara, Japan.

*Corresponding author.

felix.kraus@kit.edu (F. Kraus)

0000-0002-2102-4170 (F. Kraus); 0009-0007-0235-4995 (N. Blumenröhr); 0000-0001-6296-7282 (D. Tonne); 0000-0002-5065-469X (A. Streit)

© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[1]A thesaurus arranges terms in a hierarchical structure, adding term properties like synonyms and related terms. Their purpose is usually the provision of standardized terms for data annotation or analysis.

nor efficient [5]. Still, even when building project-specific thesauri, the challenge remains how to reuse terms from other resources and how to handle multiple languages.

Ontology Matching (OM) aims to tackle this challenge by identifying and aligning equivalent terms or concepts[2], and is therefore well-suited for integrating SKOS thesauri across projects and languages. However, the alignment of multilingual thesauri introduces additional complexity. Current OM systems are typically optimized for English-language input and tend to perform poorly on multilingual thesauri or non-English SKOS terms. This limitation becomes especially evident in DH contexts [7], where multilingual contexts, specialized terminology, historical languages, and diverse scripts are prevalent. A promising strategy to address this limitation is to translate non-English thesauri into English before alignment. While classical or neural Machine Translation (MT) methods face significant challenges in DH contexts, Large Language Models (LLMs) offer a compelling alternative for translation (see Section 2.2 for discussion).

In this work, we propose to use the translation pipeline WOKIE (**W**ell-translated **O**ptions for **K**nowledge Management in **I**nternational **E**nvironments)[3] as a preprocessing step. This pipeline enables high-quality translations that balance throughput, cost and translation quality, allowing existing OM systems to benefit immediately without requiring any modifications.

In this work, we present the following key contributions:

- We introduce an automated translation preprocessing step for OM.
- We adapt two DH benchmarks [4] from recent Ontology Alignment Evaluation Initiative (OAEI) campaigns to evaluate WOKIE as preprocessing step for OM.
- We perform a comprehensive analysis to demonstrate that English pretranslation of SKOS thesauri leads to measurable improvements across multiple OM systems.

2. Related Work

2.1. SKOS Thesauri in the Digital Humanities

Table 1 lists monolingual and multilingual thesauri where OM results could benefit from English translation. These resources are not part of our evaluation due to the lack of multilingual expert-generated reference alignments, but they illustrate typical real-world scenarios for which our method is relevant.

Table 1
Examples of groups of monolingual and multilingual thesauri.

Domain	Thesaurus	Main language
Archaeology	iDAI.world Thesaurus[5]	German
	PACTOLS[6]	Multilingual
Geography	Acteurs internationaux[7]	French
	China Biographical Database Project (CBDB)[8]	Chinese
	iDAI.world Thesaurus	German
	LNB ģeogrāfisko nosaukumu ontoloģija[9]	Latvian
	travel!digital Thesaurus[10]	German
Library	iDAI.world Thesaurus	German
	Systematik Basler Bibliographie[11]	German
Plants	AGROVOC[12]	Multilingual
	Kassu - Kasvien suomenkieliset nimet[13]	Finnish

[2]Although SKOS defines concept as "an idea or notion; a unit of thought" [6], we use "term" instead within this work to avoid ambiguity.
[3]https://github.com/FelixFrizzy/WOKIE
[4]https://github.com/FelixFrizzy/DH-benchmark, https://github.com/FelixFrizzy/DH-benchmark-multiling

2.2. Translation Systems

Manual translation of thesauri is usually impractical due to limited resources or a lack of available expert knowledge. Early efforts to support automated translation of thesaurus labels combined web services with databases of multilingual data [8]. These solutions were semi-automatic and did not scale well. More general approaches include statistical and neural MT. These techniques typically require well-prepared datasets, such as bilingual corpora for supervised learning or large monolingual, discipline-specific corpora for unsupervised learning. In the DH context, the scarcity of such resources [9] prevents successful application of these methods, especially on domain mismatch of training and application data [10]. Recent advances in LLMs have opened up new possibilities for MT [11]. While they require no fine-tuning or training on the user's side, they come with challenges such as longer processing times, the risk of hallucinated outputs, and non-deterministic results.

2.3. Limitations of Current Ontology Matching Systems

Although SKOS is widely used, we have previously shown that SKOS thesauri and the DH are not well represented in existing OAEI tracks [7]. Using our DH track, only six out of 17 evaluated matching systems produced meaningful results when applied with SKOS data. This limited compatibility poses a barrier for DH projects [12]. Furthermore, in our multilingual Archaeology track, most OM systems performed poorly when tested on monolingual data [12].

Looking at the systems listed above, we did not find information about language translation in the corresponding publications. In cases where source code was available, we investigated further. LogMap includes implementations for both Google Translate and Microsoft Translator, AgreementMakerLight (AML) supports only the latter. For the remaining systems, no source code was available.

3. Translation Pipeline Overview

This section outlines the core components of WOKIE. A full technical description of the software and its configuration options can be found in [13][14].

WOKIE translates SKOS thesauri using a pipeline that integrates primary translation services, confidence calculators, and LLM-based refinement, as illustrated in Figure 1. For each term and selected property (default is `skos:prefLabel`), the system retrieves translation candidates from one or more user-defined services. Each label is processed independently, and multiple services can be used to meet a minimum number of required outputs.

To determine a confidence of the translation, WOKIE first applies a frequency-based method that selects the candidate most commonly returned across services. If this candidate reaches a user-defined confidence threshold, it is accepted without further processing. Otherwise, the system uses an LLM to generate a context-sensitive alternative. Context is derived from available term definitions, descriptive metadata, or a general thesaurus description set by the user. If the LLM's suggestion matches a primary candidate, this version is selected. If not, the LLM is used to evaluate all available candidates, including its own, and selects the most appropriate, again factoring in the context. This step ensures ambiguous terms are resolved while minimizing the risk of hallucinated translations. In case of invalid or unparseable LLM output, WOKIE defaults back to the frequency-based choice. Once a final translation is selected, it

[5] https://isl.ics.forth.gr/bbt-federated-thesaurus/DAI/en/?clang=de
[6] https://isl.ics.forth.gr/bbt-federated-thesaurus/PACTOLS/en/
[7] https://data.legilux.public.lu/vocabulaires/international-actor/fr/
[8] https://isl.ics.forth.gr/bbt-federated-thesaurus/CBDB/en/?clang=zh
[9] https://dati.lnb.lv/onto/nll_geo/lv/
[10] https://vocabs.acdh.oeaw.ac.at/traveldigital_thesaurus/en/index/D?clang=de
[11] https://skosmos.bartoc.org/700/en/?clang=de
[12] https://agrovoc.fao.org/browse/agrovoc/en/
[13] https://finto.fi/kassu/en/?clang=fi
[14] Accepted as full paper for the Sixth Conference on Computational Humanities Research (CHR2025), in press.

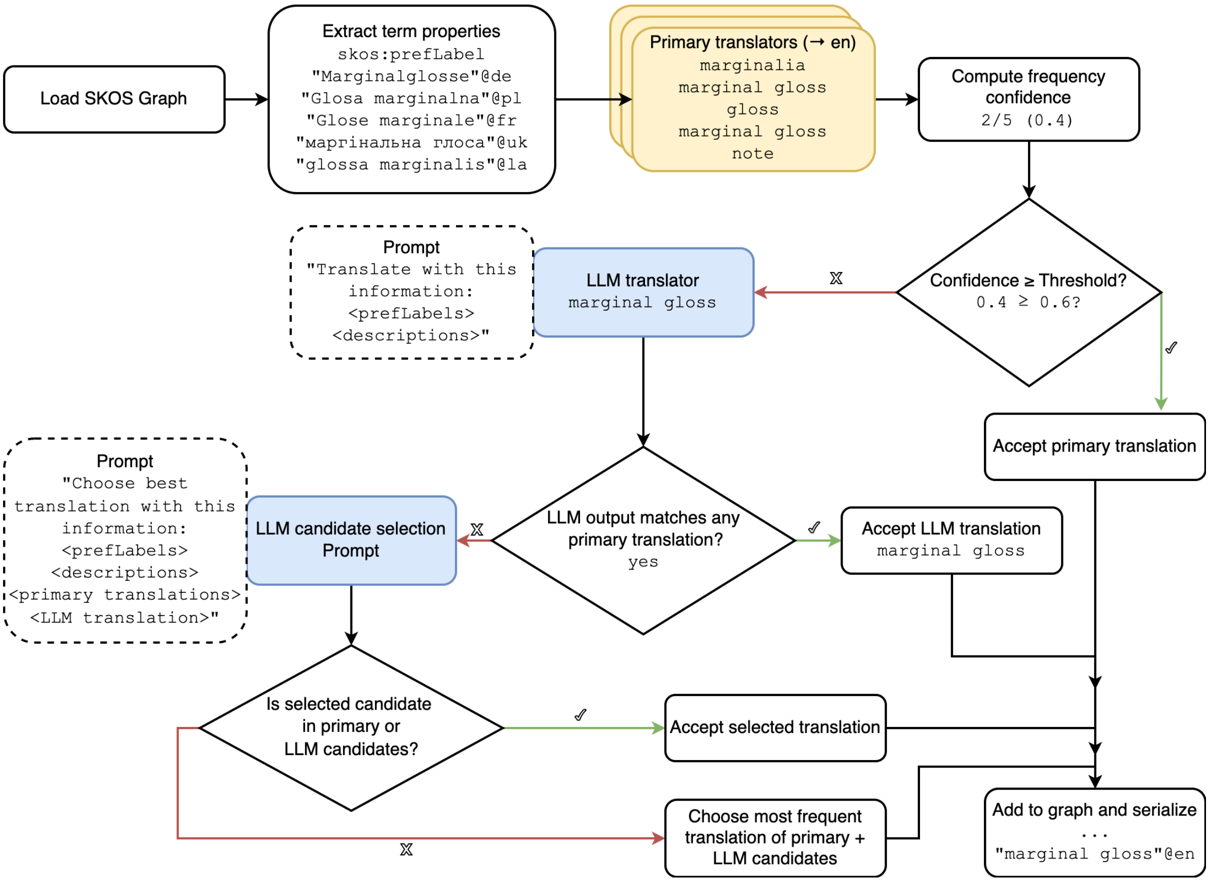

Figure 1: Simplified flowchart diagram of WOKIE's translation steps using "marginal gloss" as an example. The involvement of external translation services is marked in yellow, the use of LLMs blue. Positive decisions follow the right-hand path (green arrow), negative decisions left-hand path (red arrow).

is inserted into the SKOS graph as a new language-labelled literal. The resulting thesaurus remains a valid SKOS file, differing from the input only by the added translated labels.

WOKIE is designed to be easy to use: it requires no model training, data preparation, or in-depth knowledge of SKOS or ontology engineering. Its implementation as a Python tool allows it to run on standard hardware and to be integrated smoothly into existing OM workflows. In previous evaluations, the pipeline proved to be an advancement over simpler methods like using translations services only. Moreover, WOKIE demonstrated strong translation performance on DH terms [13], filling the gap of SKOS translation tools.

4. Approach

The evaluation is centred around the performance of OM systems on non-English thesauri compared to an additional pretranslation step.

4.1. Data Preparation

Base of the evaluation are thesauri used in the DH track (see Table 2), as there exist high-quality gold standard alignments. For each test case, we removed all English lexical content (identified by `xml:lang="en"`), except for the thesaurus description. The latter was kept to provide general context to the LLM in cases where a term lacks a definition or description. We then used WOKIE to translate all non-English labels into English. Eventually, we had two variants of each test case: one containing only the non-English labels, and one with added translations, see Table 3. In case a term had only English labels and would therefore be without label after removal of English content, it was excluded from the experiments and also removed from the reference alignment. For this reason, we did not directly compare the results with unaltered thesauri from [12] because the references do not match, which would introduce a bias. Additionally, thesauri that were entirely in English were left unchanged and used directly for matching.

Furthermore, we used test cases from the multilingual archaeology track. In this dataset, each test case contains only one language per thesaurus (e.g., French–Italian pairs), which allows for a targeted evaluation of monolingual thesauri matching, see Table 4.

Table 2
Thesauri used for the experiments.

Resource	Domain[15]	Version / Date	#concepts[16]	language (ISO 639)
DEFC Thesaurus[17]	Archaeology	-	~800	de, en, la
PACTOLS thesaurus for archaeology[18]	Archaeology	- / 2021-05-18	~60,000	ar, de, en, es, fr, it, nl
Iron-Age-Danube thesaurus[19]	Archaeology	1 / 2018-11-07	~6900	de, en, hr, hu, sl
iDAI.world Thesaurus[20]	Arch. / cult. hist.	1.2 / 2022-02-10	~290	de, en, es, fr, it
PARTHENOS Vocabularies[21]	Arch. / cult. hist.	- / 2019-05-07	~4200	en
OeAI Thesaurus - Cultural Time Periods[22]	Cultural history	1.0.0 / 2022-11-23	~400	de, en
DHA Taxonomy[23]	DH/CS	- / 2018-04-03	~120	en
UNESCO[24]	DH/CS	- / 2024-06-03	~4500	ar, en, fr, es, ru
TaDiRAH[25]	DH/CS	2.0.1 / 2021-07-22	~170	de, en, es, fr, it, pt, sr

Table 3
Properties of dataset based on DH track

Domain	Source	Target	#True Positives (TPs)
Archaeology	DEFC (632 terms[26], no English)	PACTOLS (70 terms, no English)	6
	DEFC (translated)	PACTOLS (translated)	6
	iDAI (2600 terms, no English)	PACTOLS (70 terms, no English)	18
	iDAI (translated)	PACTOLS (translated)	18
	Iron-Age-Danube (290 terms, no English)	PACTOLS (70 terms, no English)	6
	Iron-Age-Danube (translated)	PACTOLS (translated)	6
	PACTOLS (70 terms, no English)	PARTHENOS (800 terms, English only)	13
	PACTOLS (translated)	PARTHENOS (English only)	13
Cultural History	iDAI (270 terms, no English)	PARTHENOS (200 terms, English only)	53
	iDAI (translated)	PARTHENOS (English only)	53
	OeAI (400 terms, no English)	PARTHENOS (English only)	48
	OeAI (translated)	PARTHENOS (English only)	48
DH / CS	DHA (115 terms, no English)	UNESCO (490 terms, no English)	12
	DHA (translated)	UNESCO (translated)	12
	TaDiRAH (170 terms, no English)	UNESCO (490 terms, no English)	5
	TaDiRAH (translated)	UNESCO (translated)	5

4.2. Translation

For the translation step, we used an Apple M1 chip with 16GB RAM and the optimal WOKIE configuration parameter set as proposed in [13]:

- Primary translation services: Lingvanex, Google Translate, ModernMT, Microsoft Translator, Yandex, Argos, Reverso, PONS

[15]This is the field to which the thesaurus was grouped within the used dataset.
[16]Number of terms in the primary language before it was preprocessed to be included in the dataset.
[17]https://vocabs.dariah.eu/defc_thesaurus/en/
[18]https://isl.ics.forth.gr/bbt-federated-thesaurus/PACTOLS/en/
[19]https://vocabs.dariah.eu/iad_thesaurus/en/
[20]https://isl.ics.forth.gr/bbt-federated-thesaurus/DAI/en/
[21]https://vocabs.dariah.eu/parthenos_vocabularies/en/
[22]https://vocabs.acdh.oeaw.ac.at/oeai-cp/en/
[23]https://vocabs.dariah.eu/dha_taxonomy/en/
[24]https://vocabularies.unesco.org/browser/thesaurus/en/
[25]https://vocabs.dariah.eu/tadirah/en/
[26]The number of terms varies depending on the branch used for the respective domain.

Table 4
Properties of dataset based on arch multiling track

Source	Target	#True Positives (TPs)
iDAI (German only)	PACTOLS (English only)	17
iDAI (translated from German only)	PACTOLS (English only)	17
iDAI (English only)	PACTOLS (French only)	6
iDAI (English only)	PACTOLS (translated from French only)	6
iDAI (English only)	PACTOLS (Italian only)	6
iDAI (English only)	PACTOLS (translated from Italian only)	6
iDAI (French only)	PACTOLS (Italian only)	4
iDAI (French only)	PACTOLS (translated from Italian only)	4
iDAI (translated from French only)	PACTOLS (Italian only)	4
iDAI (translated from French only)	PACTOLS (translated from Italian only)	4

- LLM: Gemini 2.0 Flash (temperature 0)
- Threshold: 0.6
- Minimum number of translations: 5

4.3. Ontology Matching

OM was performed using the Matching EvaLuation Toolkit (MELT) framework[27] which is primarily used in OAEI campaigns. Based on our prior results [7, 12], we selected the following matching systems which support SKOS input: AML [14], ATMatcher [15], LogMap [16], LogMap Bio, LogMap KG and Matcha [17].

5. Evaluation Results

5.1. Overview

In total, we performed 205 matching tasks using six different OM systems, based on eight test cases. Besides specific tests, we performed 72 tasks where we compared the results of non-English thesauri with their fully translated variants. Out of these, 53 cases (73.6%) showed an increase in F1-score after translation, with a mean improvement of 0.275 ± 0.21. In 9 cases (12.5%), translation had no impact. The remaining 10 cases (13.9%) showed a decrease in F1-score, with a mean decline of -0.09 ± 0.09. An overview of the most and least impactful cases is provided in Table 5. All used thesauri and OM results are published as Zenodo record[28].

Table 5
Top positive and negative changes in F1-score after WOKIE translation.

Test Case Pair	**System**	**ΔF1-Score**	**Effect**	**Comment**
DEFC-PACTOLS	Matcha	+0.70	Positive	From 0.13 to 0.83 after translation
OeAI-PARTHENOS	LogMap KG	+0.63	Positive	From near-zero to 0.67
DEFC-PACTOLS	LogMap	−0.40	Negative	Increase in FP dominates, recall stable
Iron Age Dan.-PACTOLS	LogMap	−0.23	Negative	Increase in FP dominates, recall stable

5.2. DH Track

The results using the DH track are presented in Figure 2. In most cases, translation significantly improved the F1-scores across matching systems. For instance, in the OeAI–PARTHENOS test case,

[27] https://github.com/dwslab/melt
[28] https://doi.org/10.5281/zenodo.16607808

all systems except Matcha initially performed poorly. After translation, all but one system reached F1-scores between 0.61 and 0.67.

However, the improvements were not consistent across all test cases. For the ones involving PACTOLS, and for TaDiRAH-UNESCO, some systems showed a decrease in F1-score after translation. This decrease was primarily due to an increase in false positives (FPs) after translation, leading to a decreasing precision while recall remained stable. A representative example is the Iron Age Danube–PACTOLS test case with LogMap, shown in Figure 3. Here, the number of true positives (TPs) remained constant at 13 to 14 out of 17 correct alignments, but the number of FPs more than doubled when translating both thesauri. This led to a decline in precision and, consequently, in the F1-score. The other cases of decreasing F1-scores show a similar pattern. Examining the FPs in these cases more closely shows that the systems classify mostly related terms as matches such as *stone - stone circle* or *grave - gallery grave* which were not identified before translation. An increase of FPs consisting of solely unrelated term pairs is not introduced by the translation step.

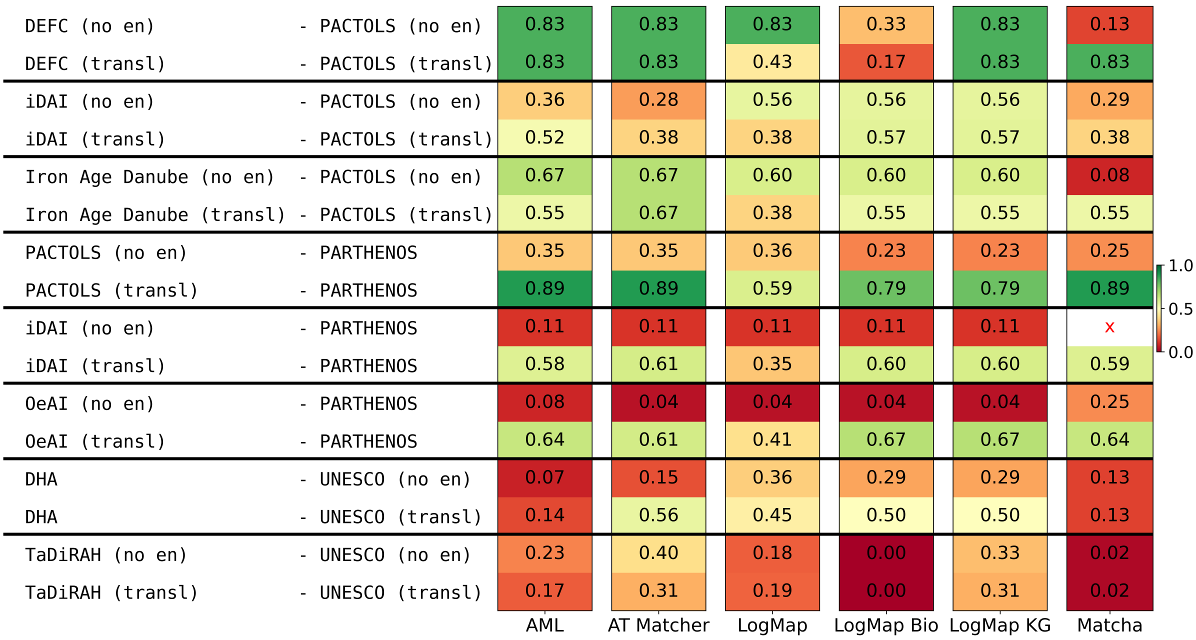

Figure 2: F1-scores across various test cases and matching systems. Each pair of rows represents related test cases: the upper row shows the original test case without English labels, while the lower row corresponds to its translated counterpart. The cell marked "x" indicates that the matching system failed on execution.

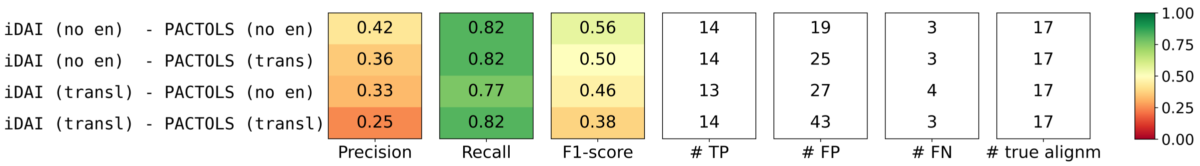

Figure 3: Additional metrics for the test cases with iDAI and PACTOLS using LogMap, including the number of true positives (TPs), false positives (FPs), false negatives (FNs) and total number of correct alignments.

5.3. Multilingual Archaeology Track

Figure 4 presents the results from the multilingual archaeology track, using monolingual thesauri with and without translation. Translation improved alignment performance across almost all systems and test cases. The only exception is AML, which failed on two test cases even after translation. Matcha showed particularly notable improvements: in two cases where it previously encountered code errors, a successful execution and increased F1-scores were observed after translation. For example, in the iDAI–PACTOLS (en–it) case, the F1-score rose to 0.50. Furthermore, all matching systems failed on the language combination French Italian. However, after translating both thesauri, systems achieved F1-scores between 0.09 and 0.22. The largest improvement occurred in the English French pair, where the top F1-score improved from 0.29 to 0.60.

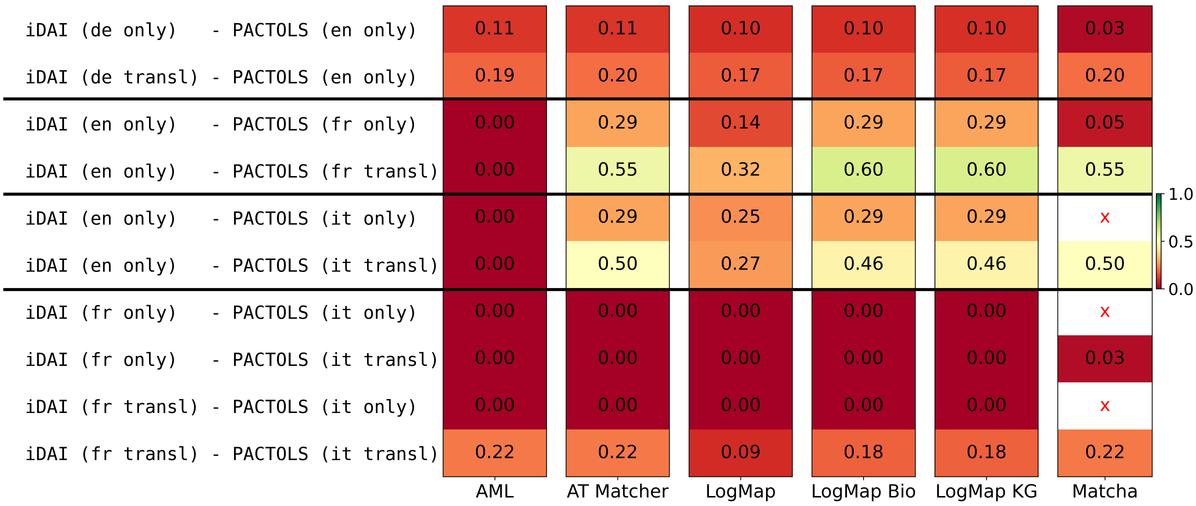

Figure 4: F1-scores for monolingual test cases. Within each group, the top row uses monolingual labels; the subsequent rows are WOKIE-translated equivalents. ”x” indicates matcher execution failure.

5.4. Confidence Levels

So far, we have treated each found alignment as being semantically identical, regardless of the confidence score reported by the matcher. Figure 5 shows the distribution of confidence scores for TP and FP matches before (left) and after (right) translation. In all systems, the proportion of TPs in the (0.9, 1.0] range increased after translation. A particular interesting case is AT Matcher, where all but one TP had a confidence above 0.9 after translation. This suggests that applying a confidence threshold could be a useful refinement to exclude alignments below that threshold, potentially reducing the number of FPs. We did not apply a global threshold, as absolute confidence values vary widely across systems. Moreover, confidence computation is typically undocumented and system-specific, making it difficult to define a meaningful threshold. This highlights a broader need for future work on standardizing confidence metrics in OM systems.

6. Discussion and Conclusion

The evaluation in this work involved thesauri covering 14 languages across two scripts, with detailed investigations covering a subset of these languages. While this does not include all existing languages, the goal was not to achieve full language coverage, but to assess the effect of translations on OM. Although WOKIE was used as the translation method, the results are not tied to this specific tool. Alternative approaches may yield even better results when adapted to SKOS and a specific use-case.

Looking at the F1-score after translation, they are on the lower end for certain test cases, which might be deemed insignificant. As explained in Section 4.1, a quantitative comparison to previously obtained results is not possible. Despite these constraints, the scores remain in line with previous results on similar DH datasets [12], where even top-performing systems achieved only moderate F1. In practical settings, even a small number of additional correct matches can improve subsequent tasks such as thesaurus merging or enrichment.

Another observation is the increase in FPs after translation in some cases. While this negatively affects precision, not all FPs are necessarily irrelevant. Some of them might represent semantically

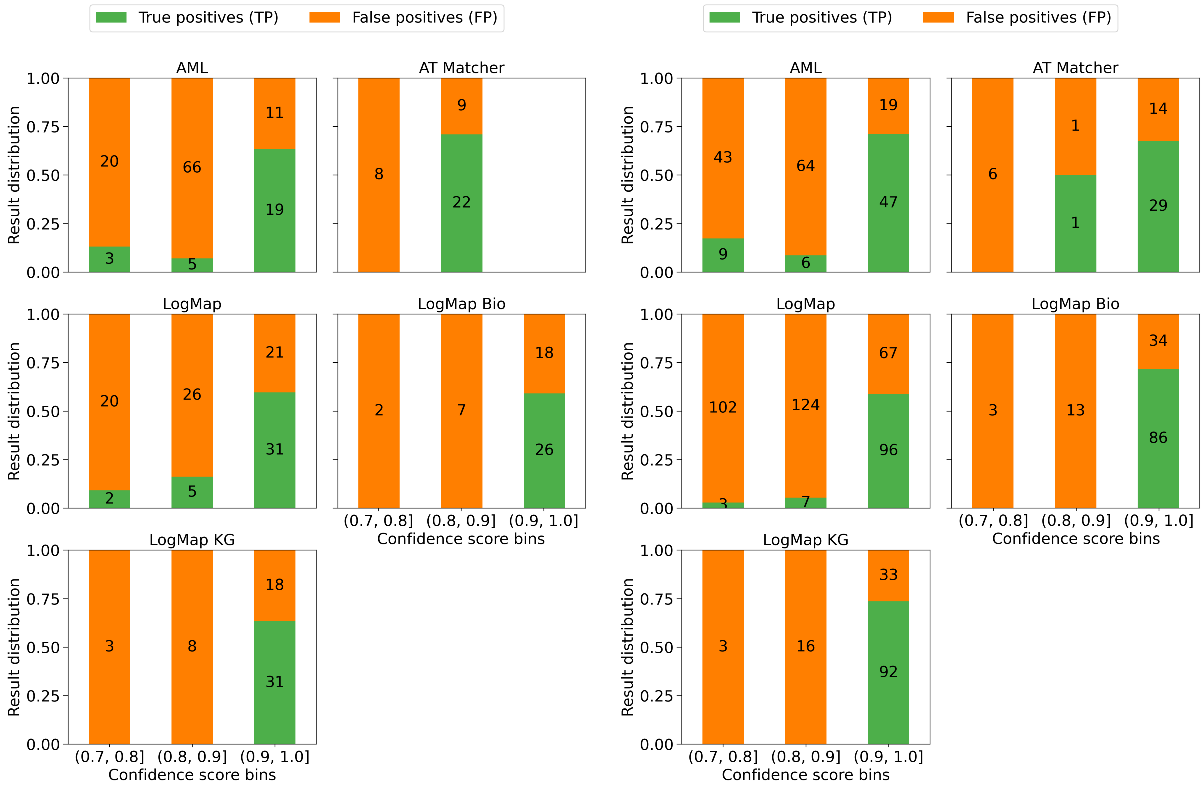

Figure 5: Distribution of matcher confidence scores for true positive (TP) and false positive (FP) alignments before (left) and after (right) translation, aggregated across all used test cases of the DH track.

related terms that fall outside the scope of strict reference alignments. In mapping or merging workflows, such matches might still provide value.

In conclusion, this work demonstrates that pretranslating non-English SKOS thesauri significantly improves the performance of existing OM systems. By decoupling translation from matching, we enable such results without requiring changes on the OM systems. Therefore, our approach is simple to integrate and effective, especially in domains where standard translation systems struggle. Future work will explore translating additional SKOS properties and adapting the pipeline for full ontologies.

Acknowledgments

This research was funded by the research program “Engineering Digital Futures” of the Helmholtz Association of German Research Centers, and the Helmholtz Metadata Collaboration Platform (HMC).

Declaration on Generative AI

Declaration following the taxonomy of ceur-ws.org/genai-tax.html: During the preparation of this work, the authors used ChatGPT and LanguageTool for the following: improve writing style, grammar and spelling check, image generation (code improvements only). After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication’s content.

References

[1] B. Haslhofer, A. Isaac, R. Simon, Knowledge Graphs in the Libraries and Digital Humanities Domain, in: S. Sakr, A. Zomaya (Eds.), Encyclopedia of Big Data Technologies, Springer International Publishing, Cham, 2018, pp. 1–8. doi:10.1007/978-3-319-63962-8_291-1.

[2] J. Pastor-Sánchez, F.-J. Martinez-Mendez, J. Rodríguez, Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives, Information Research, ISSN 1368-1613, Vol. 14, Nº. 4, 2009 14 (2009).
[3] J. Euzenat, P. Shvaiko, Ontology Matching, second edition ed., Springer, Berlin, Heidelberg, 2013. doi:10.1007/978-3-642-38721-0.
[4] A. Morvillo, M. Mecella, Integrating multiple knowledge graphs in Digital Humanities, in: ST4DM 2024: Semantic Technologies for Data Management, Twente, Italy, 2024.
[5] M. Piotrowski, G. Colavizza, F. Thiery, K.-C. Bruhn, The Labeling System: A New Approach to Overcome the Vocabulary Bottleneck, in: DH-CASE II: Collaborative Annotations on Shared Environments: Metadata, Tools and Techniques in the Digital Humanities - DH-CASE '14, ACM Press, Fort Collins, CA, USA, 2014, pp. 1–6. doi:10.1145/2657480.2657482.
[6] A. Miles, S. Bechhofer, SKOS Simple Knowledge Organization System Reference, W3C Recommendation, World Wide Web Consortium, United States, 2009.
[7] F. Kraus, N. Blumenröhr, G. Götzelmann, D. Tonne, A. Streit, A Gold Standard Benchmark Dataset for Digital Humanities, in: E. Jiménez-Ruiz, O. Hassanzadeh, C. Trojahn, S. Hertling, H. Li, P. Shvaiko, J. Euzenat (Eds.), Proceedings of the 19th International Workshop on Ontology Matching, volume 3897 of *CEUR Workshop Proceedings*, CEUR, Baltimore, MD, USA, 2024, pp. 1–17. doi:10.5445/IR/1000178023.
[8] M. Espinoza, A. Gómez-Pérez, E. Mena, LabelTranslator - A tool to automatically localize an ontology: 5th European Semantic Web Conference, ESWC 2008, The Semantic Web (2008) 792–796. doi:10.1007/978-3-540-68234-9_60.
[9] O. Suissa, A. Elmalech, M. Zhitomirsky-Geffet, Text analysis using deep neural networks in digital humanities and information science, Journal of the Association for Information Science and Technology 73 (2022) 268–287. doi:10.1002/asi.24544.
[10] Y. Kim, M. Graça, H. Ney, When and Why is Unsupervised Neural Machine Translation Useless?, in: A. Martins, H. Moniz, S. Fumega, B. Martins, F. Batista, L. Coheur, C. Parra, I. Trancoso, M. Turchi, A. Bisazza, J. Moorkens, A. Guerberof, M. Nurminen, L. Marg, M. L. Forcada (Eds.), Proceedings of the 22nd Annual Conference of the European Association for Machine Translation, European Association for Machine Translation, Lisboa, Portugal, 2020, pp. 35–44.
[11] W. Jiao, W. Wang, J.-t. Huang, X. Wang, S. Shi, Z. Tu, Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine, 2023. doi:10.48550/arXiv.2301.08745. arXiv:2301.08745.
[12] M. A. N. Pour, A. Algergawy, E. Blomqvist, P. Buche, J. Chen, P. G. Cotovio, A. Coulet, J. Cufi, H. Dong, D. Faria, L. Ferraz, S. Hertling, Y. He, I. Horrocks, L. Ibanescu, S. Jain, E. Jiménez-Ruiz, N. Karam, F. Kraus, P. Lambrix, H. Li, Y. Li, P. Monnin, H. Paulheim, C. Pesquita, A. Sharma, P. Shvaiko, M. Silva, G. Sousa, C. Trojahn, J. Vatasčinová, B. Yaman, O. Zamazal, L. Zhou, Results of the Ontology Alignment Evaluation Initiative 2024, in: Proceedings of the 19th International Workshop on Ontology Matching (OM 2024), Baltimore, USA, November 11, 2024. Ed.: E. Jiménez-Ruiz, O. Hassanzadeh, C. Trojahn, S. Hertling, H. Li, P. Shvaiko, J. Euzenat, volume 3897 of *CEUR Workshop Proceedings*, CEUR, Baltimore, MD, USA, 2024, pp. 64–97. doi:10.5445/IR/1000179469.
[13] F. Kraus, N. Blumenröhr, D. Tonne, A. Streit, Mind the Language Gap in Digital Humanities: LLM-Aided Translation of SKOS Thesauri, 2025. doi:10.48550/arXiv.2507.19537. arXiv:2507.19537.
[14] D. Faria, C. Pesquita, E. Santos, M. Palmonari, I. F. Cruz, F. M. Couto, The AgreementMakerLight Ontology Matching System, in: R. Meersman, H. Panetto, T. Dillon, J. Eder, Z. Bellahsene, N. Ritter, P. De Leenheer, D. Dou (Eds.), On the Move to Meaningful Internet Systems: OTM 2013 Conferences, Springer, Berlin, Heidelberg, 2013, pp. 527–541. doi:10.1007/978-3-642-41030-7_38.
[15] S. Hertling, H. Paulheim, ATBox results for OAEI 2020, in: P. Shvaiko, J. Euzenat, E. Jiménez-Ruiz, O. Hassanzadeh, C. Trojahn (Eds.), CEUR Workshop Proceedings, volume 2788, RWTH Aachen, Aachen, Germany, 2020, pp. 168–175.
[16] E. Jiménez-Ruiz, B. Cuenca Grau, LogMap: Logic-Based and Scalable Ontology Matching, in: L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. Noy, E. Blomqvist (Eds.), The Semantic Web – ISWC 2011, Springer, Berlin, Heidelberg, 2011, pp. 273–288. doi:10.1007/

978-3-642-25073-6_18.
[17] D. Faria, M. C. Silva, P. Cotovio, P. Eugénio, C. Pesquita, Matcha and Matcha-DL results for OAEI 2022, in: P. Shvaiko, J. Euzenat, E. Jiménez-Ruiz, O. Hassanzadeh, C. Trojahn (Eds.), Proceedings of the 17th International Workshop on Ontology Matching (OM 2022), volume 3324 of *CEUR Workshop Proceedings*, CEUR, Hangzhou, China, 2022, pp. 197–201.

Abbreviations

AML AgreementMakerLight

DH Digital Humanities

FP false positive

LLM Large Language Model

MELT Matching EvaLuation Toolkit

MT Machine Translation

OAEI Ontology Alignment Evaluation Initiative

OM Ontology Matching

SKOS Simple Knowledge Organization System

TP true positive