

Jackalope Plus Performance: Benchmarking and Competitors

Denys Kaduk, MD^{1,4}, Bohdan Khilchevskiy, MD^{1,2}, Maksym Trofymenko, MD¹, Tetiana Nesmiian MD, LLM^{1,3}, Polina Talapova, MD, PhD^{1,2}, Max Ved¹, Inna Ageeva¹

¹ IT company SciForce, Kharkiv, Ukraine

² Kharkiv National Medical University, Kharkiv, Ukraine

³ Kharkiv National Pedagogical University named after H. S. Skovoroda, Kharkiv, Ukraine

⁴ V. N. Karazin Kharkiv National University, Kharkiv, Ukraine

Introduction

In the healthcare domain, where data plays an increasingly crucial role, accurate and efficient mapping of medical terminologies is fundamental. This mapping ensures seamless data exchange (interoperability) between different healthcare information systems, facilitates clinical research efforts, and ultimately improves patient care. The advent of the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) has further emphasized the significance of terminology mapping.

While manual mapping remains an option, it is a laborious and time-consuming process, hindering large-scale data standardization initiatives. To address this challenge, several automated mapping tools have emerged, aiming to streamline the process and enhance efficiency^{1,2}. This study benchmarks Jackalope Plus, a prominent mapping tool, against manual mapping and Usagi, another automated option, to assess their relative strengths and weaknesses.

The following sections delve into the methodologies employed for the benchmarking process, the findings gleaned from the analysis, and the implications for selecting the optimal mapping tool in the context of healthcare data integration.

Methods

We decided to benchmark Jackalope Plus against manual mapping and Usagi mapping. For this purpose, we assembled three domain experts (testers) familiar with all mapping methods and assigned tasks based on their preferences. The benchmark utilized a newly created dataset comprising unmapped concepts from MeSH and CIEL, as well as some concepts created by our team expert. The test set was provided to the testers prior to the mapping process, formatted as a CSV file with a semicolon (;) delimiter containing 40 unmapped concepts. Each mapping method followed specific steps:

- *Manual*: The process begins by selecting a list of unmapped concepts, which are then mapped using Athena's search capabilities, constrained by the current limitations on SNOMED CT.
- *Usagi*: The pre-mapping phase involves downloading Usagi and the necessary vocabulary archives, updating CPT names, and creating an index within Usagi. For mapping, a list of unmapped codes is selected and uploaded as a comma-separated CSV file only. The tool filters based on the desired vocabulary or domain, after which mappings are reviewed and approved. Queries are created to add new concepts and mappings to the same set of tables as Athena. The time measurement excludes the steps involved in the installation process.
- *Jackalope Plus*: Users start by selecting unmapped codes and adding a column to create a file compatible with Jackalope. After logging into Jackalope, users upload the CSV file and select the appropriate delimiter. Subsequently, a mapping process was completed

Results

Results are presented in Table 1. The time for preparation is not included in the final results due to the variability of the source data.

Table 1. Results

Time in minutes	Method		
	Manual	Usagi	Jackalope
Time required for dataset preparation	4	5	1
Time for mapping process	104	35	27
Results			
Correct mapping	29	21	31
Ambiguous/not full mapping	5	7	6
Wrong mapping	6	11	3

The benchmarking process revealed significant insights into the efficiency and accuracy of the mapping tools compared. Both Jackalope Plus and Usagi demonstrated the ability to significantly reduce the time required for the mapping process compared to manual methods. This time efficiency is critical in large-scale data standardization efforts, where speed and accuracy are paramount.

However, the quality of the results varied between the tools. Jackalope Plus consistently produced more accurate mappings than Usagi. This superior accuracy can be attributed to Jackalope Plus's advanced algorithms and streamlined workflow, which minimize errors and enhance precision. The domain experts reported that Jackalope Plus not only reduced the time spent on mapping but also delivered higher-quality results, making it a more reliable tool for integrating medical data into standardized frameworks.

Conclusion

This benchmarking exercise has shed light on the relative merits of various medical terminology mapping methods: manual mapping, Usagi, and Jackalope Plus. While manual mapping offers the advantage of accuracy and organization, its significant time consumption renders it impractical for large-scale projects. Usagi, on the other hand, being an open-source tool, excels in filtering and indexing functionalities but necessitates additional post-processing steps and separate database integration.

Jackalope Plus emerges as a compelling choice, demonstrably reducing mapping time compared to both manual and Usagi methods. Moreover, it surpasses Usagi in accuracy due to its sophisticated algorithms and streamlined workflow. These factors combined make Jackalope Plus a reliable and efficient tool for integrating medical data into standardized frameworks.

Selecting the most suitable mapping method hinges on the specific requirements of the project. For smaller-scale endeavors where meticulous accuracy is paramount, manual mapping

might be considered. However, for large-scale data standardization initiatives that prioritize both speed and accuracy, Jackalope Plus stands out as the preferred option. Its ability to streamline the mapping process, minimize errors, and seamlessly integrate with databases makes it a valuable asset in the healthcare informatics landscape.

References

1. OHDSI. Usagi: A tool for automated mapping of coding systems. Available from: <https://github.com/OHDSI/Usagi>. Accessed 21 June 2024.
2. OHDSI. Jackalope Plus: Showcase 2023. Available from: https://www.ohdsi.org/2023showcase-24/?_ga=2.194360047.357006080.1718820387-675270364.1690810023&_gl=1*1xocnz0*_ga*Njc1MjcwMzY0LjE2OTA4MTAwMjM.*_ga_BHVF662WPC*MTcxODg4MjI2My45My4xLjE3MTg4ODI3NTMuMjM4wLjA.*_ga_PNYYSZVRVX*MTcxODg4MjI2My45My4xLjE3MTg4ODI3NTMuMjM4wLjA. Accessed 21 June 2024.