

Jackalope: A software tool for meaningful post-coordination for ETL purposes

Eduard Korchmar^{1,4}, Polina Talapova MD^{1,2}, Mariia Kolesnyk¹, Denys Kaduk MD^{1,3},
Inna Ageeva¹, Max Ved¹

¹ IT company SciForce, Kharkiv, Ukraine

² Kharkiv National Medical University, Kharkiv, Ukraine

³ V. N. Karazin Kharkiv National University, Kharkiv, Ukraine

⁴ Goldsmiths, University of London, London, United Kingdom

Background

Post-coordination using the SNOMED CT logic can be a feasible solution to represent some of the concepts that can not be directly mapped to existing Standard concepts in Standardized Vocabularies⁶. The goal of the first step of the toolset development should be to create a system that is usable and provides consistent results, while demanding not more than knowledge of post-coordination in SNOMED CT from the end user. The end result of the system should be the automated creation of local Standard concepts in independent CDM instances, with interoperability and sustainability in mind. This can be achieved by evaluating SNOMED Compositional Grammar expressions² and processing them against SNOMED CT ontological representation, generating local standard concepts inside the existing SNOMED Vocabulary hierarchy³.

Methods

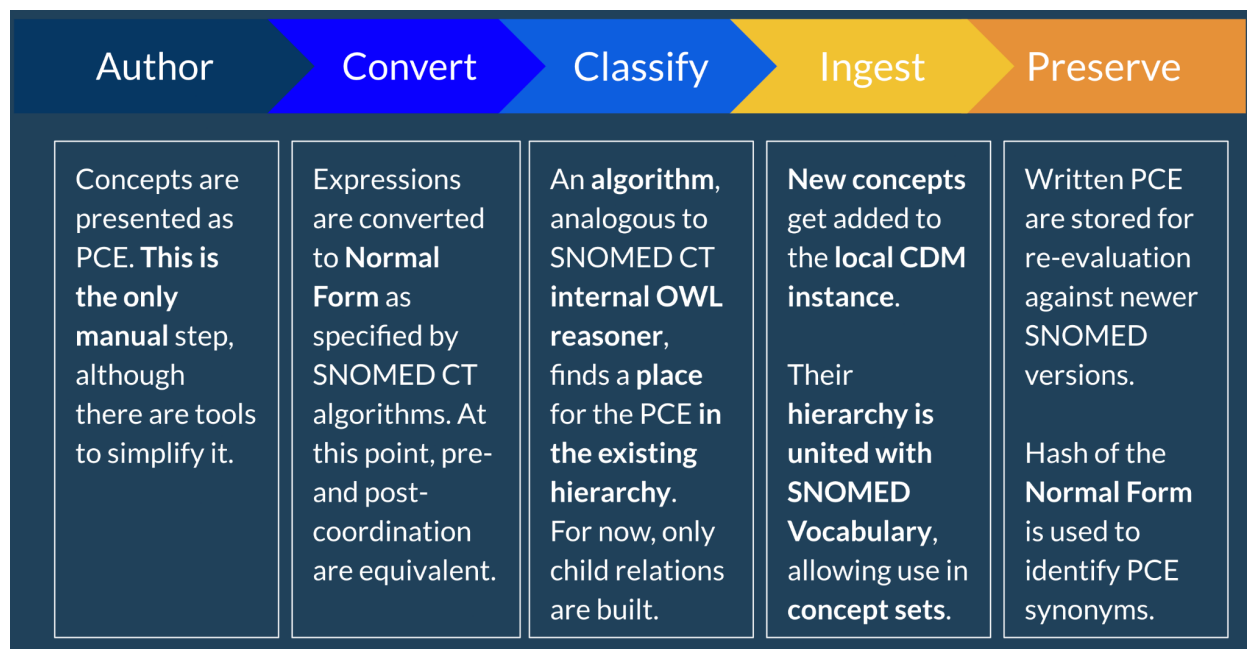


Chart 1: Implementation model

To prove the feasibility of our approach, we collected 8 examples of unmappable concepts from source datasets of OHDSI community members, as well as 7 examples of OMOP Extension concepts, which are

standard concepts created and maintained manually to respond to community demand, or imprecisely standardized concepts from ICD10 family of vocabularies. We prototyped a software tool^{1,4} to place the concepts, represented as SNOMED post-coordinated expressions, into the SNOMED hierarchy, introducing seamless changes into OMOP CDM Standardized Vocabularies on the local deployed instance.

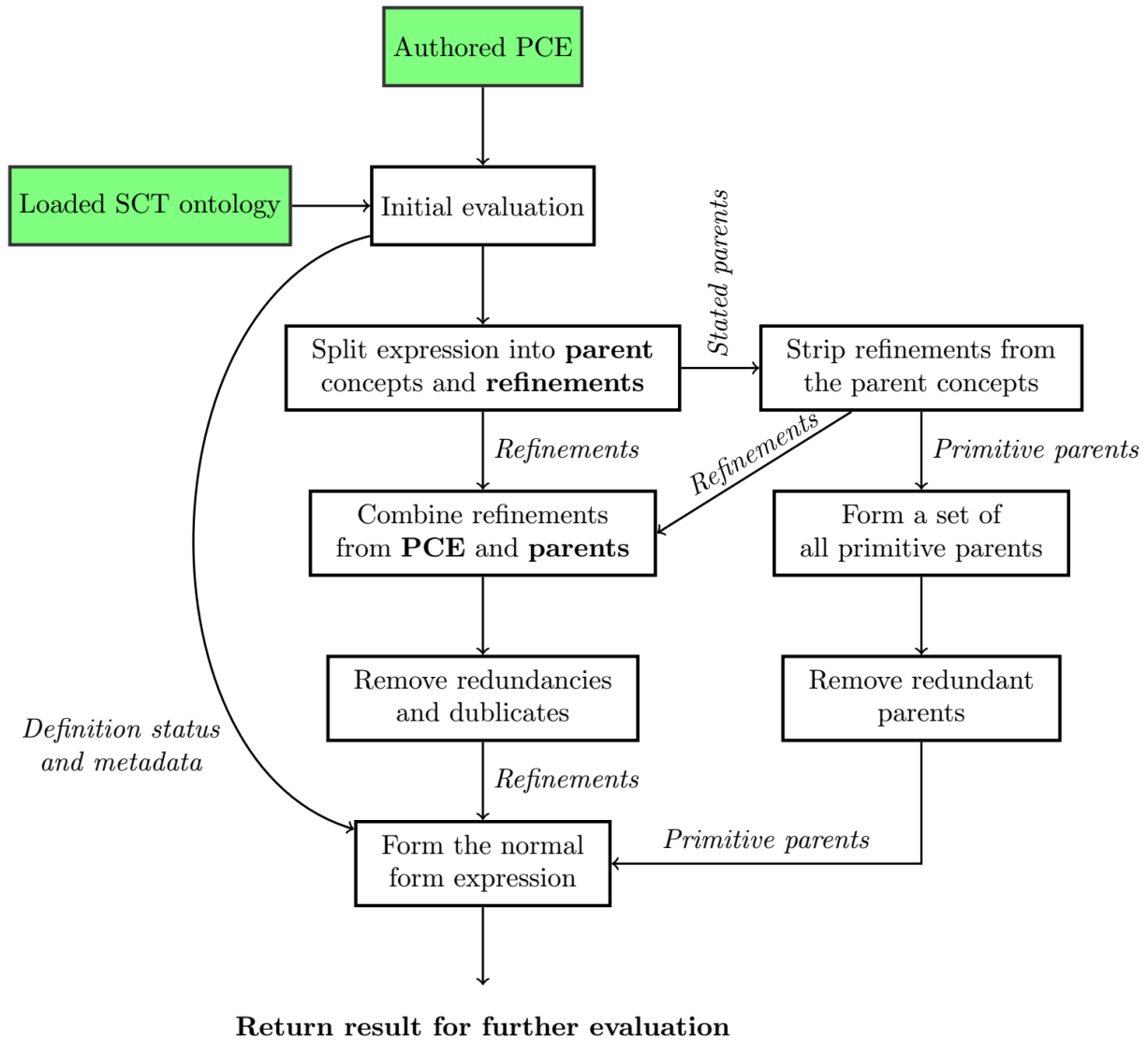


Chart 2: Extracting Normal Form from a post-coordinated expression

PCE evaluation was done through computation of Necessary Normal Form and evaluating it against existing concepts in the hierarchy^{7,8}. If an existing concept is isomorphic with a subgraph of expression's semantic representation, it is considered to be an ancestor of said expression.

Results

We have created a functional algorithm (see Chart 1.) that processes representations of clinical ideas in

the format of SNOMED CT post-coordinated expressions, and makes changes in the OMOP CDM database, hosted on the backend server. This compatibility makes it possible to use added synthetic concepts in local OMOP CDM instances, sustainably preserving the full clinical meaning of source vocabulary concepts. Compatibility is also preserved with other existing CDM instances as well future SNOMED CT and Standardized Vocabulary versions.

Please find the source expressions and results of their evaluation in the neighboring file in the Symposium directory.

Conclusion

SNOMED post-coordination is a viable option to preserve the full meaning of the source data in cases where using simple mapping would lead to an unacceptable loss of data. Implementing it as an optional step of the ETL process can improve standardization coverage, and it can also be used to better implement vocabulary ingestion into OMOP. Further work to develop a visual tool for building convention compliant SNOMED post-coordination expressions is required⁹.

References/Citations

1. [Jackalope source code](#)
2. [SNOMED Post-coordination/Compositional Grammar guidance](#)
3. [SNOMED Vocabulary ingestion process](#)
4. [Graph computation in Jackalope](#)
5. [SNOMED release format](#)
6. [Post-coordinated expression application](#)
7. [Normal form as a mean of semantic evaluation](#)
8. [Normal form theoretical basis for SNOMED RT – predecessor of the modern SNOMED](#)
9. [Practice of PCE authoring interface](#)