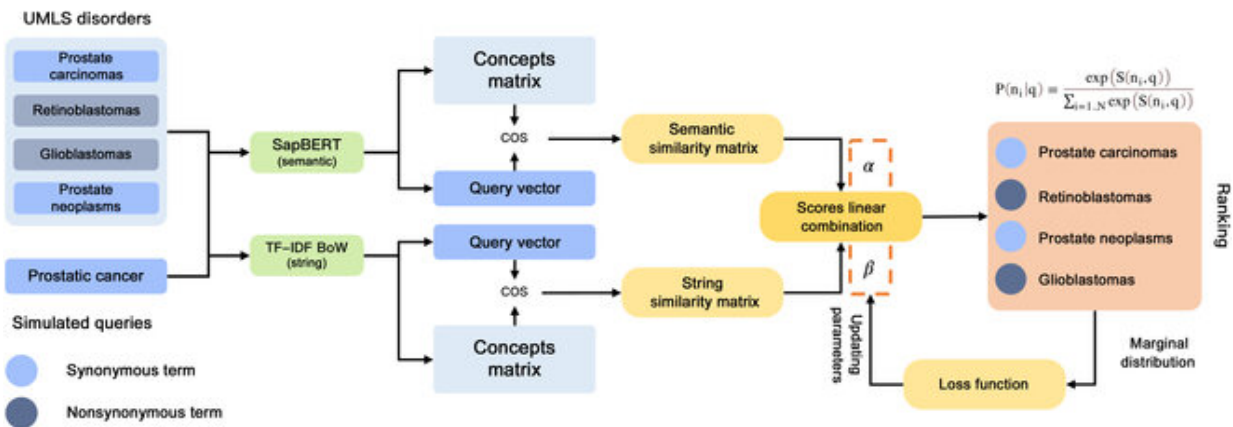


# Mapping Chinese medical entities to the Unified Medical Language System

May 8 2023



Overview of the training process for optimizing linear combination parameters. Credit: *Health Data Science* (2023). DOI: 10.34133/hds.0011

An effective strategy for mapping Chinese medical entities to the Unified Medical Language System (UMLS) has been described in a recent study. In addition, the study has established an evaluation dataset based on real-world data for linking Chinese medical entities to systems in other languages.

The results were published in *Health Data Science*.

UMLS is one of the well-developed medical terminology systems that make up the cornerstone of health informatics research and

technologies. However, the lack of a high-quality Chinese medical ontology makes it challenging to process medical texts and documents for computation.

"To overcome this challenge, we sought to identify the optimal approach for mapping Chinese medical entities to UMLS concepts in the current stage," said Taijiao Jiang, author and professor at Guangzhou Laboratory (GL), Guangzhou Medical University, China. "With this strategy in place, it is now feasible to normalize medical entities across [different languages](#)."

"The study investigated three mapping methods, namely a string-based, a semantic-based, and a string and semantic similarity-combined strategy," shared Lizong Deng, coauthor and associate professor at Institute of Systems Medicine (ISM), Chinese Academy of Medical Sciences & Peking Union Medical College, China. "In addition, cross-lingual pre-trained language models were applied."

"We leveraged character-level and semantic-level similarity using a linear combination method, with the help of multi-source translation and a pre-trained language model to map Chinese medical entities to UMLS, an important basis for a well-organized terminology system," explained Luming Chen, coauthor at GL. "This linear combination strategy demonstrated the best performance on the cross-lingual medical entity linking task despite inadequately developed Chinese medical terminology systems."

Further, the reported method can be applied to downstream tasks to automatically map Chinese medical terms to standard UMLS concepts, thereby facilitating fine-grained medical knowledge representation and other advanced intelligent medical applications.

Moreover, by tapping web-based translation engines, the described

approach can be used in mapping non-English entities to UMLS, thus extending UMLS's coverage. In return, the UMLS with multilingual components expedites using medical informatics tools in other languages.

"To further validate the results of the mapping strategies on a larger and more diverse dataset, we will establish a new cross-lingual dataset that covers a wider range of semantic types in medical terminology," said Yifan Qi, coauthor at ISM. "We will investigate the limitations of the mapping strategies, including cases where the translations are inaccurate or when there is a lack of English counterparts in the UMLS."

The main goal for future research is to develop a desirable [language](#) model that fully integrates medical concepts at the string and semantic levels.

Finally, the ultimate goal is to fine tune cross-lingual medical entity linking and integrate the mapping strategy into the development of Chinese medical ontology.

**More information:** Luming Chen et al, Mapping Chinese Medical Entities to the Unified Medical Language System, *Health Data Science* (2023). [DOI: 10.34133/hds.0011](https://doi.org/10.34133/hds.0011)

Provided by Health Data Science

Citation: Mapping Chinese medical entities to the Unified Medical Language System (2023, May 8) retrieved 20 April 2024 from <https://medicalxpress.com/news/2023-05-chinese-medical-entities-language.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private

study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.