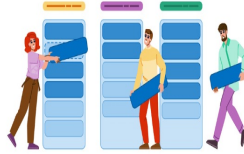


---

## Leveraging RWE to Automate Drug Indication Taxonomy Construction with LLMs



---

Acquiring, normalising, and formalising medical knowledge is fundamental in medical informatics, particularly focusing on drug indications. It highlights the challenges in curating drug indications and normalising them for consistent interpretation. Existing methods mainly involve manual or semi-automated processes using tools like MetaMap and ontologies like SNOMED-CT. However, these methods have limitations, prompting the need for automated approaches.

[This recent study published in JAMIA](#) proposes a novel approach using large language models (LLMs) and real-world evidence (RWE) to automate the construction of drug indication taxonomies. It aims to extract indication terms from drug labels, establish subsumption relations between indications, and create a standardised taxonomy, termed Drug Indication Standardised Taxonomy Induced by Large Language Models (DISTILL). The study utilises GPT-4 as the foundation language model for this purpose.

Overall, the research aims to address the challenges in drug indication normalization by leveraging advanced AI technologies and real-world evidence to automate the process and improve semantic interoperability in clinical research and decision support systems.

### Deriving DISTILL: An Integrated Approach for Drug Indication Taxonomy Construction

The process of deriving DISTILL, a drug indication taxonomy, is described in detail, integrating GPT-4 and real-world evidence (RWE). The workflow involves multiple subtasks, including drug indication extraction and taxonomy learning.

For drug indication extraction, FDA-approved product labels from DailyMed are utilised. Indications are extracted using GPT-4 with a few-shot prompting technique, and then normalised and de-duplicated. Terms with similar meanings are identified and retained based on cosine similarities and GPT-4 evaluations.

The drug indication taxonomy learning process involves generating high-level categories and assigning indication terms to concepts based on subsumption relations. Information of concepts is calculated using cosine similarities between drug representations derived from real-world patient data. Subcategories are created iteratively based on information gain, with intra-procedural and post-procedural rules ensuring taxonomy consistency.

Evaluation of DISTILL involves quantitative topological analysis and qualitative comparison with SNOMED-CT. The accuracy of concept-to-concept and concept-to-term relations is assessed through expert evaluation.

### Extracting, Comparing, and Evaluating a Drug Indication Taxonomy

Key findings provide detailed statistics on the extraction of drug indication terms, highlighting the process and outcomes of the DISTILL taxonomy construction, and comparing it with SNOMED-CT.

Drug Indication Term Extraction: Initially, 4190 distinct indication terms were extracted from FDA-approved product labels. After post-processing, 2909 terms remained, linked to RxNorm drugs. The median number of terms per active moiety is 2.

Characteristics of DISTILL: DISTILL consists of 24 high-level categories, with each category containing various indication terms. The topological characteristics of DISTILL's sub-taxonomies for three high-level categories are analysed, indicating depth and width metrics. The taxonomy aims for specificity regarding drug indications.

Comparison with SNOMED-CT: Comparisons are made between DISTILL and SNOMED-CT for three high-level categories: Cardiovascular diseases, Endocrine system diseases, and Genitourinary system diseases. While SNOMED-CT offers a more granular classification, DISTILL provides a smaller taxonomy with comparable coverage. Notable overlaps and discrepancies between the two classifications are highlighted, indicating areas of alignment and divergence.

Performance Evaluation of DISTILL: The accuracy of DISTILL's concept-to-concept subsumption relations is generally high, with inter-rater reliability scores indicating good reliability. However, concept-to-term subsumption relation accuracies vary, with lower reliability scores. Consistency scores for GPT-4 judgments on concept-to-term subsumption relations are provided, indicating varying levels of consistency based on evaluator agreement on GPT-4's accuracy.

Overall, the results showcase the effectiveness of DISTILL in providing a drug indication taxonomy, though there are areas for improvement, particularly in concept-to-term relations. Comparisons with SNOMED-CT highlight differences in granularity and coverage, emphasising the unique contributions and challenges of each classification system.

### **Design, Evaluation, and Future Directions of a Three-Level Drug Indication Taxonomy**

The decision to implement a three-level hierarchy for creating subcategories stemmed from extensive experimentation, aiming to strike a balance between depth and granularity. This depth allows for a taxonomy comparable to industry standards while ensuring specificity in drug indication classification.

The study aligns with key principles for controlled medical taxonomies, including granularity, expansion facilitation, polyhierarchy, and avoidance of "not elsewhere classified" concepts. The pipeline accommodates the continuous expansion of content and allows for polyhierarchy, where concepts can have multiple parent concepts.

Evaluation of the taxonomy's performance indicates higher accuracy and reliability for concept-to-concept subsumption relations compared to concept-to-term relations. Errors in the latter are identified and attributed to shared children issues and inversion of concept-term subsumption relations.

The study acknowledges limitations, such as the availability of RWE for information gain computation and the use of only one LLM (GPT-4). Suggestions for future research include exploring the potential of non-deterministic LLM outputs and comparing different LLMs for taxonomy construction.

In conclusion, generative AI, in conjunction with RWE, can support taxonomy development activities but may not fully support end-to-end processes. The proposed pipeline provides a framework applicable beyond drug indications, with potential applications in various medical knowledge acquisition, normalisation, and classification tasks. Further evaluation is needed to assess its support for downstream tasks and its generalizability.

**Source:** [JAMIA](#)

**Image Credit:** [iStock](#)

Published on : Thu, 30 May 2024