

New paper published in JAMIA explores results of SNOMED CT entity linking challenge

SNOMED has documented the development of entity linking models to link spans of free-text data in clinical notes with specific topics in SNOMED CT.

LONDON, LONDON, UNITED KINGDOM, October 8, 2025

/EINPresswire.com/ -- SNOMED

International, along with a team of research and technology focused subject matter experts, recently

contributed to [a research paper](#) documenting the development of entity linking models to link spans of free-text data in clinical notes with specific topics in the clinical terminology SNOMED CT. The paper, which had been planned as part of the competition, was recently published in the highly respected, peer-reviewed Journal of the American Medical Informatics Association, the

outcomes of which were derived from SNOMED International's Entity Linking Challenge held in 2024.

“

There are plenty of approaches and models for doing this; but what there isn't [yet] is a high performing set of well-trained entity linking models out there that can be ported to the many use cases”

*Will Hardman, Veratai Data
Science Director*

Entity linking involves identifying and labeling the portions of clinical notes that correspond to specific medical concepts. Most of the currently existing clinical content is in free text, making it difficult to analyze and extract meaningful insights from it. Entity linking enables healthcare organizations to convert that data into a structured format that can be readily analyzed by computers, in turn stimulating the development of new medicines, treatment pathways, and better patient outcomes.

SNOMED CT is a systematically organized clinical terminology used by healthcare providers globally to facilitate the accurate recording, sharing, and analysis of clinical data. It is owned, administered and developed by the not-for-profit organization SNOMED International.

The [SNOMED CT Entity Linking Challenge](#), which ran from January to March 2024, trained



Delivering
SNOMED CT

The global language of healthcare

machine learning models to link clinical notes with specific topics based on the largest publicly available dataset of labeled clinical notes that had been de-identified and annotated with SNOMED CT concepts. It provided an opportunity to advance the development of efficient and reliable tools for automating the coding of patient data, facilitating interoperability and decision support, and improving healthcare delivery. The challenge was supported by platform host partner DrivenData, which hosts online data science competitions; AI consultancy Veratai; Physionet, the Research Resource for Complex Physiologic Signals, and an annotation team.

There were 553 registered entrants. The three winning teams submitted varied approaches: a dictionary-based method, an encoder-based method, and a decoder-based method, respectively.

In a [2024 Coded Conversations podcast](#) on the competition, SNOMED International Chief Digital Information Officer Rory Davidson and AI Advisor Will Hardman (who was also a co-organizer of the challenge and a co-author of the paper) outlined the difficulties posed by analyzing the healthcare data stored in free-text documents such as in clinician-entered patient notes, and explained how clinical terminologies such as SNOMED CT enable healthcare professionals to map free-text to medical concepts, enabling the records to be unambiguously analyzed by computers. "It's a niche field," said Will, "but it's a field that has been under study for a couple of decades. There are plenty of approaches and models out there for doing this; there are models in production ... in hospitals across the world .. but what there isn't [yet] is a really high performance set of well-trained entity linking models out there in the public domain that can be ported to all the many, many use cases."

The paper, which was co-authored by SNOMED International, Veratai and the winning teams of the competition, describes the basis of the work – a large set of 74,808 annotations curated across 272 discharge notes spanning 6,624 unique clinical concepts – and the evaluation process. It compares the approaches used by the winning solutions and highlights the most challenging factors affecting clinical entity linking models. It also describes the data set and the policy-based approach to the development of the "ground truth" data set and provides an example of its approach to scoring. Importantly, it analyzes the reasons for low-scoring concepts, and details a number of lessons learned.

Referencing the creation of annotation guidelines and the publicly available annotated data set as part of the outcomes of the challenge, Rory described the project as one that "keeps on giving," adding that the paper points to the importance of developing tools and techniques for monitoring annotation quality and for spotting and fixing potential inconsistencies. "What we found is that the data itself is more important to successful entity linking than the approach used, as our comparative analysis highlighted few meaningful distinctions in the strengths and weaknesses of each," he said. "This challenge showed that even the most advanced AI models, including large language models, are only as good as the data they learn from. This challenge was a big step forward, and we're excited to see it spark further progress along with our goals of facilitating interoperability and decision support, and to improving healthcare delivery."

The Entity Linking Challenge also supports the technologies focus of SNOMED International's 2025-2030 Strategy, which aims to promote and support quality products, services and tools leveraging current and emerging technologies, supporting seamless health information flow from the point of capture to the point of data consumption; and prioritize initiatives focused on making SNOMED CT easier to adopt and simpler to implement.

Learn more about the Entity Linking Challenge and read the paper.

Kelly Kuru

SNOMED International

+ +1 416-566-8725

comms@snomed.org

Visit us on social media:

[LinkedIn](#)

[YouTube](#)

[X](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/854324551>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2025 Newsmatics Inc. All Right Reserved.