Turbocharging Clinical NLP: Using UMLS Hierarchies to Model Word Cloud Knowledge Base for Faster Insight and Relationship Discovery with Al Assistance

Sina Madani, Asli Weitkamp Department of HealthIT, Vanderbilt University Medical Center

PURPOSE

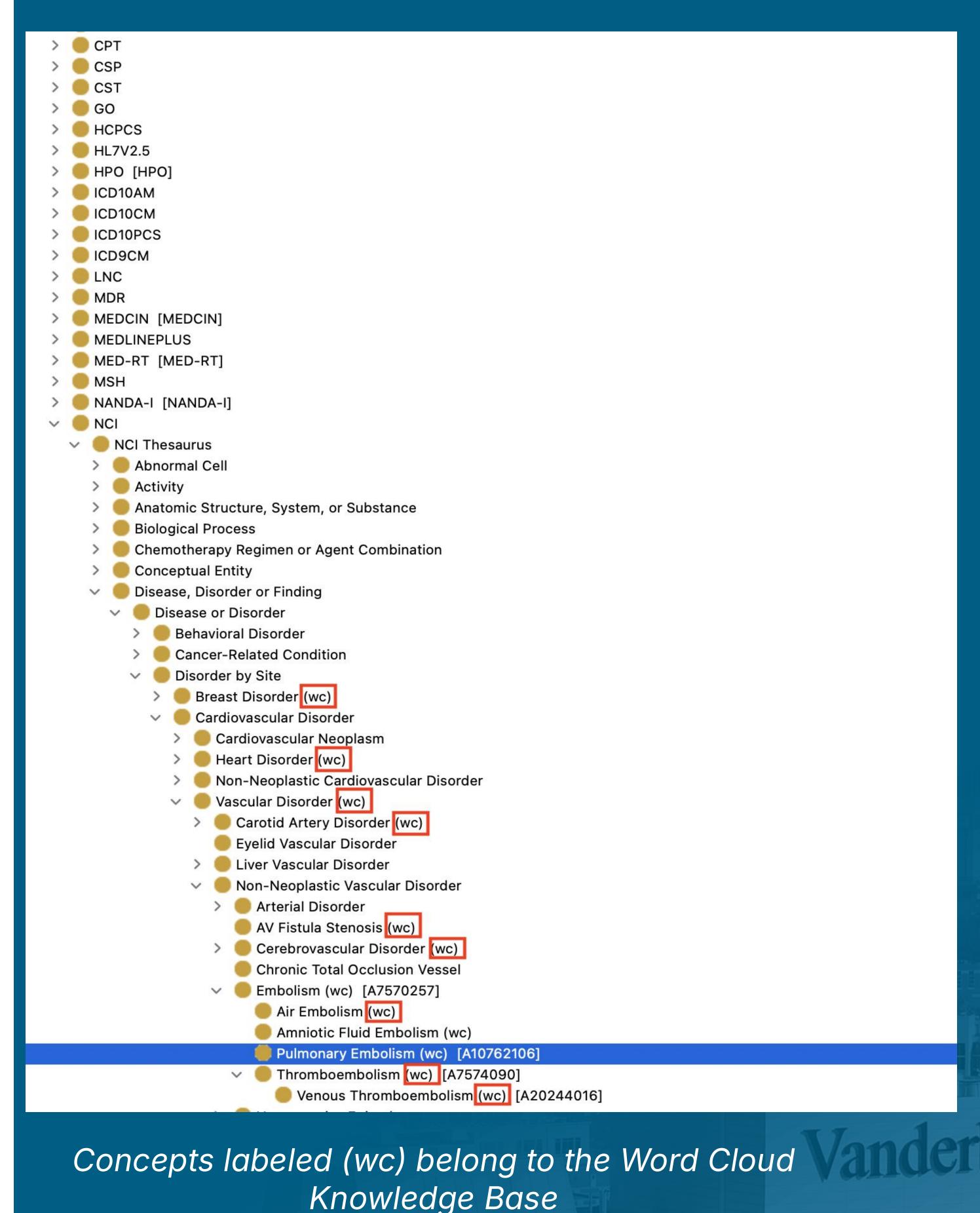
- VUMC Enterprise NLP application (World Cloud) lacks hierarchical structure
- UMLS KB provides >100 terminologies with rich relationships.
- Goal: Embed local CUIs in UMLS hierarchy → scalable, reusable ontology for NLP projects.

METHODS

- Queried UMLS MRCONSO & MRHIER tables in Databricks to filter CUIs, AUIs, and hierarchical paths.
- Used Al tools for schema interpretation, Spark SQL design, and RDF modeling
- Upward & downward hierarchical traversal for top-level and granular concept retrieval
- Graph Build Ontology & integration into TopBraid / NCBO Bioportal

Ontology enables instant expansion of clinical concepts within Word Cloud KB

Word Cloud KB Concepts like
"heart valve disease", "thromboembolism"
can be quickly expanded into narrower/broader
concepts across terminologies



Embolism (wc) [A7873197] Air embolism (wc) [A2990041] Amniotic fluid embolism (wc) [A2889579] Embolism due to and following ectopic pregnancy Embolism from thrombosis of vein of distal lower extremity Failed attempted termination of pregnancy complicated by embolism Fat embolism (wc) [A3021705] Induced termination of pregnancy complicated by embolism Miscarriage complicated by embolism Occlusion of microvasculature of skin caused by embolus Pyemic and septic embolism in pregnanc Acute occlusion of artery due to thromboembolus Thromboembolus of artery of lower limb Thromboembolus of precerebral artery Chronic thromboembolic pulmonary hypertension (wc) [A23469216] Thromboembolism of vein

Al tools like aiChat, ChatGPT, and Databricks Assistant reduced ontology build and NLP concept analysis time from days to hours

Results

- Ontology coverage:
 > 15,000 CUIs classified
- Efficiency: Ontology generated from multi-million paths in < 2 hours
- Utility: Supports instant terminology expansion in NLP pipelines

Conclusion

- Faster development:
 Al reduced manual coding/debugging time.
- More accurate retrieval:
 Consistent terminology
 expansion and relationship
 discovery
- Clinical Impact: Improves speed and precision of concept identification for enterprise NLP projects

Presenter: Sina Madani





Department of HealthIT