



Medical Language Models for Data Scientists

April 2025

An LLM-enabled Medical Terminology Server

Kate Weber, PhD
Sr Data Scientist
John Snow Labs



Agenda

- Code Systems in Healthcare
- Specific Coding Systems
- Code Mapping
- Challenges when working with code systems
- Value Sets and Concept Maps
- The John Snow Labs Terminology Server
- Terminology Server API

Health Coding Systems

[illegible]

Health Coding Systems

- Consistent representation of a health concept
- Uses
 - **Billing** – describe procedures the same way every time
 - **Precision** – resolving multiple laboratory tests with different text names
 - **Translation** – MESH D005360 means
 - “Fibula”, “Perone”, “Fíbula”, “الشظية” and “腓骨”
 - **Relationships**
 - Codeine is an ingredient in thousands of products

	concept_class_id	count
1	Ingredient	1
2	Branded Pack Box	2
3	Clinical Pack Box	2
4	Clinical Pack	50
5	Branded Pack	86
6	Clinical Drug Comp	145
7	Clinical Dose Group	183
8	Quant Branded Box	308
9	Quant Branded Drug	330
10	Quant Clinical Box	334
11	Quant Clinical Drug	357
12	Branded Dose Group	373
13	Clinical Drug Form	383
14	Clinical Drug	827
15	Branded Drug Form	896
16	Branded Drug Comp	971
17	Branded Drug	1088
18	Clinical Drug Box	1152
19	Branded Drug Box	1551
20	Marketed Product	2719

Common Systems



LOINC®



ICD-10



SNOMED
International

Challenges in Translation

- Coding systems are intended to reduce chaos
- Because there's lots of chaos to reduce
- Synonyms
- Varying Levels of Detail
- Mapping between vocabularies
- Limitations of string searching
- Multiple Languages

<input type="checkbox"/> code	<input type="checkbox"/> concept_name
V32.3	Occupant of three-wheeled motor vehicle injured in collision with two- or three-wheeled motor vehicle, unspecified occupant of three-wheeled motor vehicle injured in nontraffic accident
V35.3	Occupant of three-wheeled motor vehicle injured in collision with railway train or railway vehicle, unspecified occupant of three-wheeled motor vehicle injured in nontraffic accident
V32.9	Occupant of three-wheeled motor vehicle injured in collision with two- or three-wheeled motor vehicle, unspecified occupant of three-wheeled motor vehicle injured in traffic accident
V34.3	Occupant of three-wheeled motor vehicle injured in collision with heavy transport vehicle or bus, unspecified occupant of three-wheeled motor vehicle injured in nontraffic accident
V35.9	Occupant of three-wheeled motor vehicle injured in collision with railway train or railway vehicle, unspecified occupant of three-wheeled motor vehicle injured in traffic accident

Meta- Vocabularies



Managing the Chaos with FHIR



Fast Healthcare Interoperability Resources

- Value Sets
 - A collection of concepts from any vocabulary
- Applications
 - Billing
 - Reporting
 - Research

domain	custom_code	custom_term	custom_term_alias	custom_code_system	standard_code	standard_term	standard_code_system
Condition	CC100	Hypertension	Blood Pressure	CCS	38341003	Hypertension	SNOMED
Observation	CC200	History of Hypertention	History of blood pressure	CCS	Z86.79	History of Hypertention	ICD10CM
Means Value	CC300	HBP	high bp	CCS	LA7444-8	Hypertension	LOINC
Condition	CC400	Kidney Hypertension	kidney hp	CCS	28119000	renal hypertension	SNOMED

Managing the Chaos with FHIR



Concept Maps

- Mapping between vocabularies
- Especially local custom concepts to standard vocabularies

domain	custom_code	custom_term	custom_term_alias	custom_code_system	standard_code	standard_term	standard_code_system
Condition	CC100	Hypertension	Blood Pressure	CCS	38341003	Hypertension	SNOMED
Observation	CC200	History of Hypertention	History of blood pressure	CCS	Z86.79	History of Hypertention	ICD10CM
Means Value	CC300	HBP	high bp	CCS	LA7444-8	Hypertension	LOINC
Condition	CC400	Kidney Hypertension	kidney hp	CCS	28119000	renal hypertension	SNOMED

The John Snow Labs Terminology Server

- **Terminology searches with LLM support**
 - We match by **semantic similarity**
- Includes 24 **Common Vocabularies**
- Understands OMOP **Standard Concepts**
- **Maps** between vocabularies
- Supports **Value Sets**
- Features both **GUI** and **API** interaction

Search

API Service

Code Systems ⓘ

Value Sets ⓘ

Upload

Concept Maps ⓘ



TERMINOLOGY SERVER

Look up standard medical codes from text phrases

hypertension



Coding Results (12)

Download (0)

<input type="checkbox"/>	Code System ↑↓	Concept Name ↑↓	Concept Code ↑↓	Chunk ↑↓	Start Date ↑↓	End Date ↑↓	Domain ↑↓
<input type="checkbox"/>	NCIt	hypertension	C160273	hypertension			
<input type="checkbox"/>	LOINC	hypertension	LA7444-8	hypertension	1970-01-01	2099-12-31	Meas Value
<input type="checkbox"/>	LOINC	hypertension	LP74941-3	hypertension	1970-01-01	2099-12-31	Observation
<input type="checkbox"/>	LOINC	hypertension (high blood pressure)	LA32904-7	hypertension	1970-01-01	2099-12-31	Meas Value
<input type="checkbox"/>	LOINC	medical-hypertension	LA17551-5	hypertension	1970-01-01	2099-12-31	Meas Value
<input type="checkbox"/>	SNOMED	hypertensive disorder	38341003	hypertension	2002-01-31	2099-12-31	Condition
<input type="checkbox"/>	ICD10CM	ht - hypertension	I10	hypertension			Condition
<input type="checkbox"/>	MeSH	hypertension	D006973	hypertension			Condition
<input type="checkbox"/>	MedDRA_LLT	hypertension	10020772	hypertension			
<input type="checkbox"/>	MedDRA_PT	hypertension	10020772	hypertension			
<input type="checkbox"/>	UMLS	hypertension	C0020538	hypertension			
<input type="checkbox"/>	ICD10	essential (primary) hypertension	I10	hypertension	1990-05-01	2099-12-31	Condition

View 50 Items per page

Showing 1 - 12 of 12 items

< 1 >

Demo

API Service

FastAPI 0.1.0 OAS 3.1

/openapi.json

default



GET /api/healthcheck Ping



POST /api/upload_valuesets Upload Valuesets



GET /api/upload_status/{task_id} Get Upload Status



GET /api/valuesets Get All Valuesets



POST /search Search With Valuesets



GET /api/valueset_details/{valueset_metadata_id} Valueset Details



GET /api/download_valueset Download Valueset

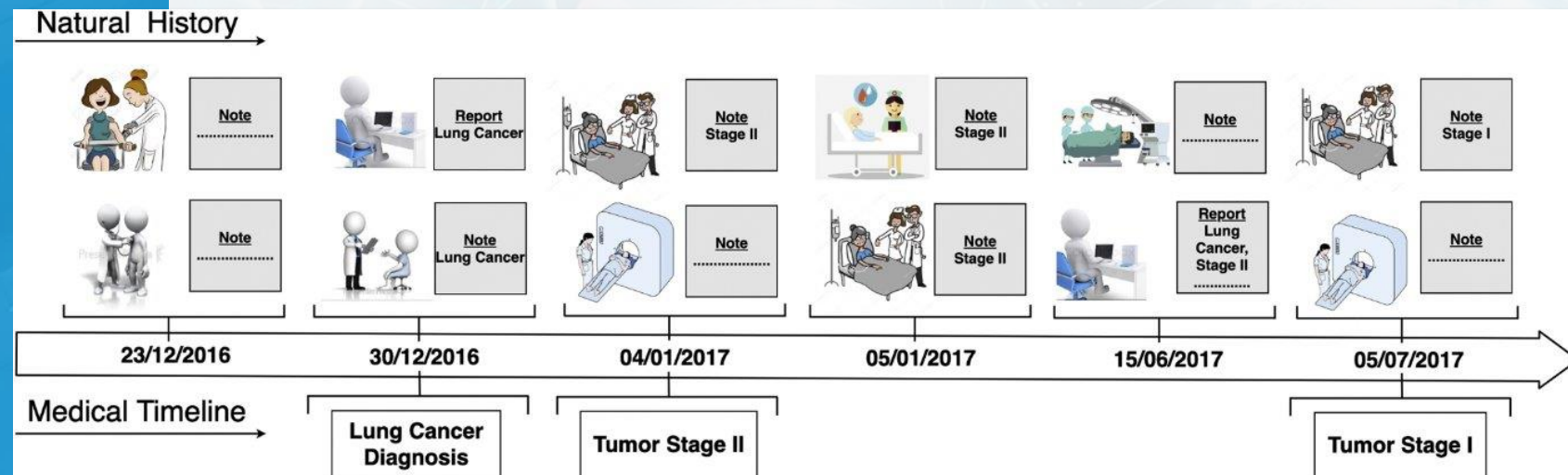


DELETE /api/delete_valueset Delete Valueset



Conclusion

- LLM-Enabled context search
- Able to search, filter, sort, and export
- 24 Built-In Vocabularies
- Continually Updated
- Supports Custom Value Sets
- Deployed in your environment
- Available now
- Key part of the JSL Patient Journeys product



Thank you!



kate@JohnSnowLabs.com



[in/katherineweber](https://www.linkedin.com/in/katherineweber)