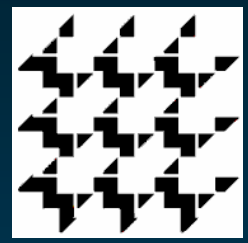
Institute for Discrete Sciences
Workshop on Associating Semantics with Graphs

Rutgers University
April 16, 2007



Unified Medical Language System

The graph behind the forest

Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA

Biomedical trees

http://www.tolweb.org/tree/

NCBI

Taxonomy Browser

Lineage (full): root; cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus

---

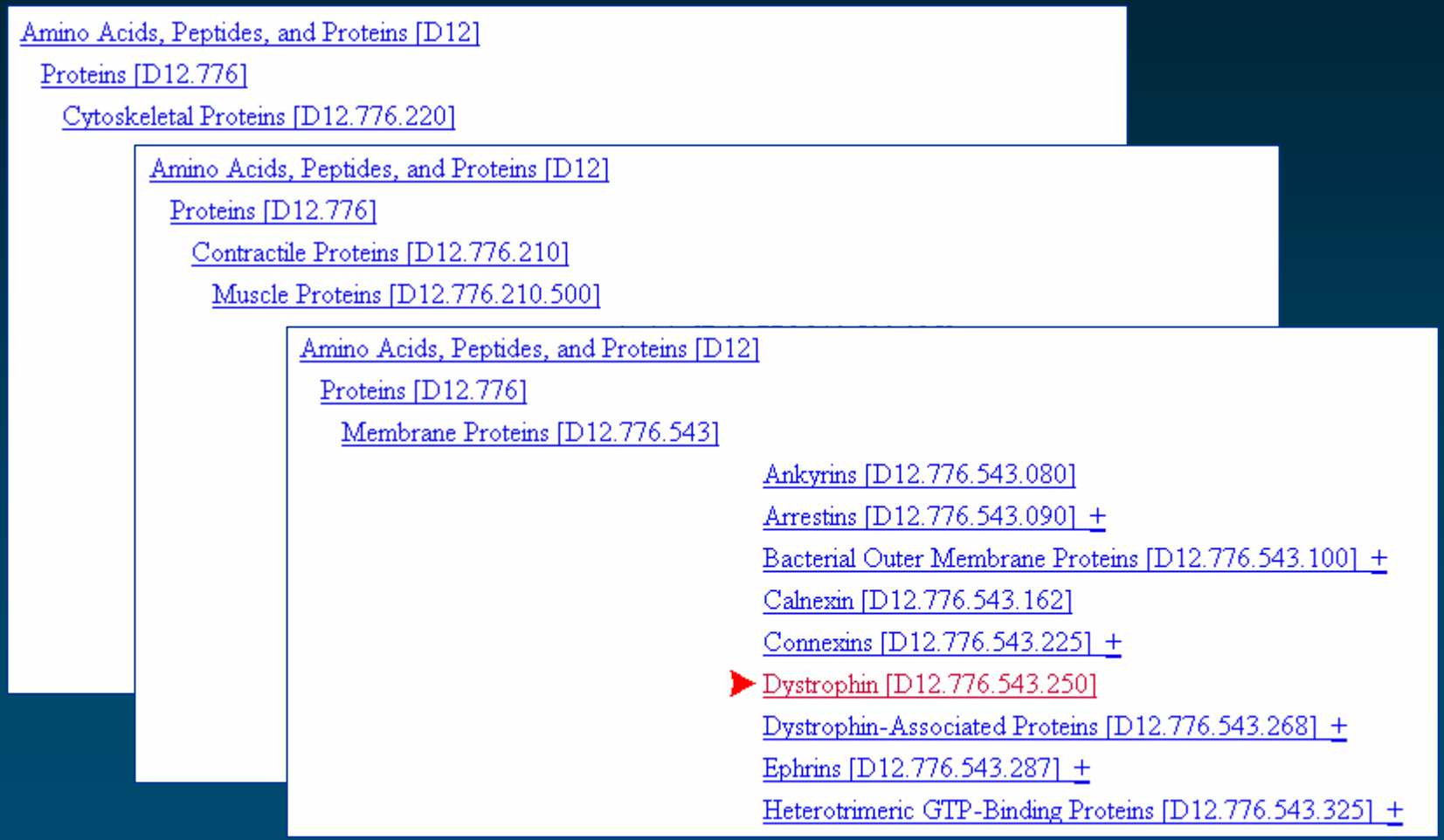
- **Mus musculus** (house mouse) *Click on organism name to get more information.*

  - **Mus musculus bactrianus** (southwestern Asian house mouse)
  - **Mus musculus castaneus** (southeastern Asian house mouse)
  - **Mus musculus domesticus** (western European house mouse)
  - **Mus musculus gentilulus**
  - **Mus musculus homourus**
  - **Mus musculus molossinus** (Japanese wild mouse)
  - **Mus musculus musculus** (eastern European house mouse)
  - **Mus musculus praetextus**
  - **Mus musculus wagneri**

http://www.ncbi.nlm.nih.gov/Taxonomy/

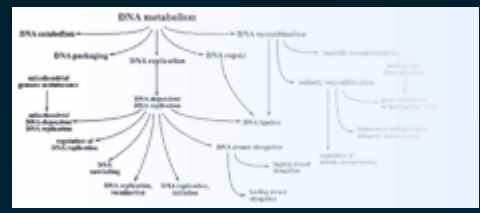
NLM

Medical Subject Headings



http://www.nlm.nih.gov/mesh/2007/MBrowser.html

Gene Ontology



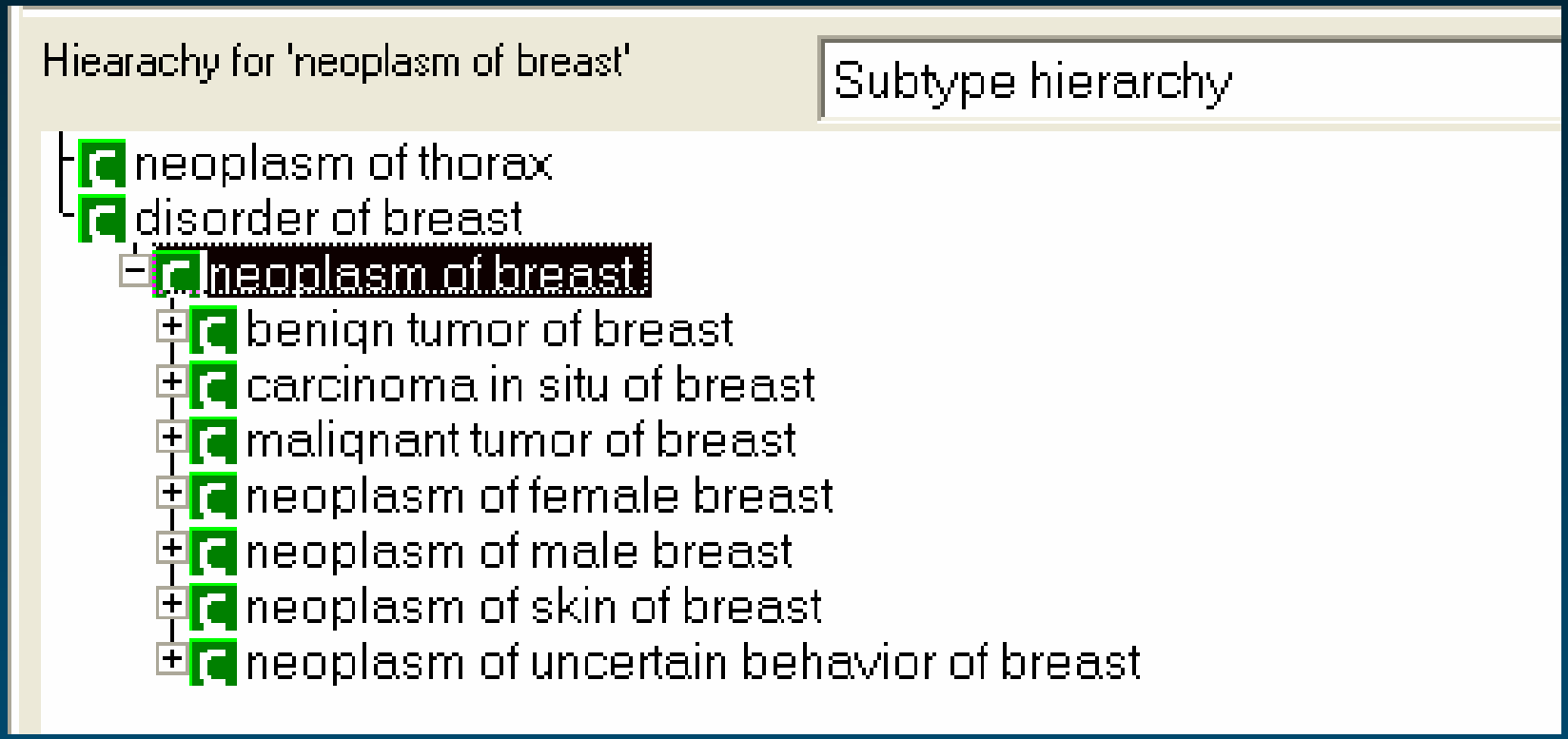
- all : all [189459]
  - GO:0008150 : biological_process [137743]
    - GO:0065007 : biological regulation [19935]
      - GO:0050789 : regulation of biological process [18154]
        - GO:0019222 : regulation of metabolic process [9816]
          - **GO:0019216 : regulation of lipid metabolic process [108]**

- GO:0008152 : metabolic process [54684]
  - GO:0044238 : primary metabolic process [44818]
    - GO:0006629 : lipid metabolic process [3940]
      - **GO:0019216 : regulation of lipid metabolic process [108]**

- GO:0019222 : regulation of metabolic process [9816]
  - **GO:0019216 : regulation of lipid metabolic process [108]**

http://amigo.geneontology.org/cgi-bin/amigo/go.cgi

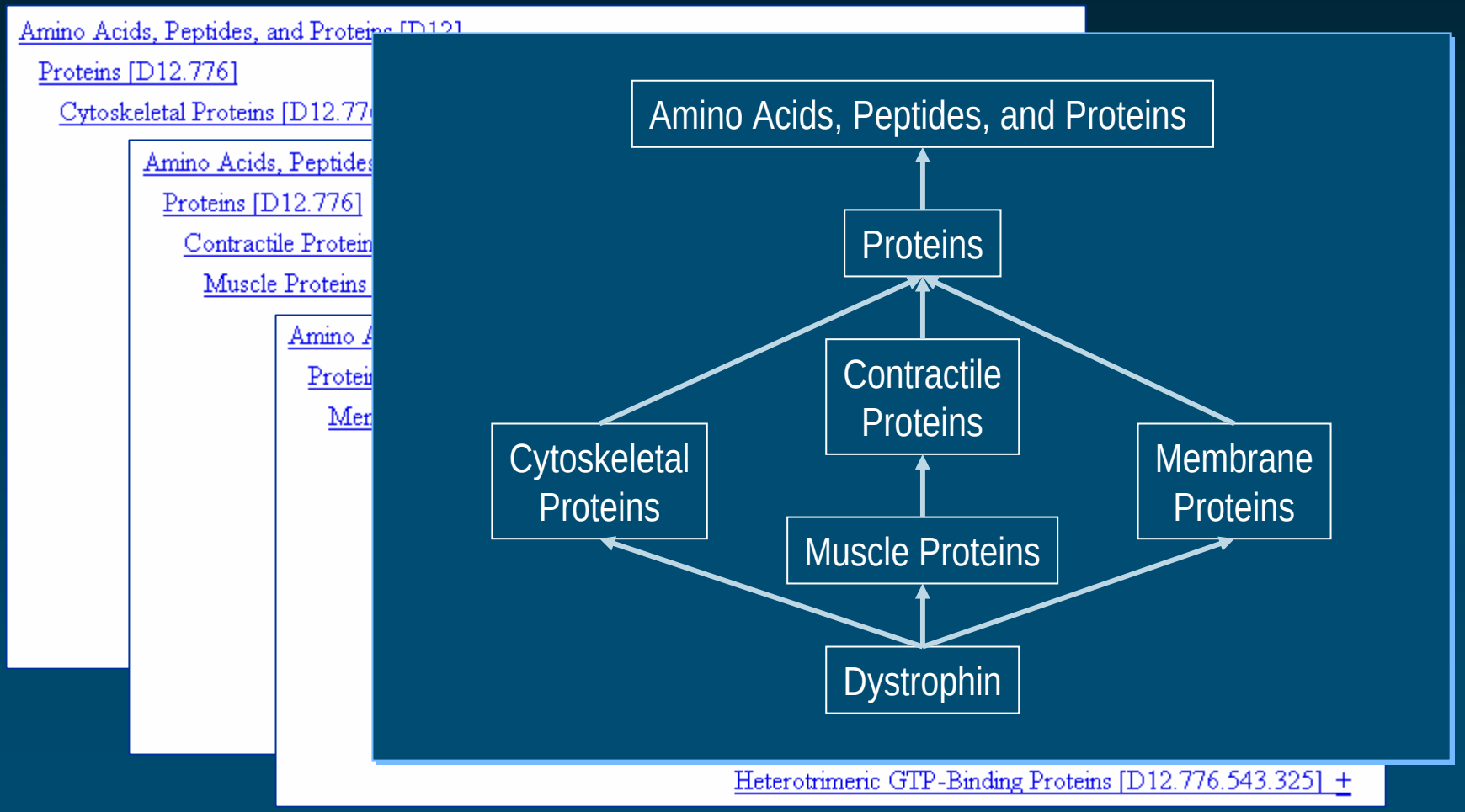
SNOMED Clinical Terms



http://www.clininfo.co.uk/clue5/clue.htm

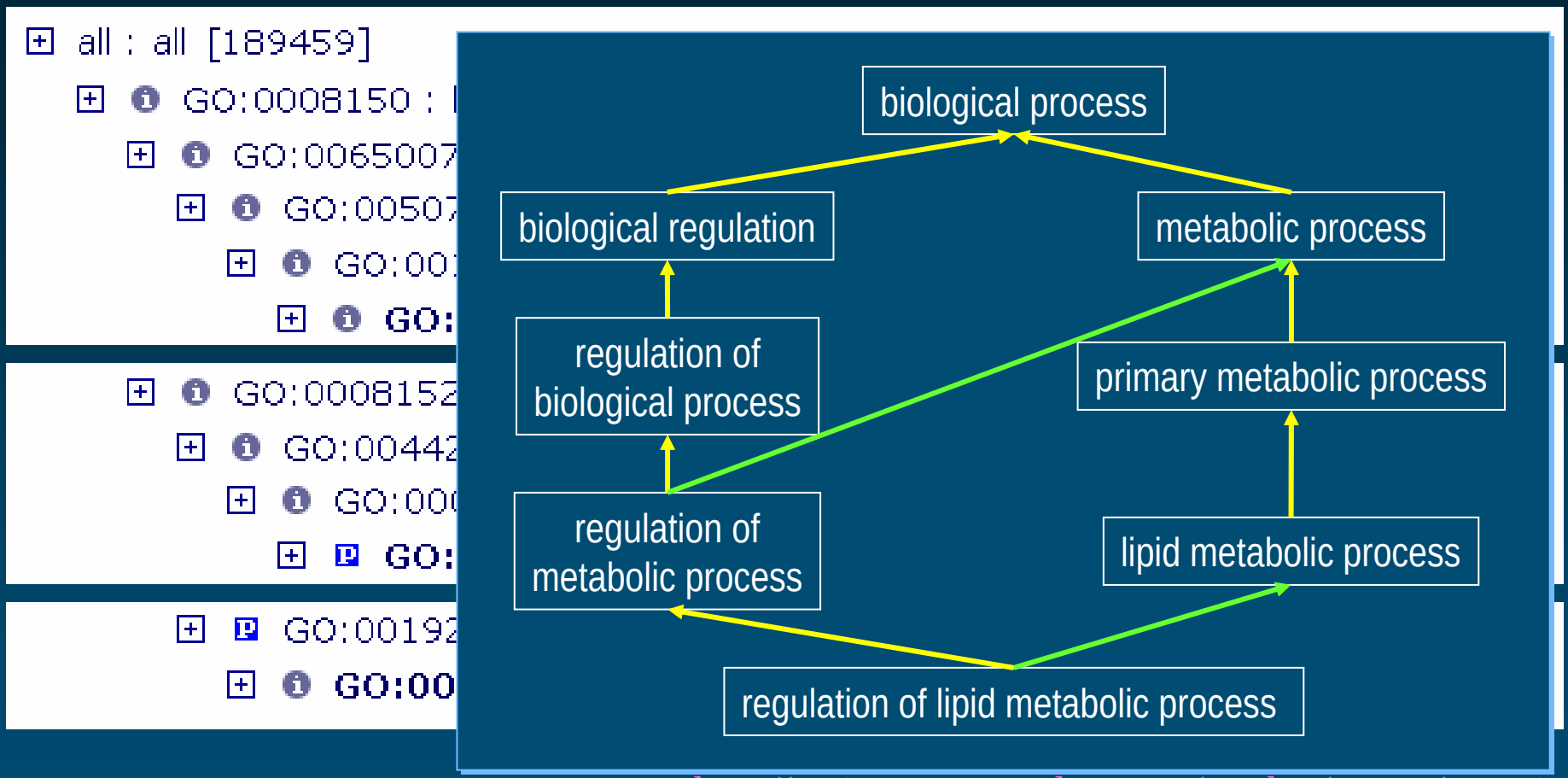
Biomedical trees revisited

Medical Subject Headings



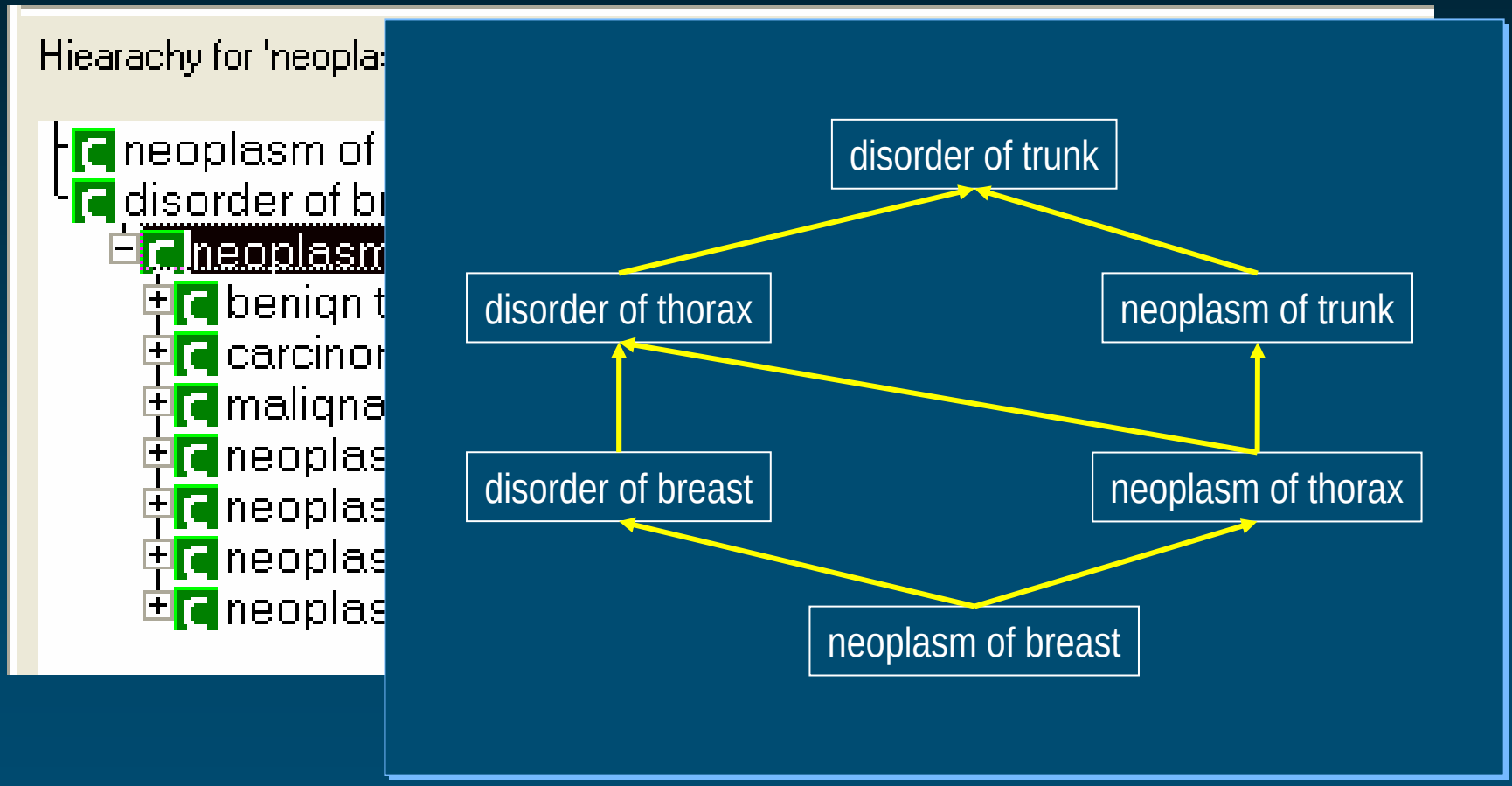
http://www.nlm.nih.gov/mesh/2007/MBrowser.html

Gene Ontology



http://amigo.geneontology.org/cgi-bin/amigo/go.cgi

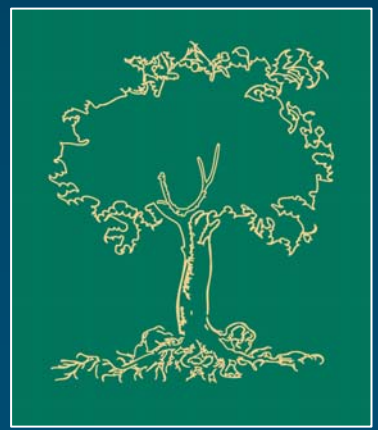
SNOMED Clinical Terms



http://www.clininfo.co.uk/clue5/clue.htm

Terminology integration

Unified Medical Language System



Addison's disease in medical vocabularies

- Synonyms
  - Addisonian syndrome — eponym
  - Bronzed disease — symptoms
  - Addison melanoderma — symptoms
  - Asthenia pigmentosa — symptoms
  - Primary adrenal deficiency — clinical variants
  - Primary adrenal insufficiency — clinical variants
  - Primary adrenocortical insufficiency — clinical variants
  - Chronic adrenocortical insufficiency — clinical variants

Organize terms

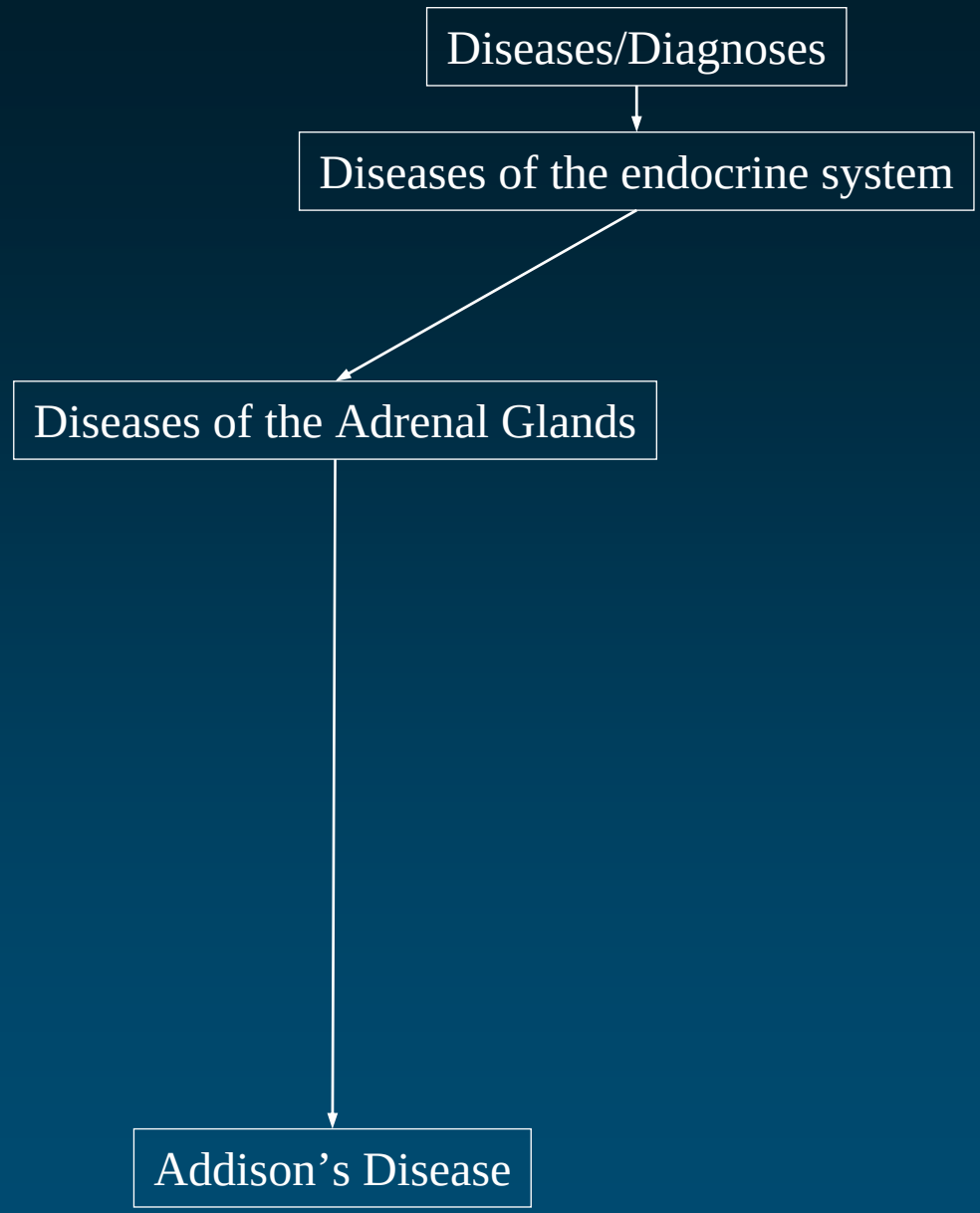
- Synonymous terms clustered into a concept
- Preferred term
- Unique identifier (CUI)

Addison Disease	MeSH	D000224
Primary hypoadrenalism	MedDRA	10036696
Primary adrenocortical insufficiency	ICD-10	E27.1
Addison's disease (disorder)	SNOMED CT	363732003

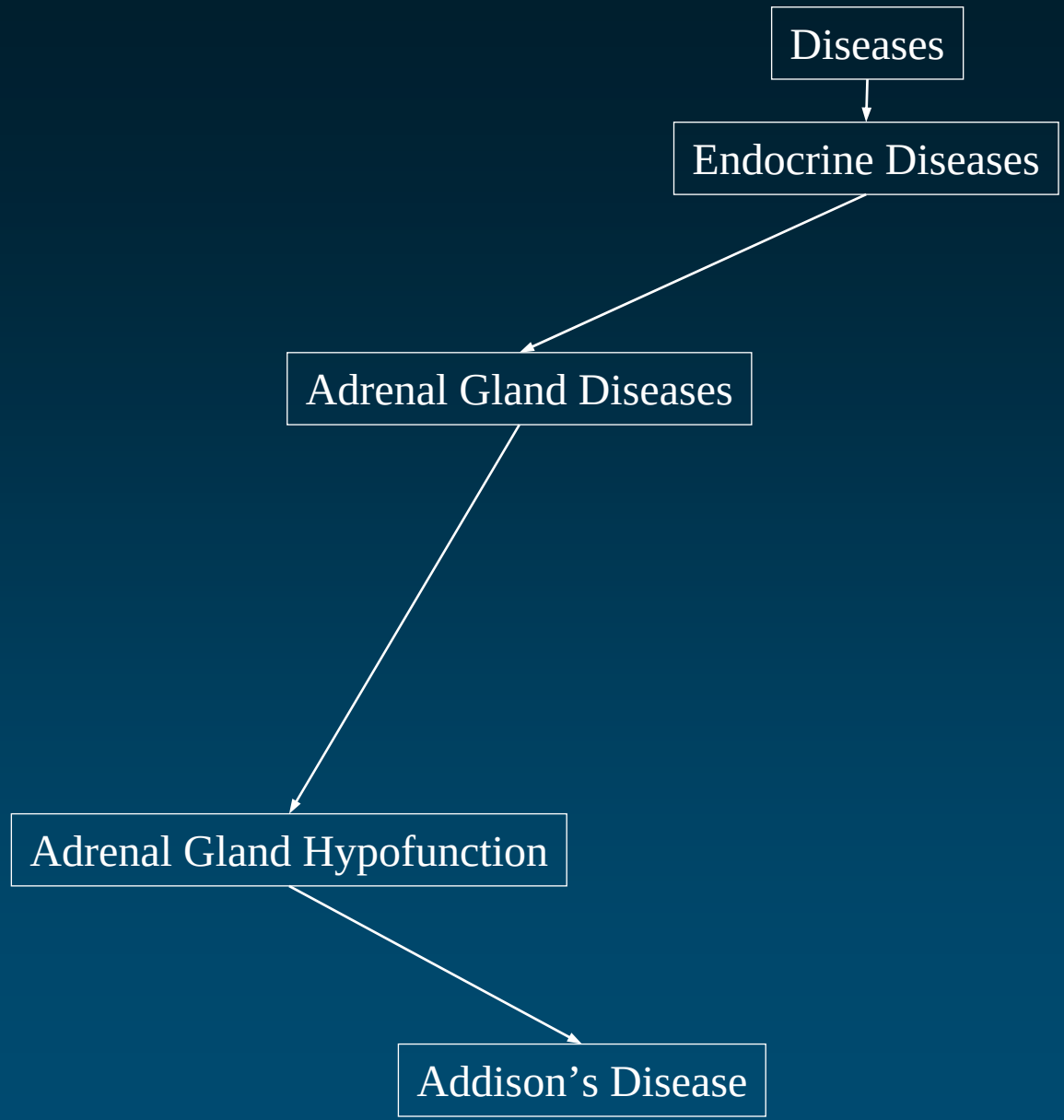
C0001403

Addison's disease

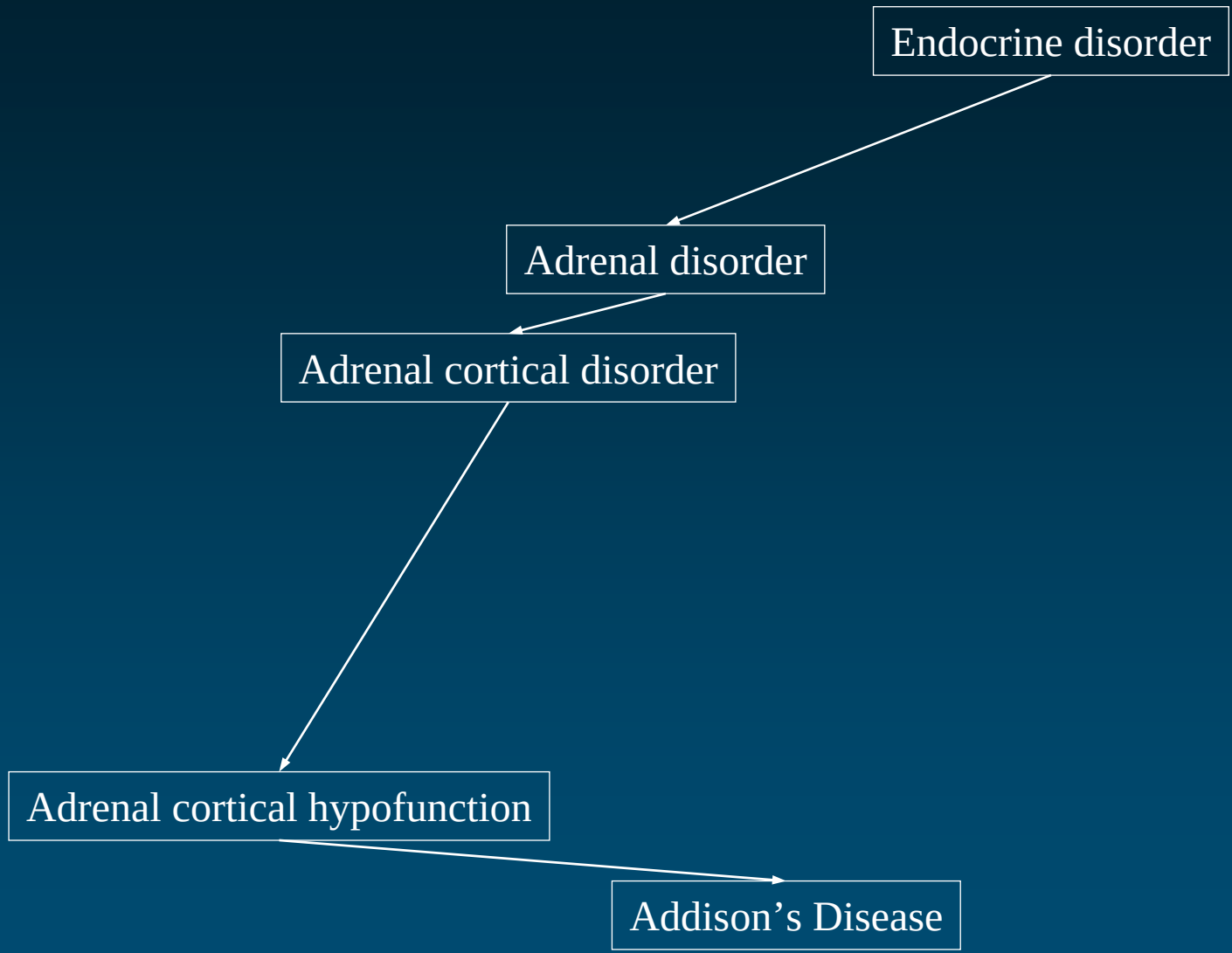
SNOMED International



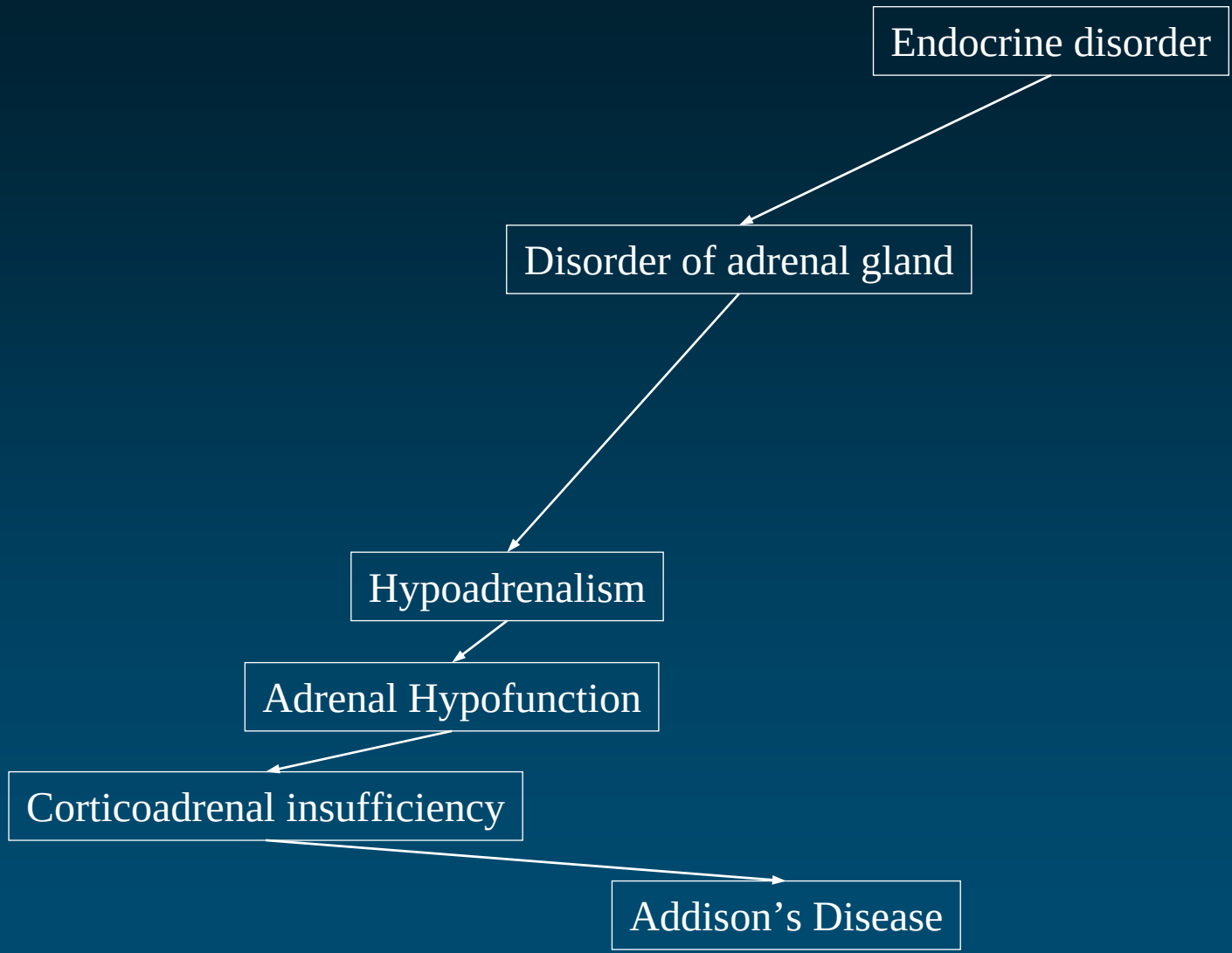
MeSH


Diseases
Endocrine Diseases
Adrenal Gland Diseases
Adrenal Gland Hypofunction
Addison’s Disease

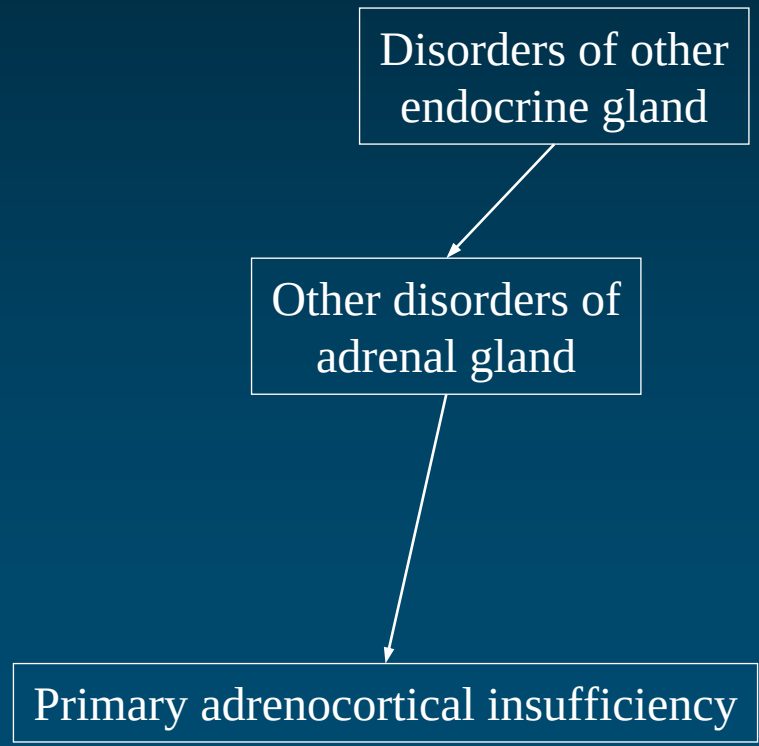
AOD


Endocrine disorder
Adrenal disorder
Adrenal cortical disorder
Adrenal cortical hypofunction
Addison’s Disease

Read Codes

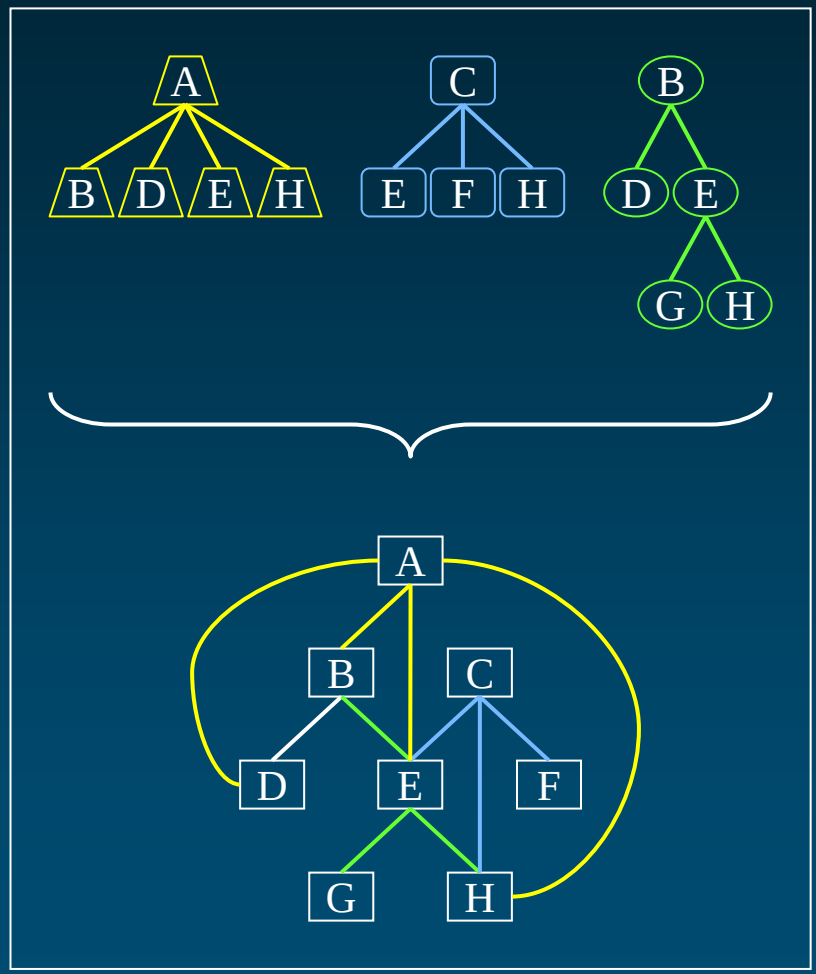

Endocrine disorder
Disorder of adrenal gland
Hypoadrenalism
Adrenal Hypofunction
Corticoadrenal insufficiency
Addison’s Disease

ICD-10


Disorders of other endocrine gland
Other disorders of adrenal gland
Primary adrenocortical insufficiency

Organize concepts

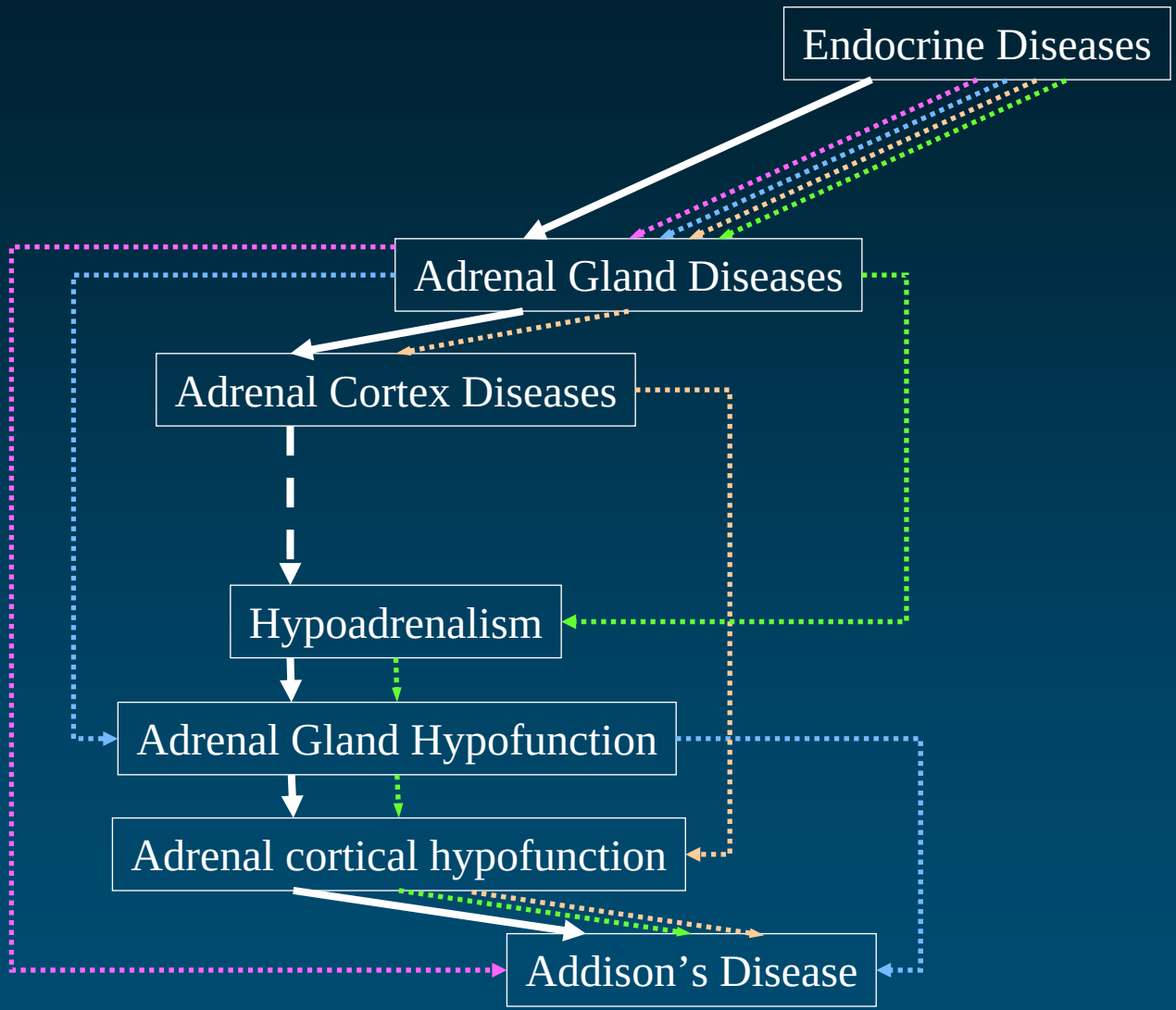
- Inter-concept relationships: hierarchies from the source vocabularies
- Redundancy: multiple paths
- One graph instead of multiple trees (multiple inheritance)

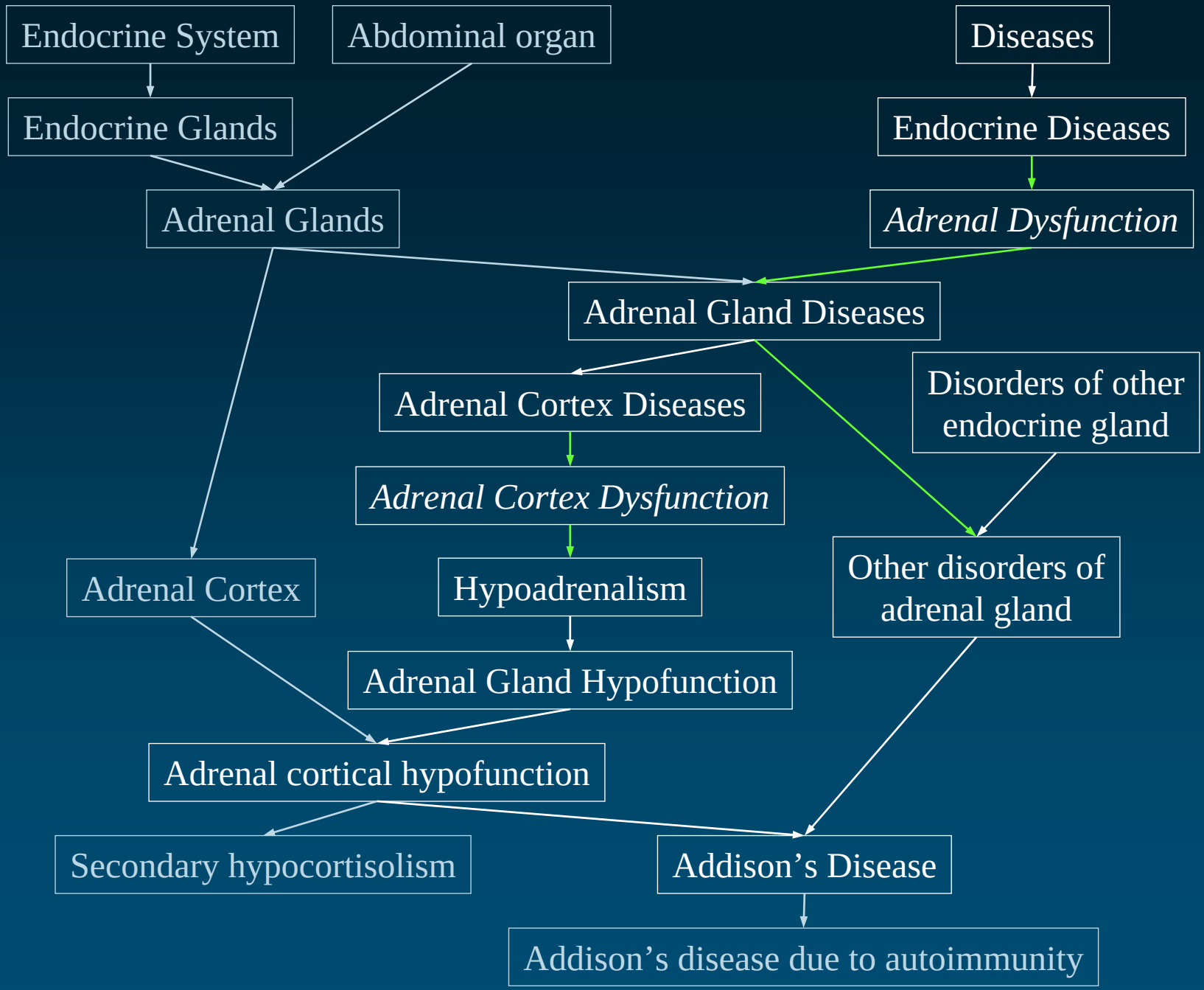


organize concepts

SNOMED
MeSH
AOD
Read Codes

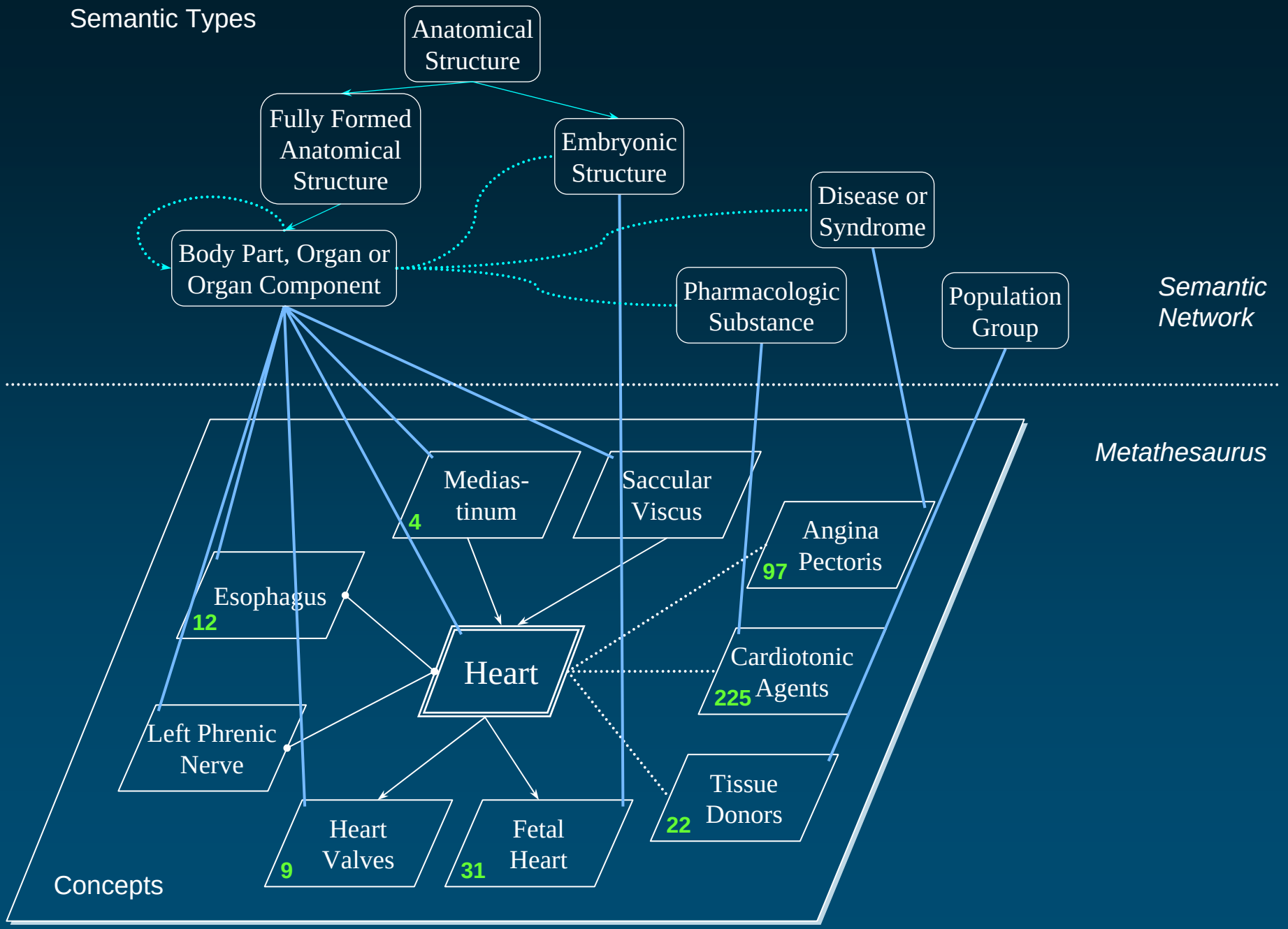
UMLS




Endocrine System
Abdominal organ
Diseases
Endocrine Glands
Endocrine Diseases
Adrenal Glands
Adrenal Dysfunction
Adrenal Gland Diseases
Adrenal Cortex Diseases
Disorders of other endocrine gland
Adrenal Cortex Dysfunction
Adrenal Cortex
Hypoadrenalism
Other disorders of adrenal gland
Adrenal Gland Hypofunction
Adrenal cortical hypofunction
Secondary hypocortisolism
Addison's Disease
Addison's disease due to autoimmunity

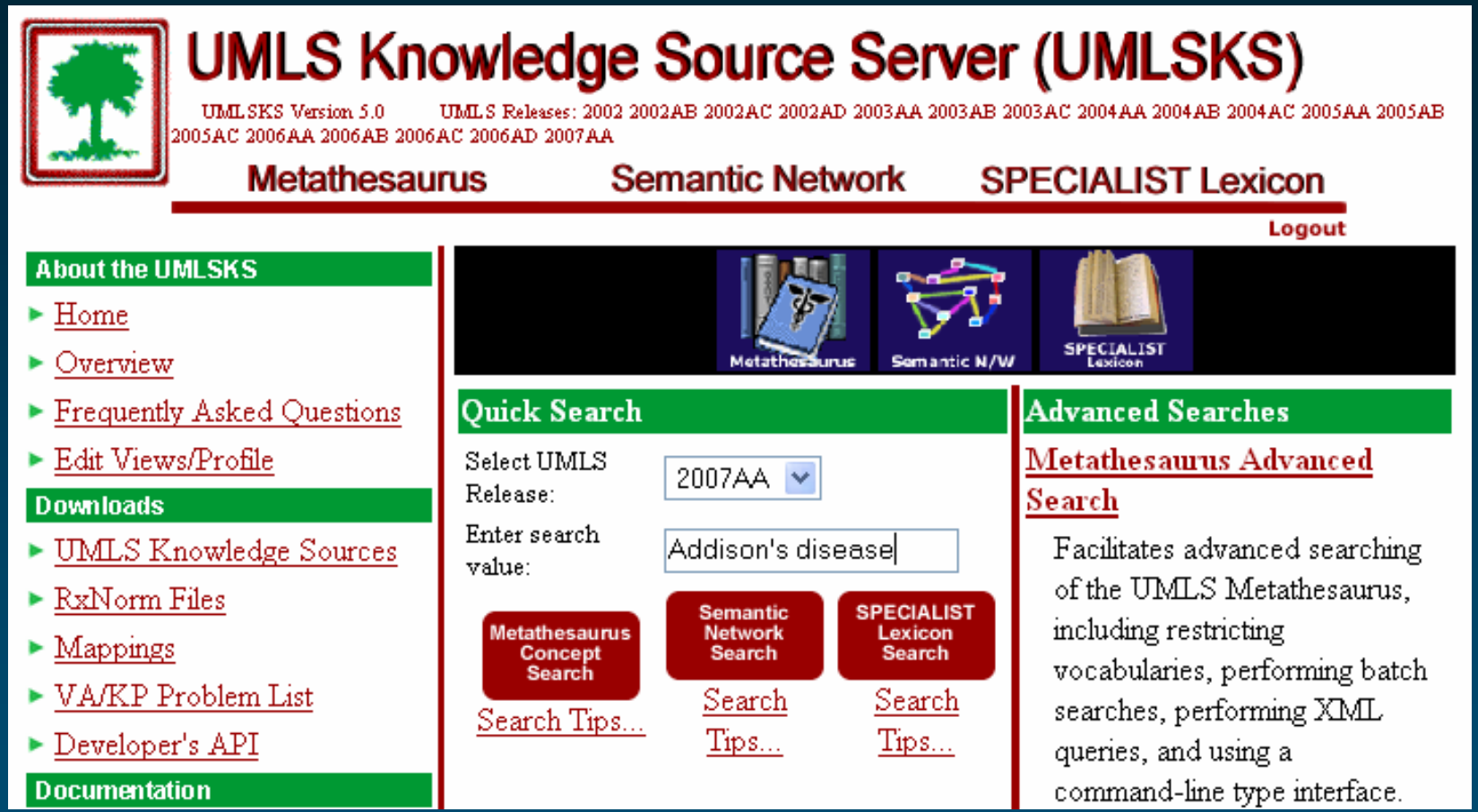
Source Vocabularies (2007AA)

- 139 source vocabularies
  - 17 languages
- Broad coverage of biomedicine
  - 5.5M names
  - 1.4M concepts
  - 16M relations
- Common presentation


Semantic Types
Anatomical Structure
Fully Formed Anatomical Structure
Embryonic Structure
Disease or Syndrome
Body Part, Organ or Organ Component
Pharmacologic Substance
Population Group
Semantic Network
Metathesaurus
Medias-tinum
4
Saccular Viscus
Angina Pectoris
97
Esophagus
12
Heart
Cardiotonic Agents
225
Left Phrenic Nerve
Tissue Donors
22
Heart Valves
9
Fetal Heart
31
Concepts

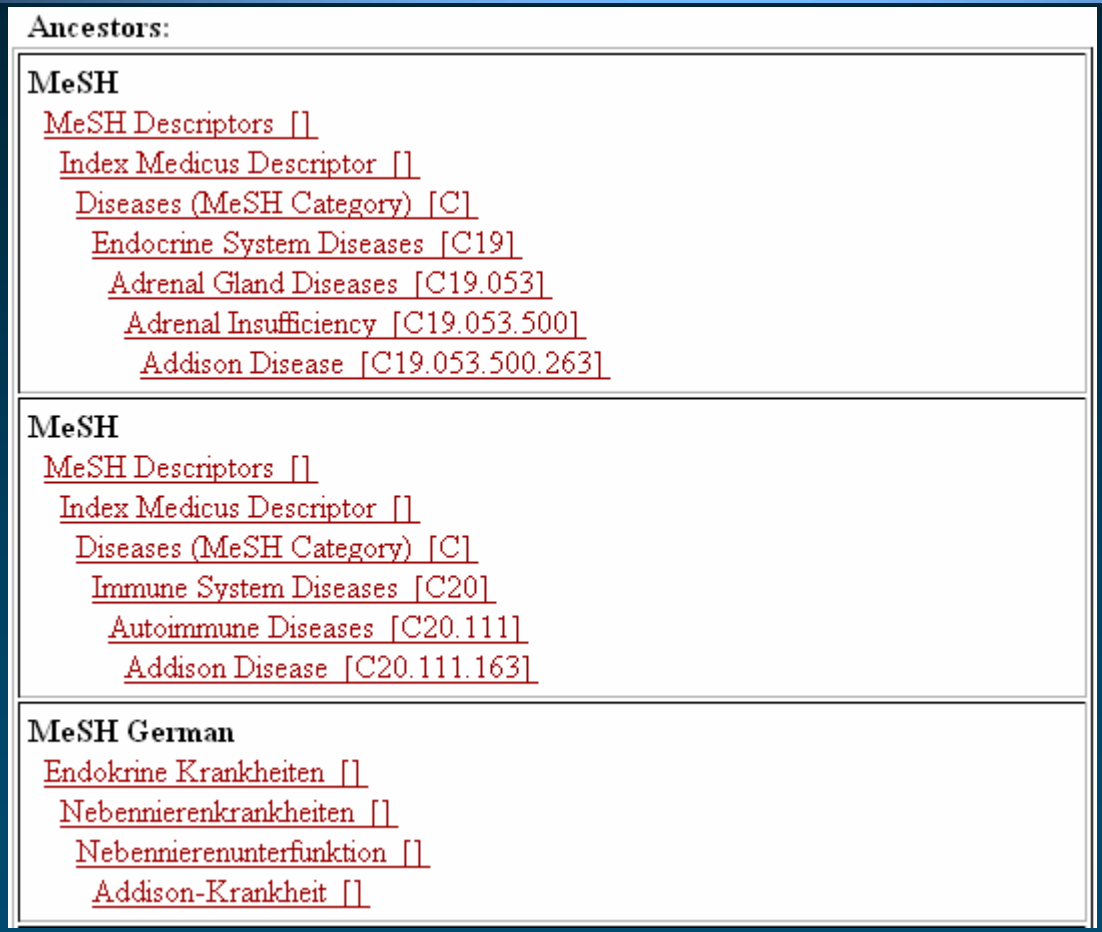
Biomedical forest vs. graph

UMLS Knowledge Source Server

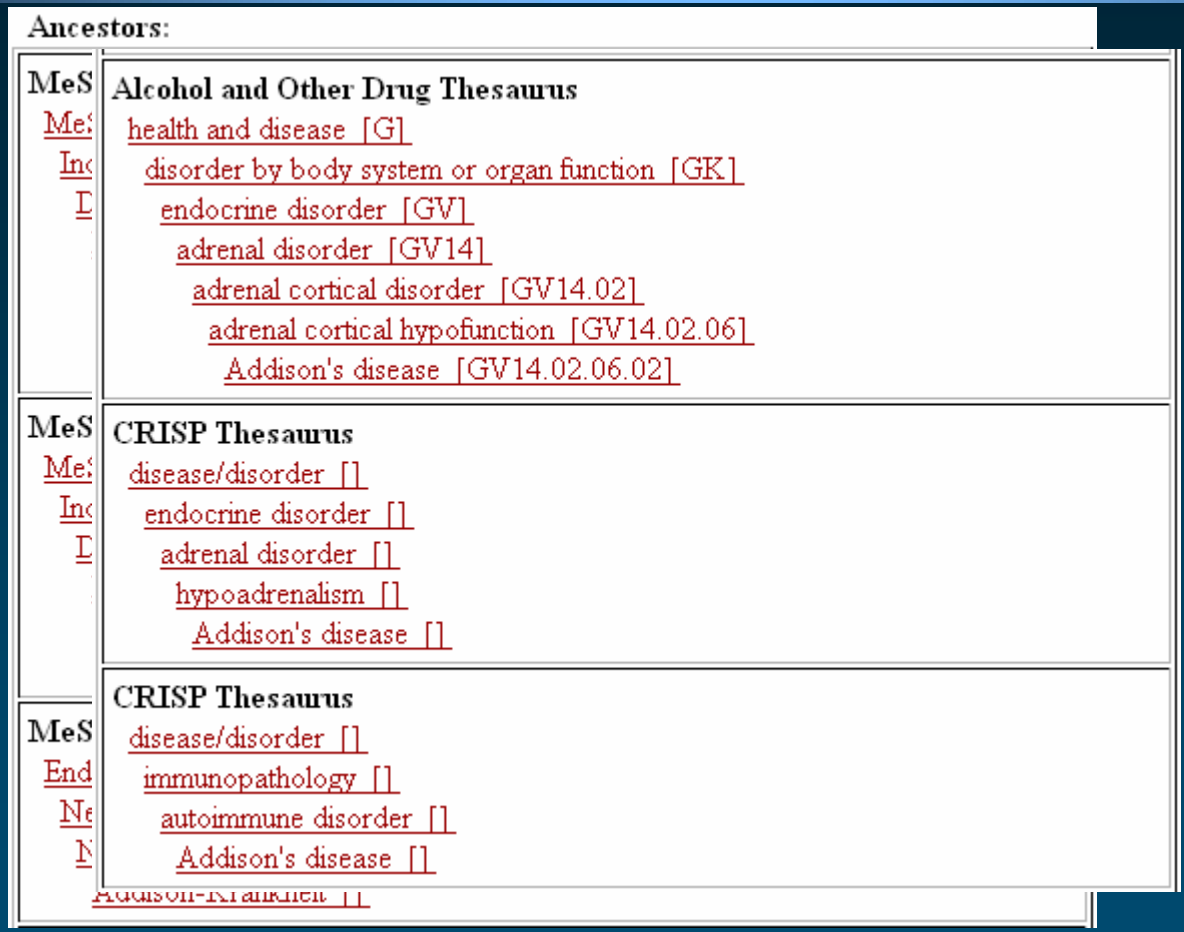


http://umlsks.nlm.nih.gov/

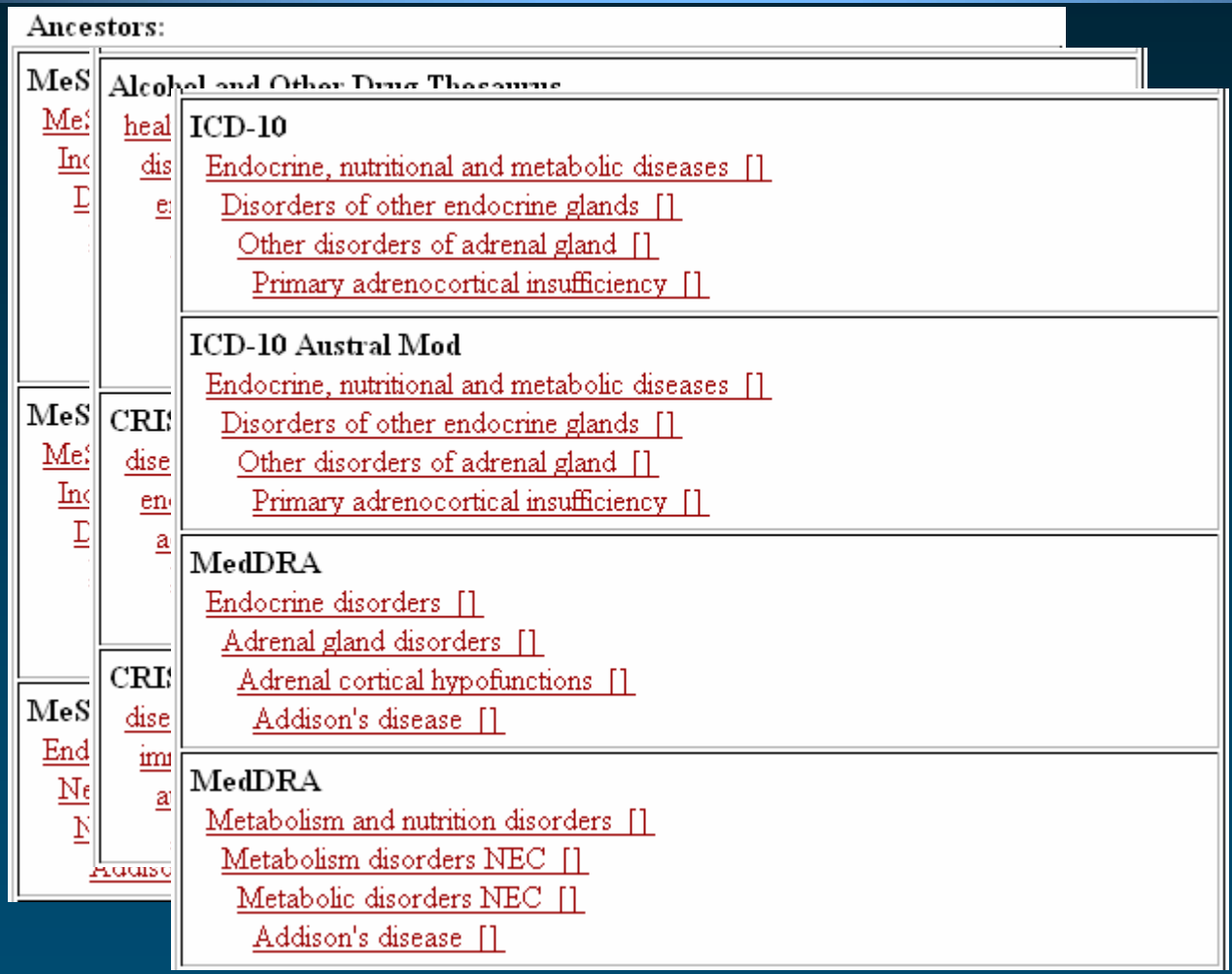
Addison's disease in UMLSKS (1)



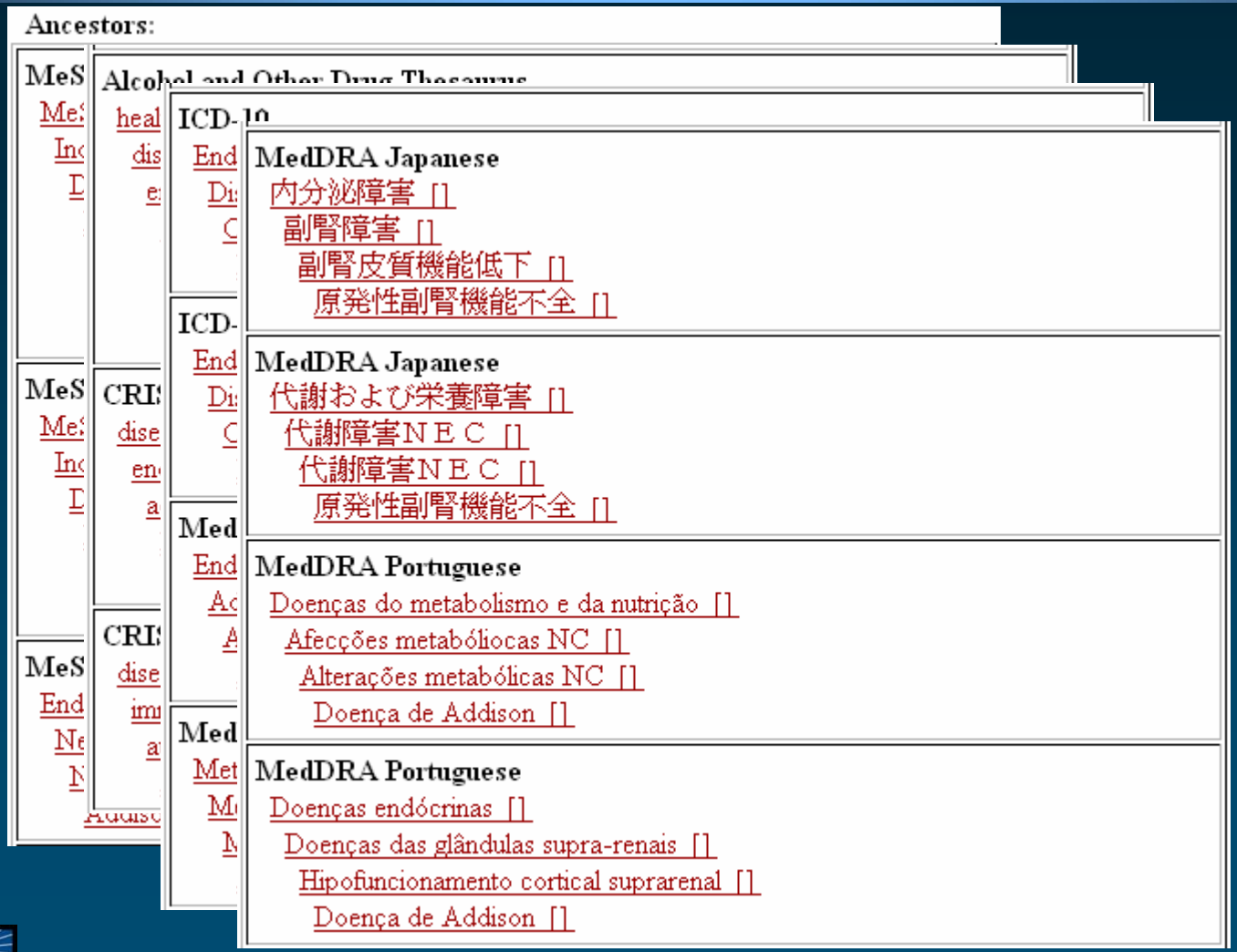
Addison's disease in UMLSKS (2)



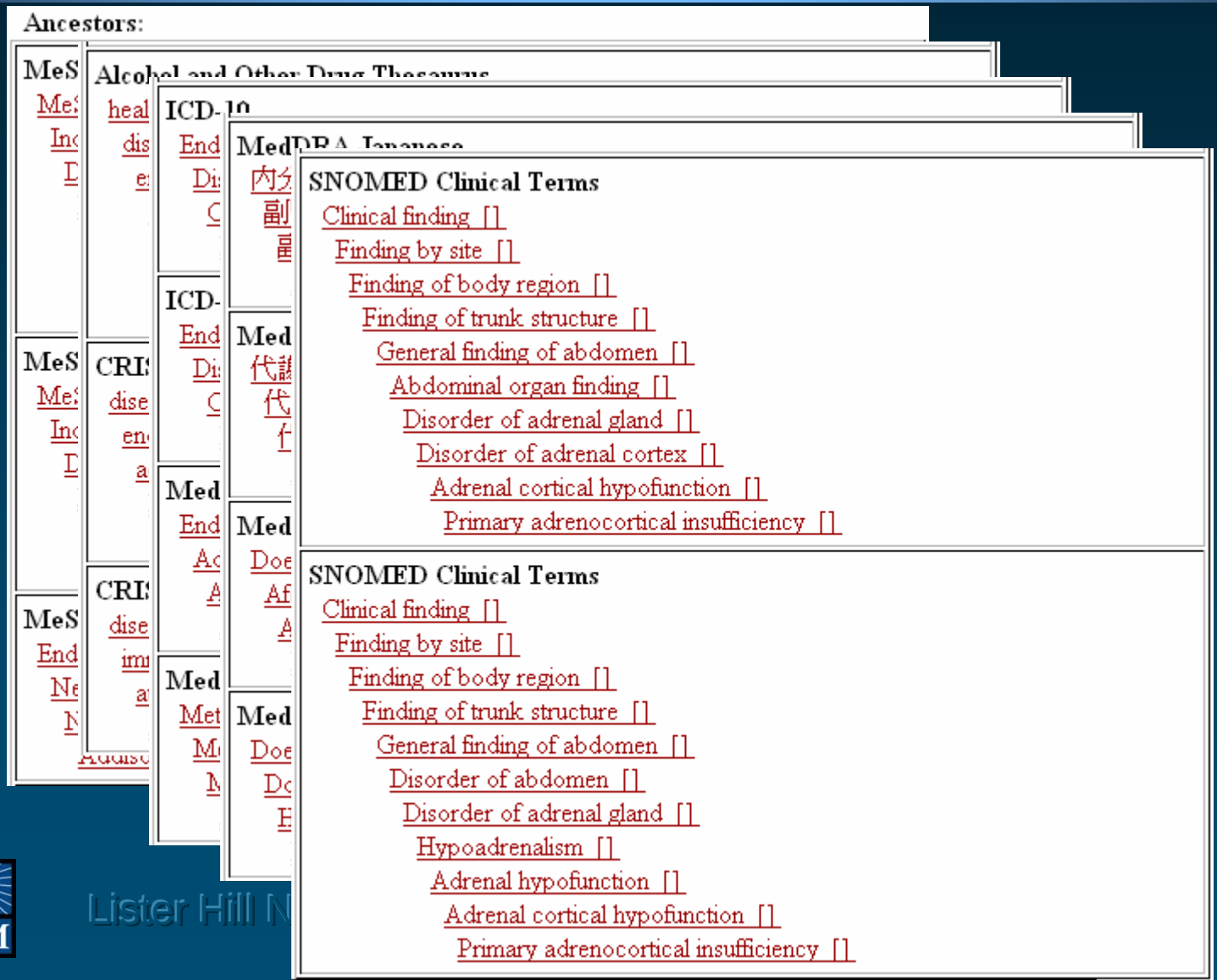
Addison's disease in UMLSKS (3)



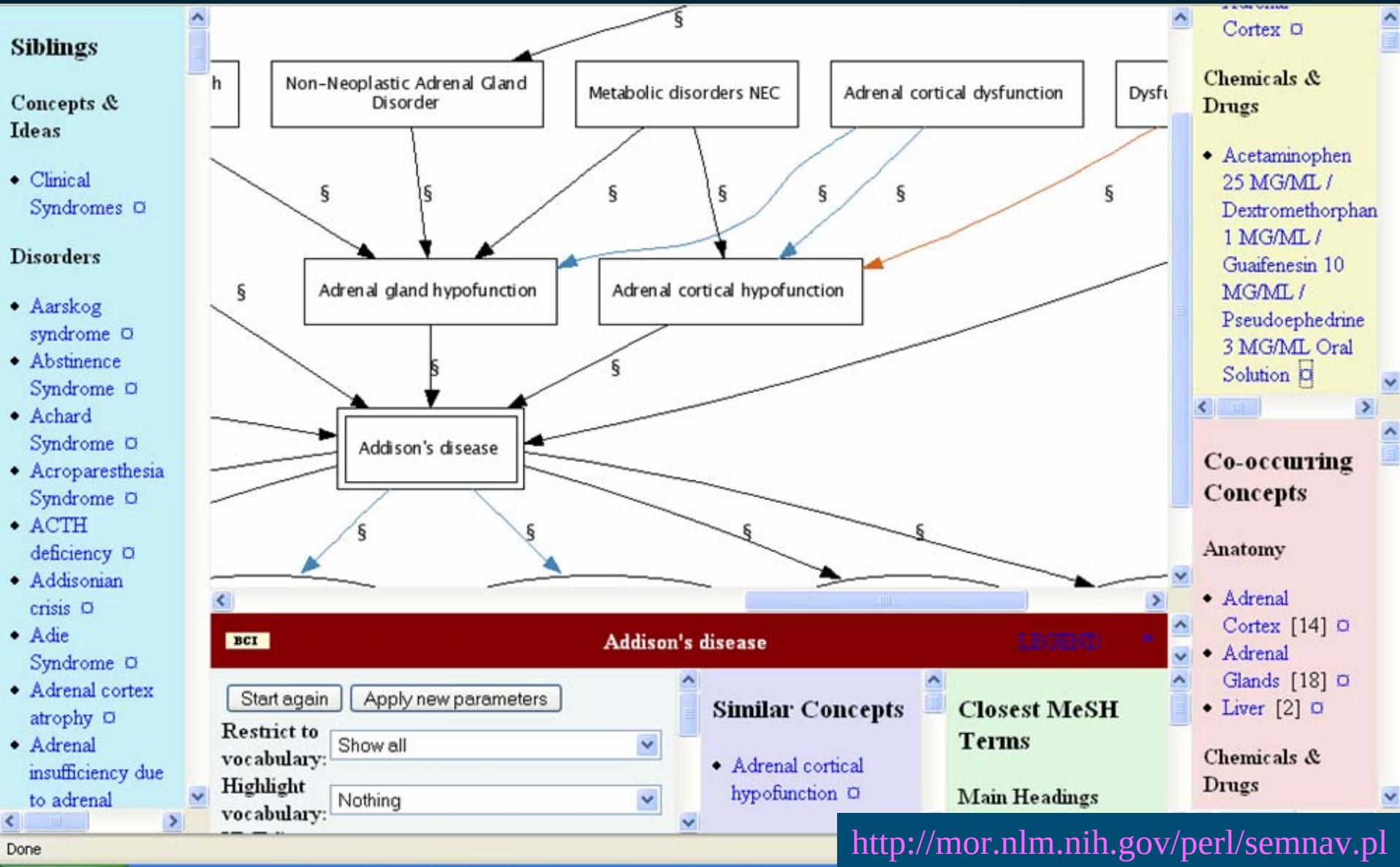
Addison's disease in UMLSKS (4)



Addison's disease in UMLSKS (5)

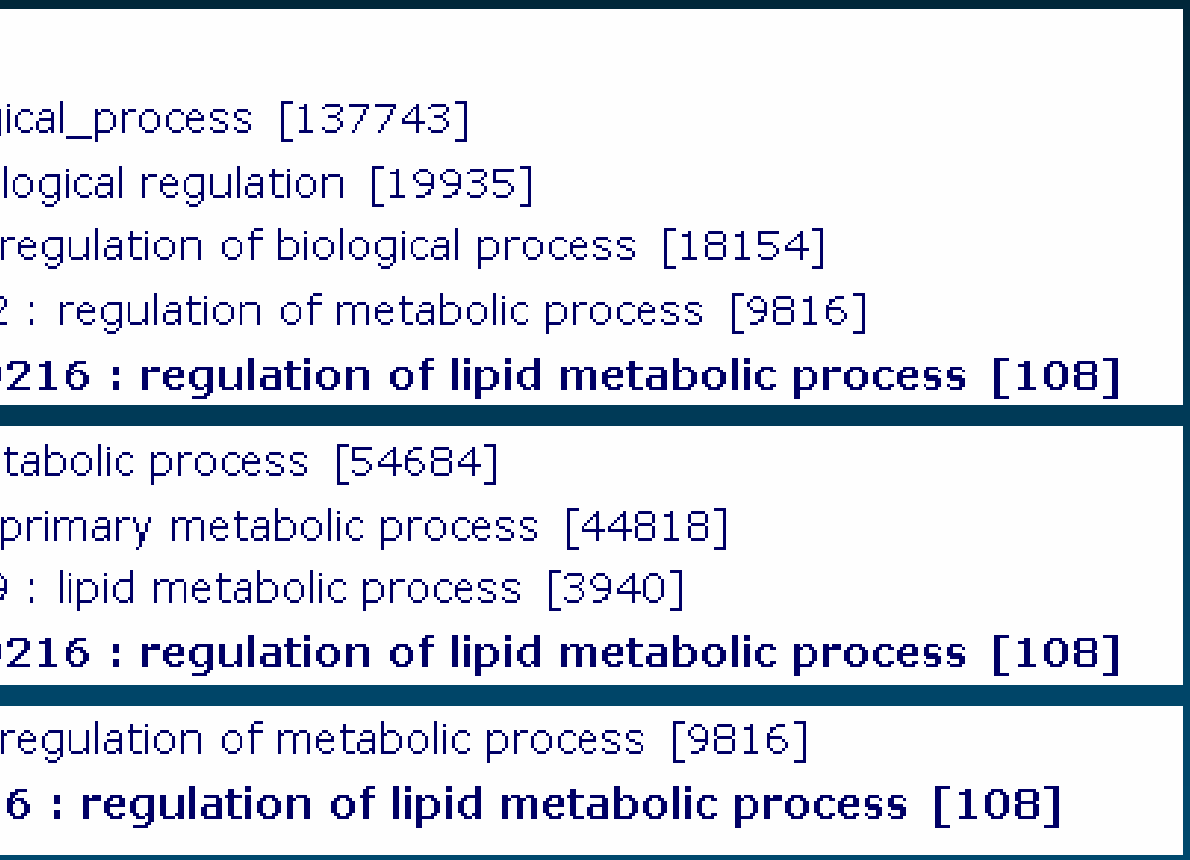


UMLS Semantic Navigator



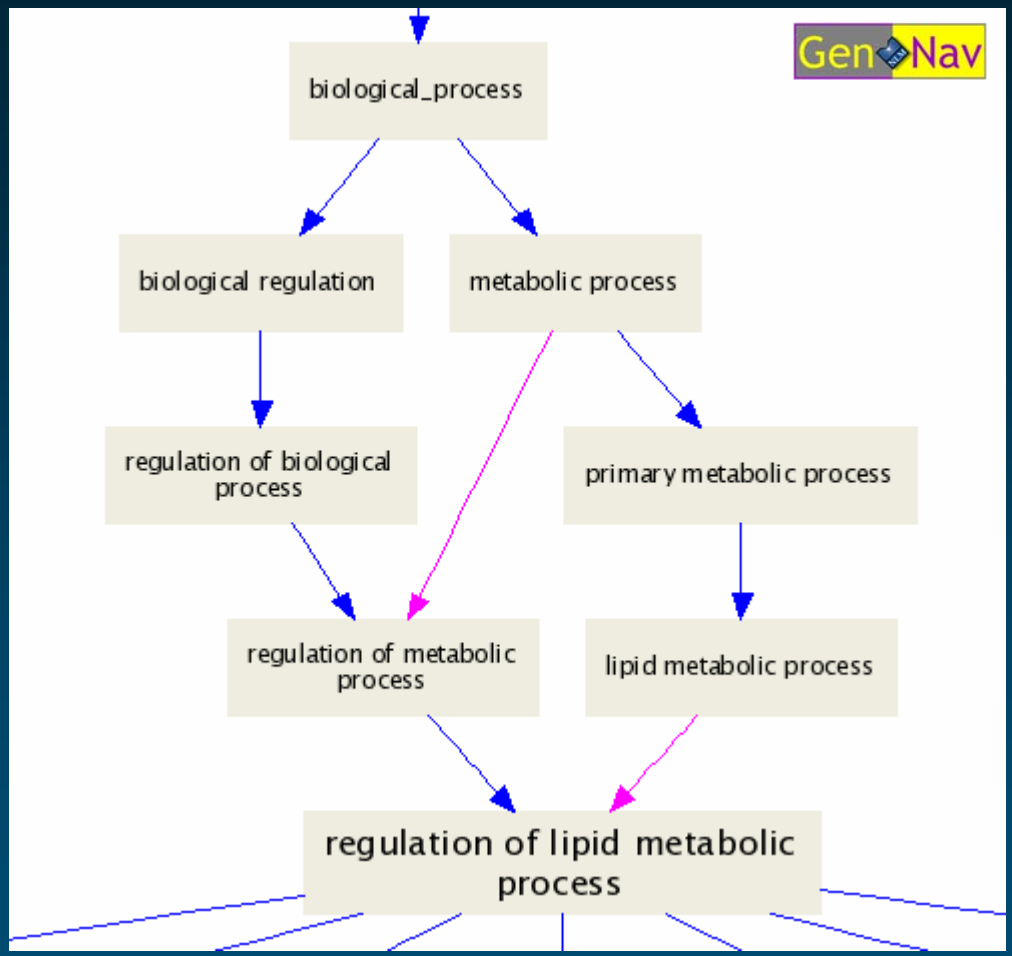
http://mor.nlm.nih.gov/perl/semnav.pl

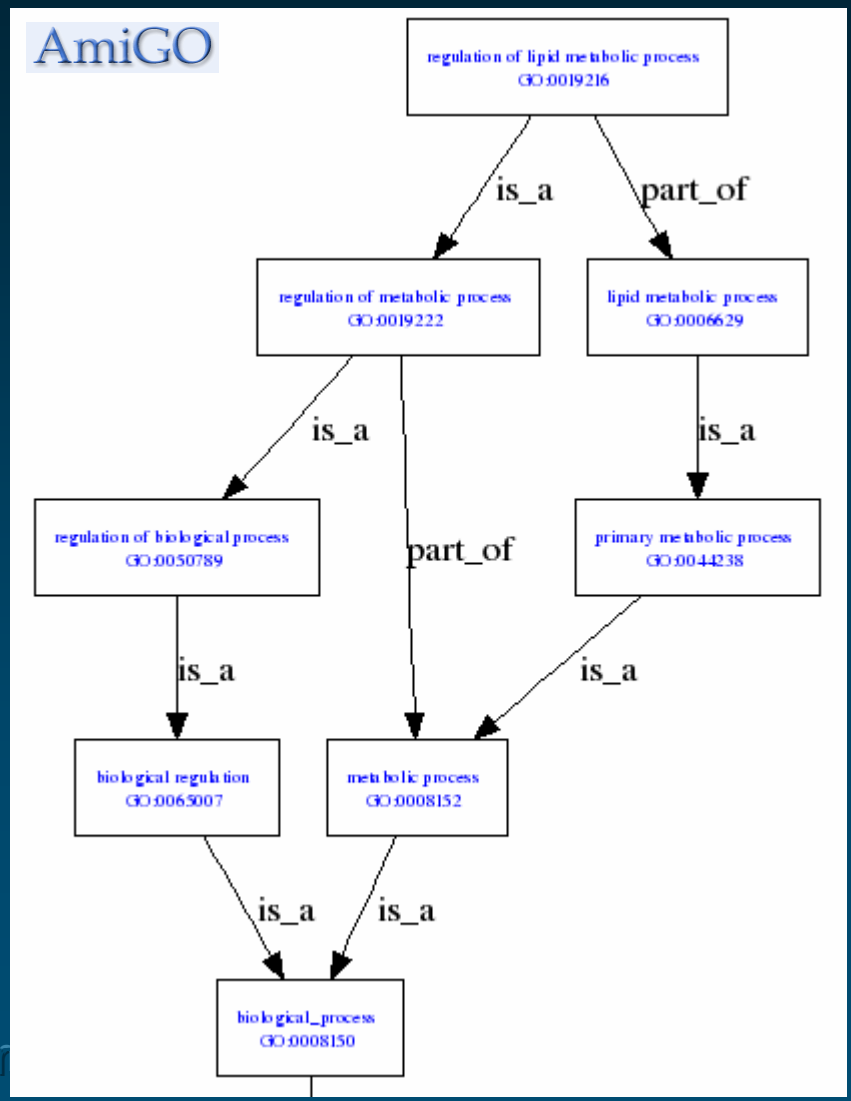
AmiGO



http://amigo.geneontology.org/cgi-bin/amigo/go.cgi

GenNav





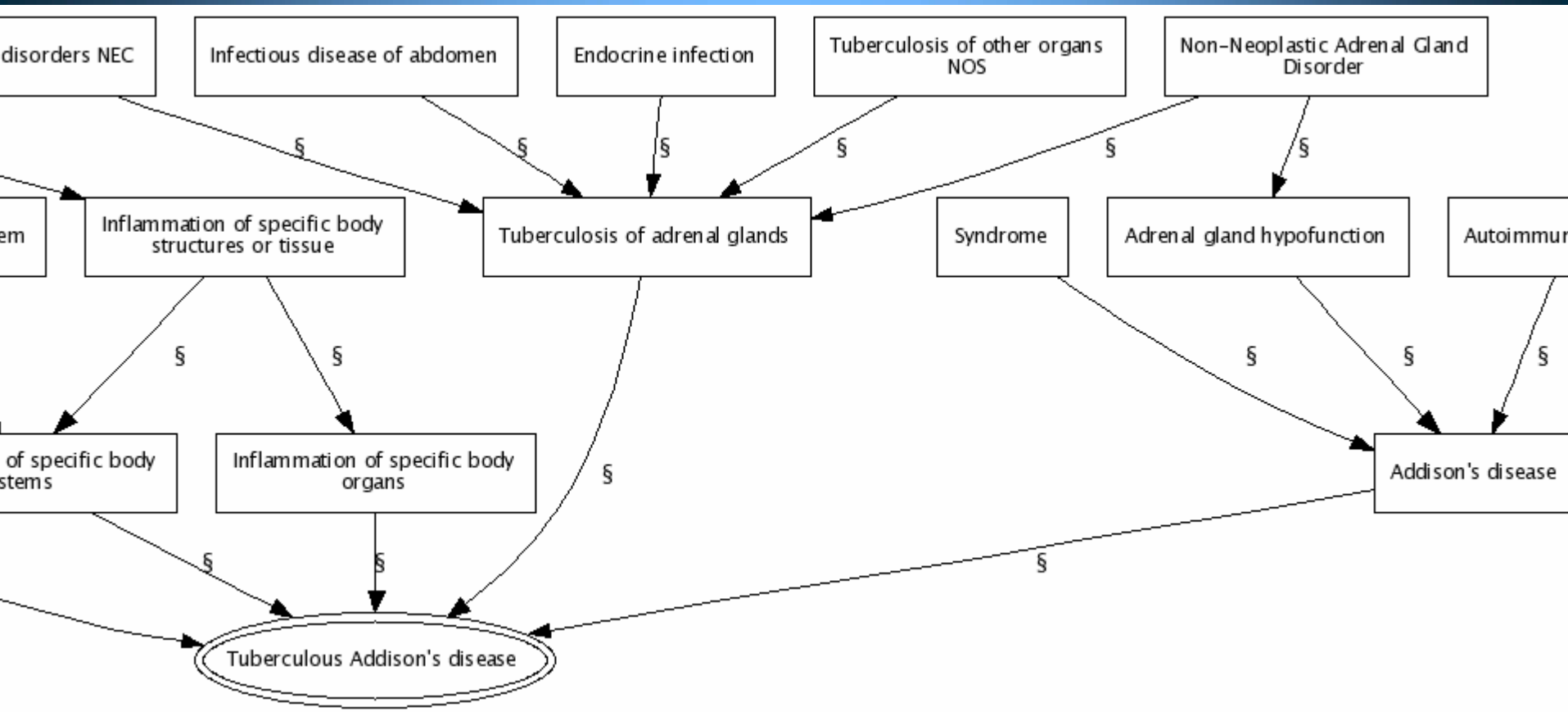
http://mor.nlm.nih.gov/perl/gennav.pl

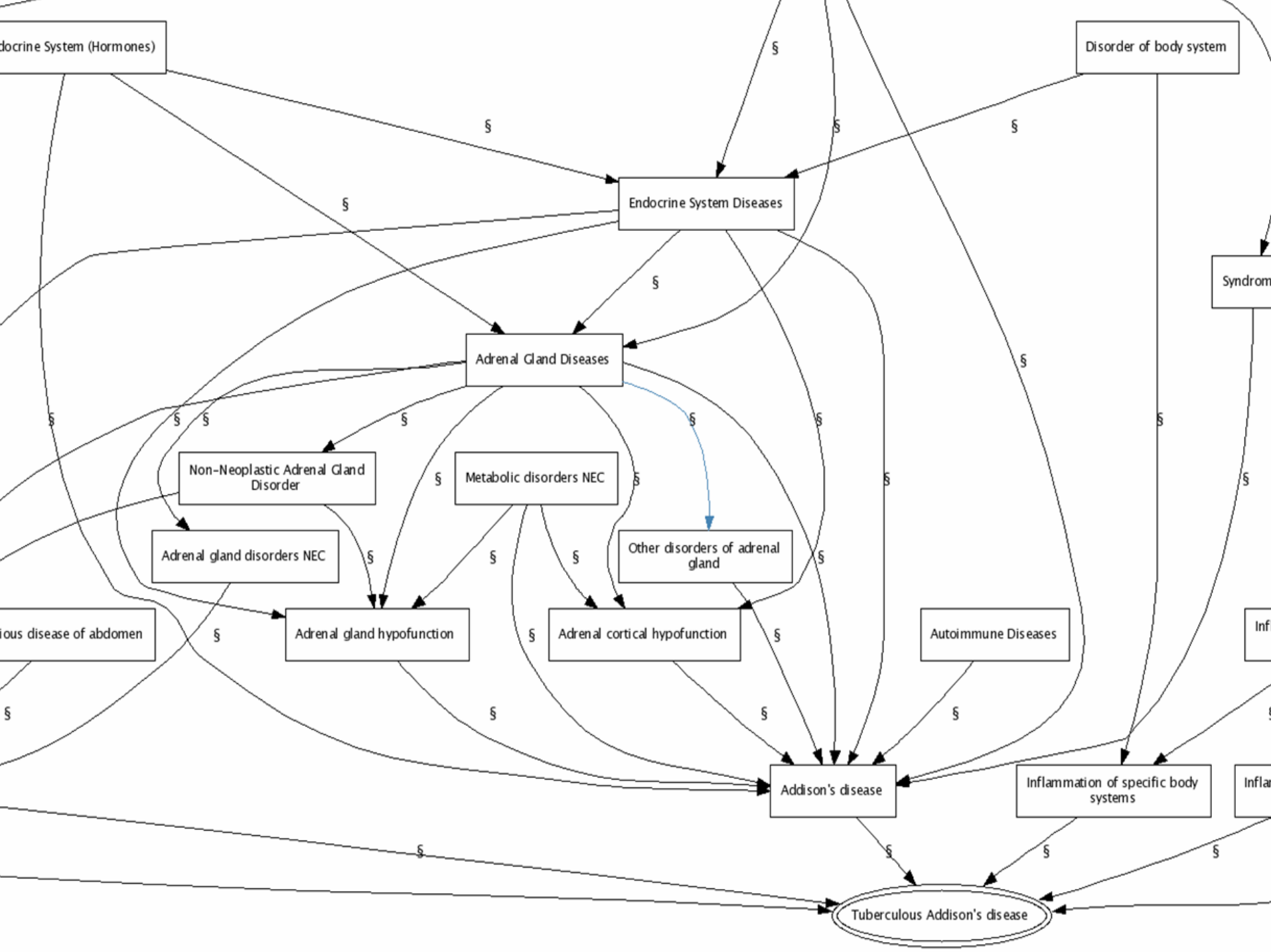
Lister Hill National Center for Bion

Semantics of the UMLS graph

Issues and challenges

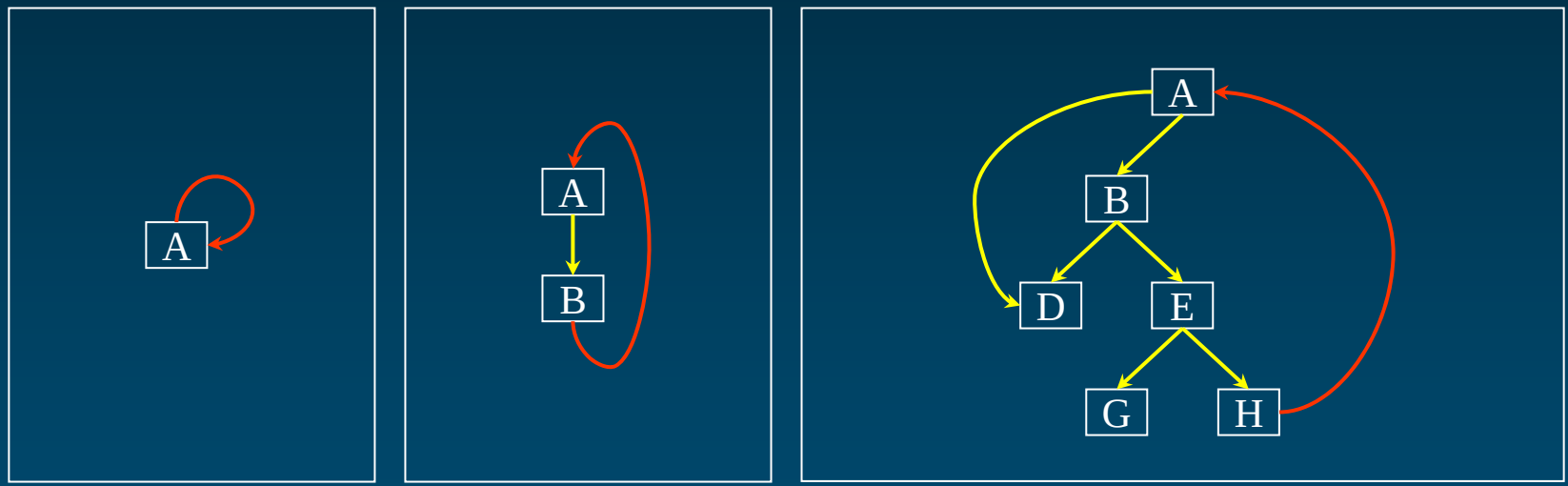
Visualization of large graphs




docrine System (Hormones)
Disorder of body system
Endocrine System Diseases
Syndrom
Adrenal Gland Diseases
Non-Neoplastic Adrenal Gland Disorder
Metabolic disorders NEC
Adrenal gland disorders NEC
Other disorders of adrenal gland
ious disease of abdomen
Adrenal gland hypofunction
Adrenal cortical hypofunction
Autoimmune Diseases
Addison's disease
Inflammation of specific body systems
Tuberculous Addison's disease

Acyclicity

"back edge" from a child concept to a parent concept



Reflexive
13,000

Direct
1800

Indirect
120

Underspecification of relationships

- Relationship “attribute” not always present
- Relations used to create hierarchies vs. hierachical relations

Which tasks?

- Information integration
- Mapping

- Depending on the degree of human involvement
  - Hypothesis generation / validation
  - Knowledge discovery
  - Automated reasoning
- Knowledge standardization
  - Common format
  - Common semantics

Which formalisms?

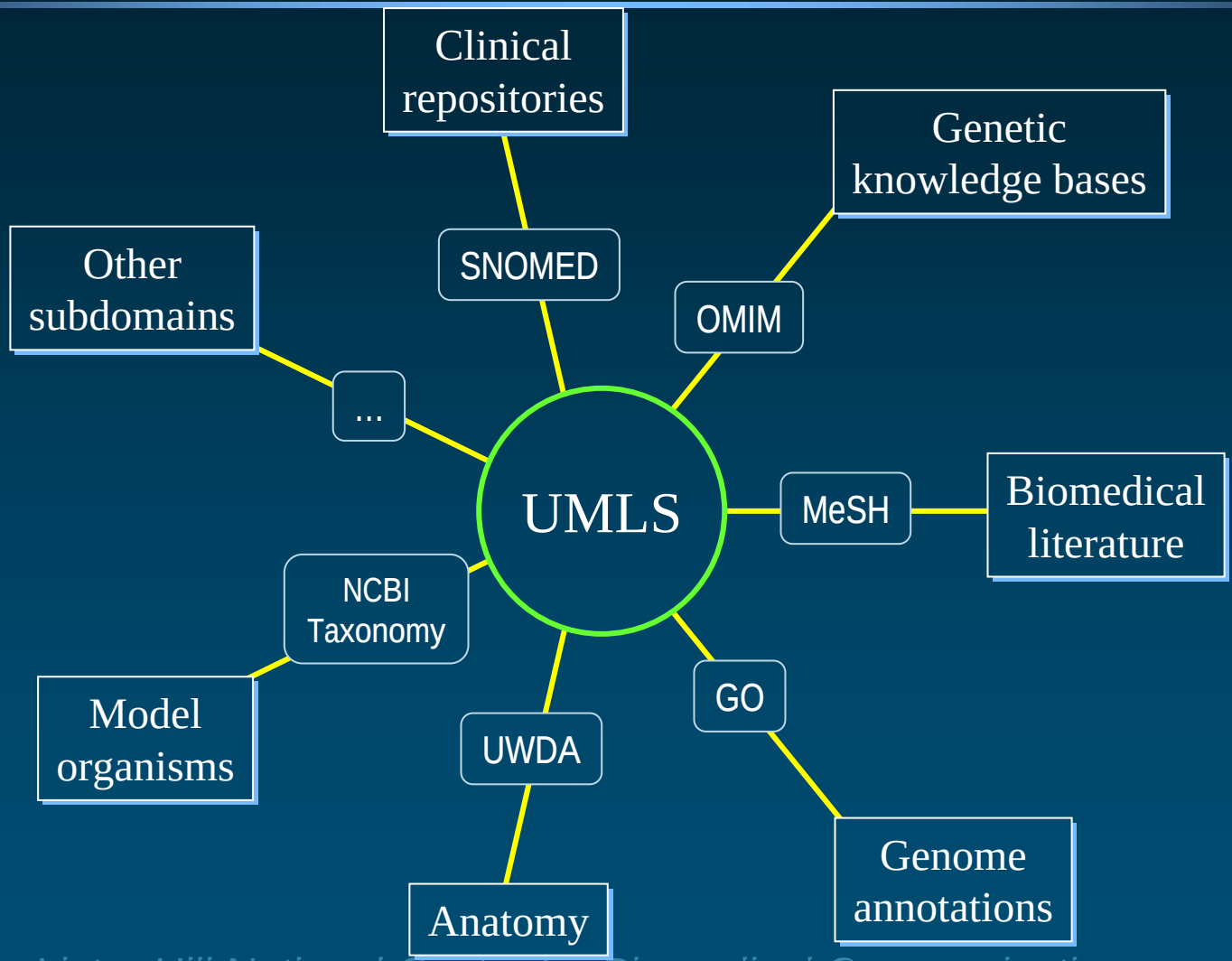
- SKOS – Thesaurus
  - Simple Knowledge Organization Schema
- RDF – Concept-Relationship-Concept triples
  - Resource Description Framework
- Description Logics / Frames
  - OWL Web Ontology Language
  - Protégé (frames / OWL)
  - OBO Open Biomedical Ontology
- Rule languages
- Formal logic

Which identifiers?

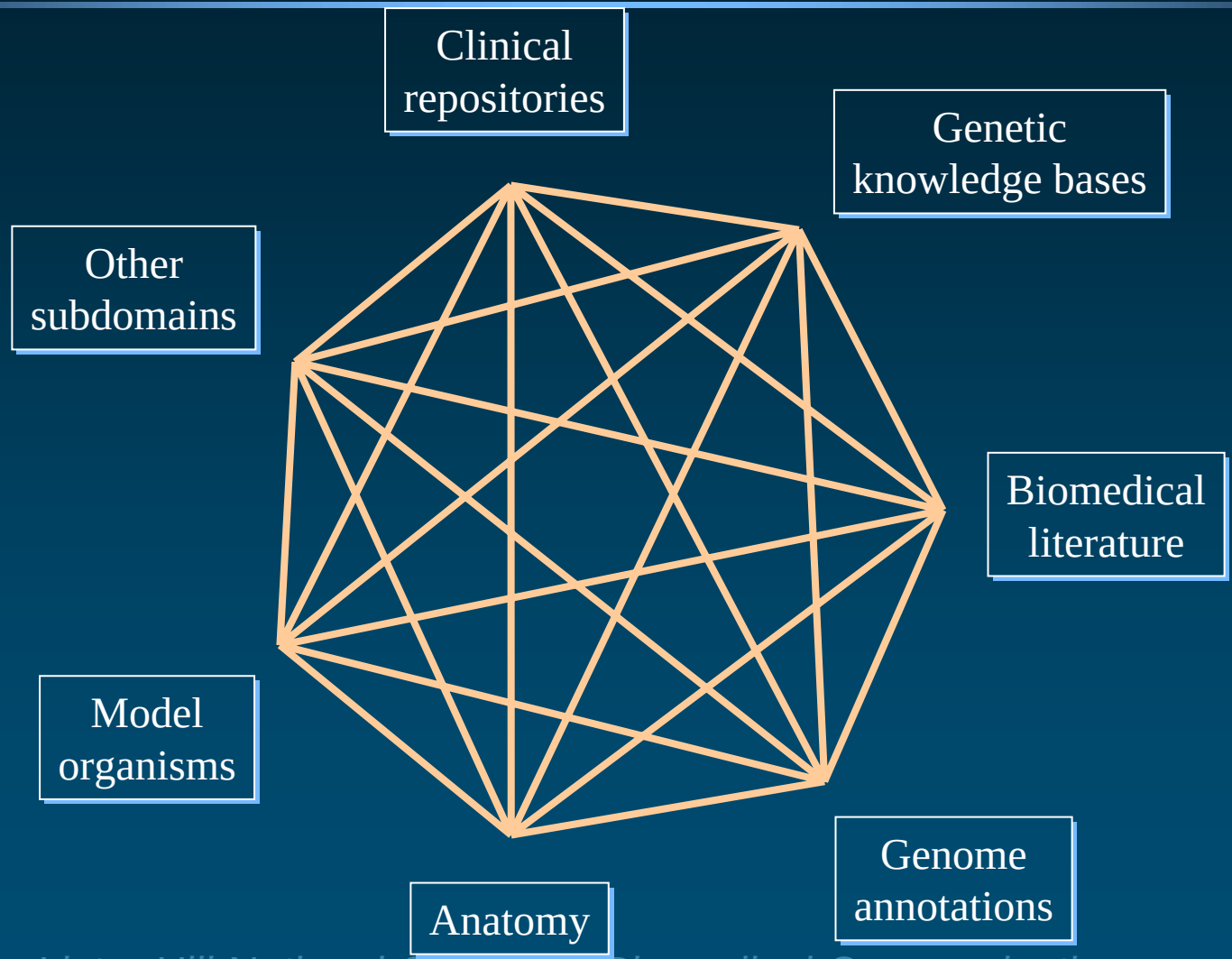
- For concepts
  - Namespaces, ontologies, knowledge bases
    - OBO – Open Biomedical Ontologies
    - UMLS – Unified Medical Language System
    - NCBI Entrez (Entrez Gene, GenBank, UniGene, …)
  - Mappings across information sources
- For relationships

Conclusions

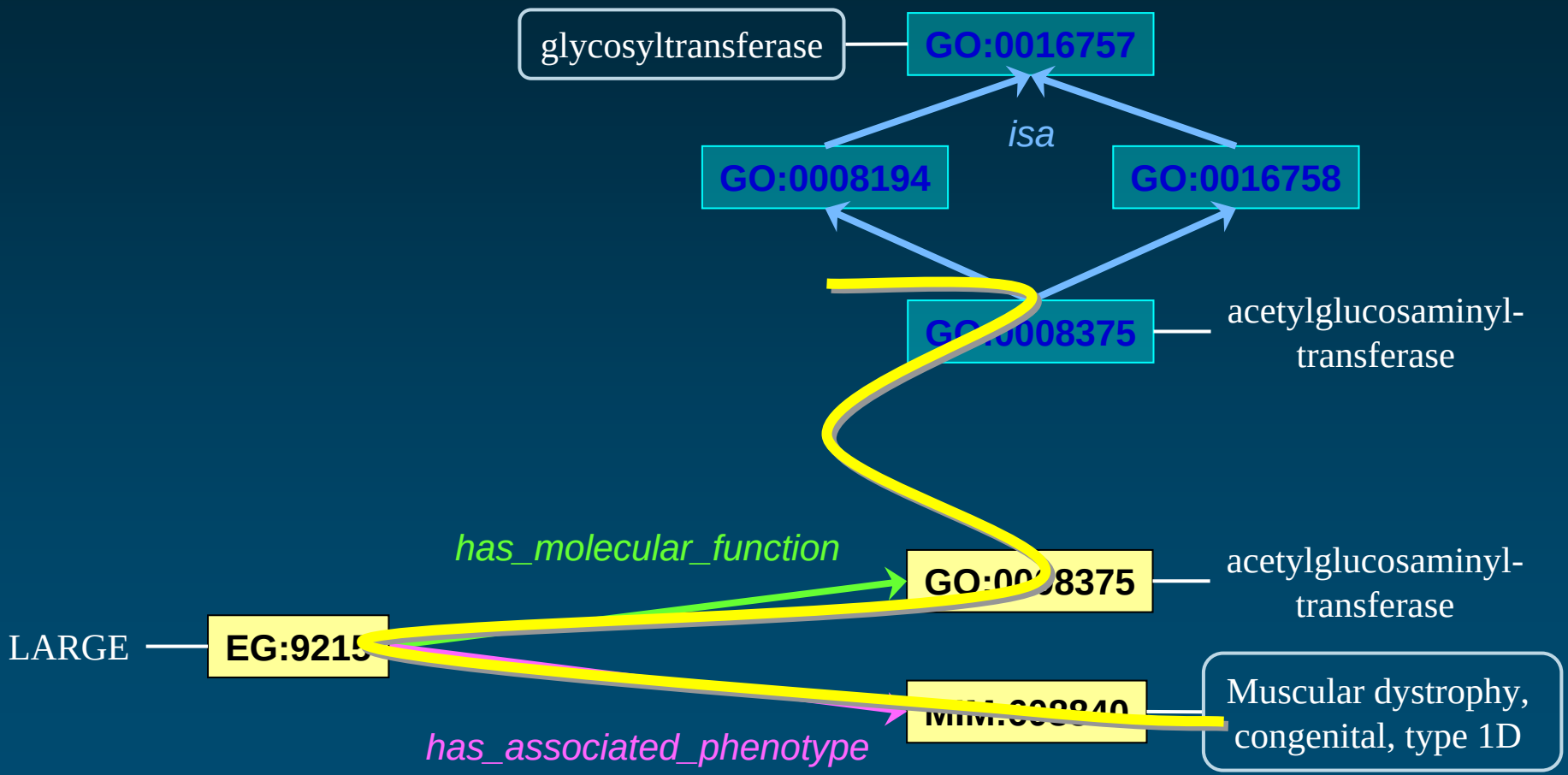
Integrating subdomains

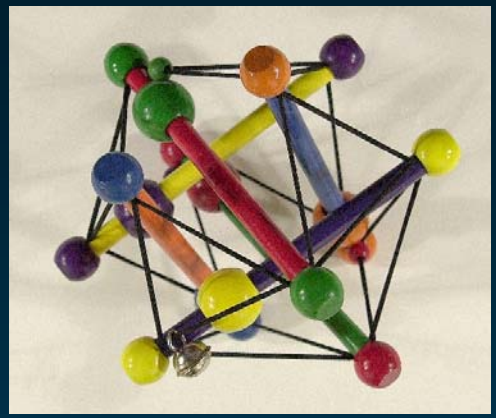


Integrating subdomains



From *glycosyltransferase* to *congenital muscular dystrophy*





Medical Ontology Research

Contact: olivier@nlm.nih.gov
Web: mor.nlm.nih.gov

NATIONAL LIBRARY OF MEDICINE

Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA

UMLS References

- UMLS
  umlsinfo.nlm.nih.gov
- UMLS browsers
  (free, but UMLS license required)
  - Knowledge Source Server: **umlsks.nlm.nih.gov**
  - Semantic Navigator:
    http://mor.nlm.nih.gov/perl/semnav.pl
  - RRF browser
    (standalone application distributed with the UMLS)

UMLS References

- Gentle introduction
  - Bodenreider O. (2004). The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Research*; D267-D270. http://mor.nlm.nih.gov/pubs/pdf/2004-nar-ob.pdf
- Seminal paper
  - Lindberg, D. A., Humphreys, B. L., & McCray, A. T. (1993). The Unified Medical Language System. *Methods Inf Med*, *32*(4), 281-91.

Biomedical information integration through RDF

- Biomedical perspective
  - Sahoo S, Zeng K, Bodenreider O, Sheth AP. (2007). From "glycosyltransferase" to "congenital muscular dystrophy": Integrating knowledge from NCBI Entrez Gene and the Gene Ontology. *Proceedings of Medinfo (in press).* http://mor.nlm.nih.gov/pubs/pdf/2007-medinfo-ss.pdf
- Semantic Web perspective
  - Sahoo S, Zeng K, Bodenreider O, Sheth AP. (2007). An experiment in integrating large biomedical knowledge resources with RDF: Application to associating genotype and phenotype information. *Proceedings of the workshop on Health Care and Life Sciences Data Integration for the Semantic Web at the 16th International World Wide Web Conference (WWW2007) (in press).* http://mor.nlm.nih.gov/pubs/pdf/2007-www_hcls-ss.pdf