University of Missouri-Columbia
Department of Computer Science
November 30, 2006

Biomedical Ontologies

What are they good for?

Olivier Bodenreider

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

Outline

- Biomedical ontologies
- Applications
- Some issues for discussion
- Current trends and future directions

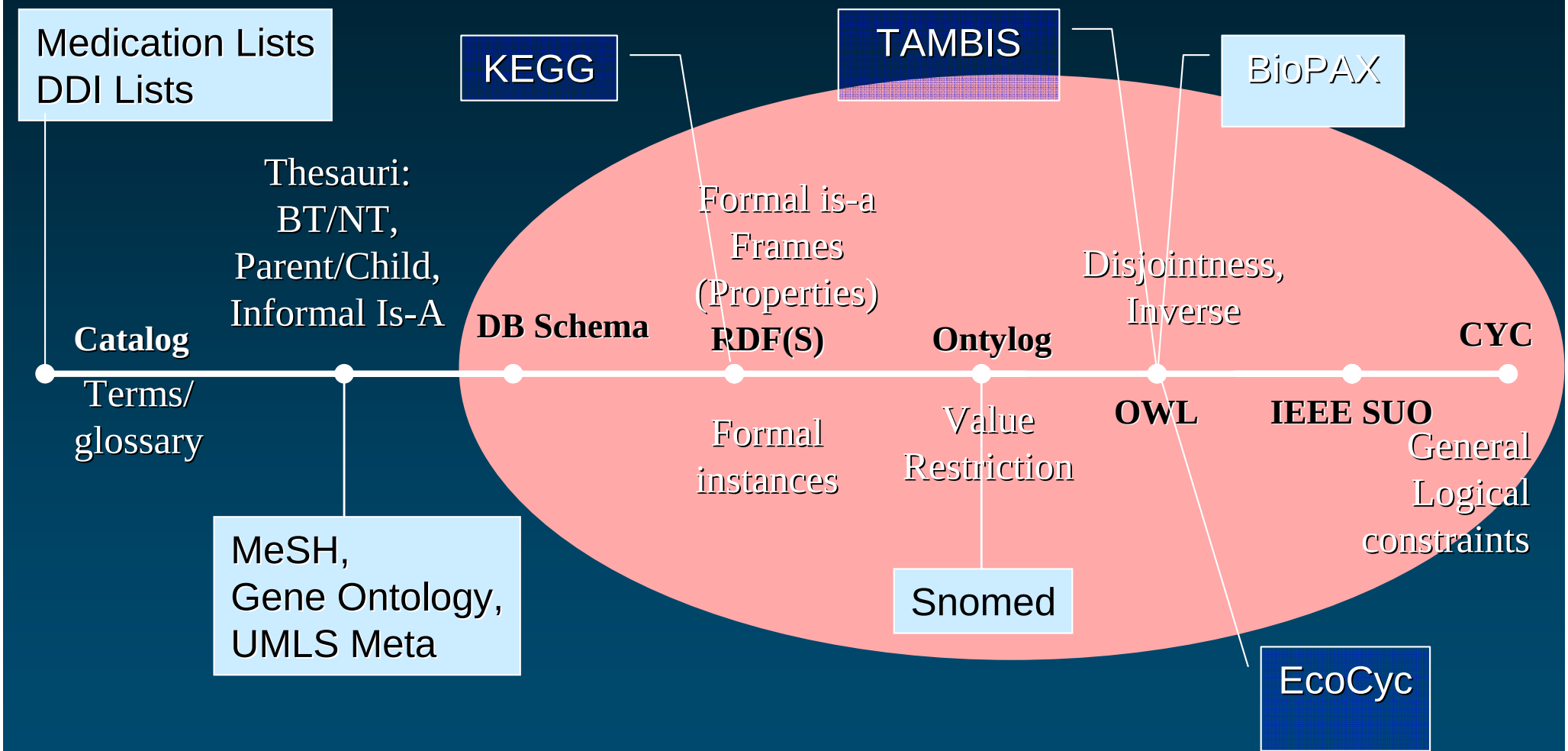
Biomedical ontologies

Ontologies

- Formal representation of a domain modeling the things in that domain and the relationships between those things

- A set of logical axioms designed to account for the intended meaning of a vocabulary [Guarino, FOIS 1998]

The Knowledge Semantics Continuum



Simple Terminologies ← → Expressive Ontologies

Ontology Dimensions based on McGuinness and Finin

Biomedical ontologies (and terminologies)

- The OBO family
  - Ontologies and terminologies
  - Gene Ontology
  - Mostly biological ontologies
- UMLS
  - Ontologies and terminologies
  - MeSH, SNOMED CT
  - Mostly clinical ontologies

Open Biological Ontologies

- Extended family of the Gene Ontology (GO)
- Collaborative development
  - http://obo.sourceforge.net/
  - 5 inclusion criteria
- National Center for Biomedical Ontology
  - http://bioontology.org/
- OBO Foundry
  - http://obofoundry.org/
  - Promote best practices in ontology development

Relation ontology

- Defines 14 core relations (isa, part of, derives from, has agent, …)
  - Textual definition
  - Formal definition
  - Comments
- To be used by other OBO ontologies

[Smith et al., Genome Biology, 2005, 6:R46]

Open Biological Ontologies (OBO)

Main Criteria Ontologies Browse Project CVS Subscribe Contact

OBO Ontology Browser

Browse the tree by clicking on the category names; click on an ontology name to view more information on it.

- anatomy
- animal natural history and life history
- chemical
- development
- ethology
- evidence codes
- experimental conditions
- genomic and proteomic
- metabolomics
- OBO relationship types
- phenotype
- taxonomic classification
- vocabularies

http://obo.sourceforge.net/

OBO format

- Used to represent many ontologies in the OBO family (Open Biological Ontologies)

http://www.godatabase.org/dev/doc/obo_format_spec.html

- Essentially a subset of OWL DL

```
[Term]
id: GO:0019563
name: glycerol catabolism
namespace: biological_process
def: "The chemical reactions and pathways resulting in the breakdown of glycerol …
subset: gosubset_prok
exact_synonym: "glycerol breakdown" []
exact_synonym: "glycerol degradation" []
xref_analog: MetaCyc:PWY0-381
is_a: GO:0006071 ! glycerol metabolism
is_a: GO:0046174 ! polyol catabolism
```

What does UMLS stand for?

- Unified
- Medical
- Language
- System

UMLS®
Unified Medical Language System®
UMLS Metathesaurus®

Motivation

- Started in 1986
- National Library of Medicine
- “Long-term R&D project”

«[...] the UMLS project is an effort to overcome two significant barriers to effective retrieval of machine-readable information.

- The first is the variety of ways the same concepts are expressed in different machine-readable sources and by different people.
- The second is the distribution of useful information among many disparate databases and systems.»

Unified Medical Language System

- SPECIALIST Lexicon
  - 200,000 lexical items
  - Part of speech and variant information
- Metathesaurus
  - 5M names from over 100 terminologies
  - 1M concepts
  - 16M relations
- Semantic Network
  - 135 high-level categories
  - 7000 relations among them

Lexical resources

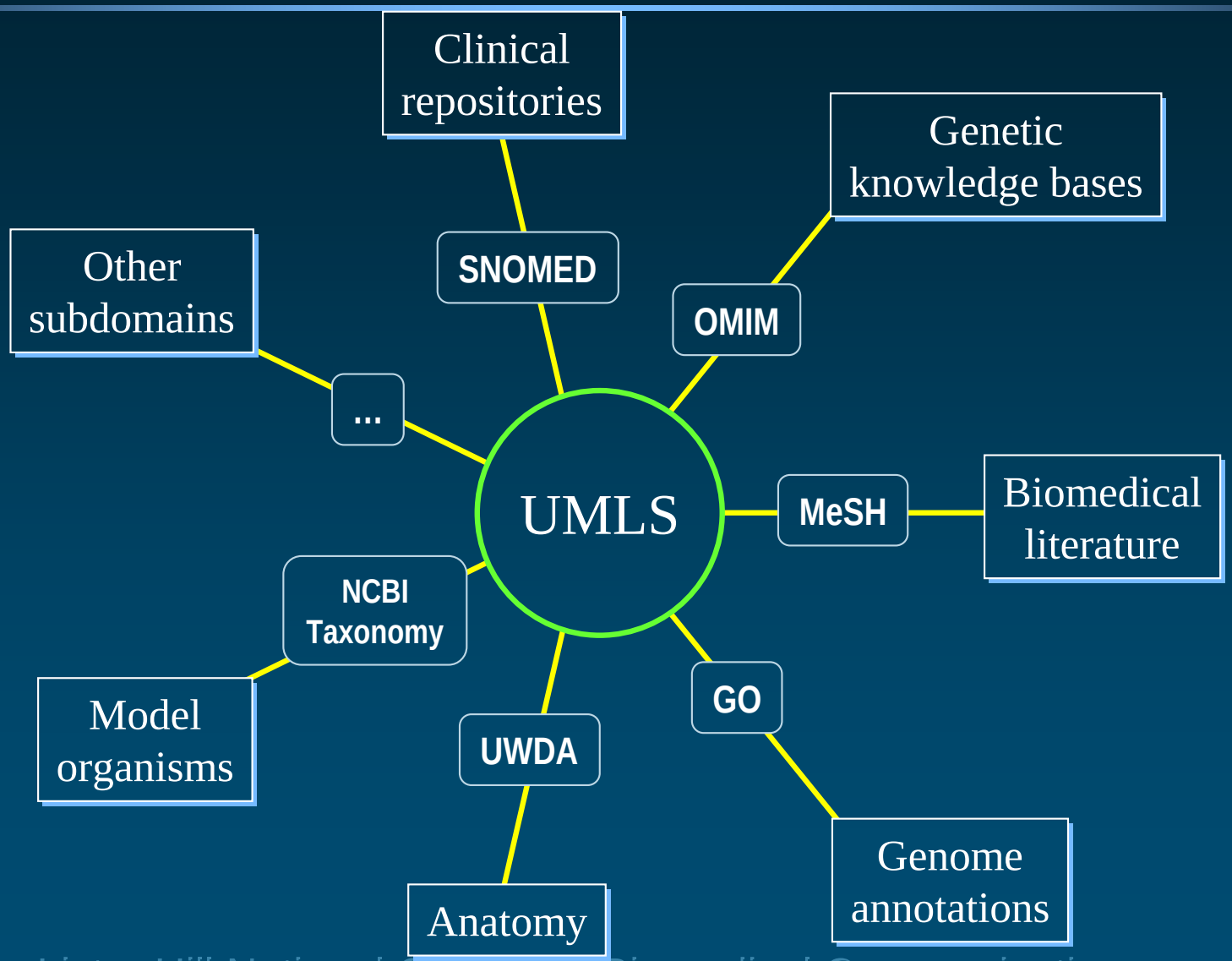
Terminological resources

Ontological resources

Source Vocabularies (2006AD)

- 138 source vocabularies
  - 17 languages
- Broad coverage of biomedicine
  - 5.4M names
  - 1.4M concepts
  - 16M relations
- Common presentation

Integrating subdomains



Addison's Disease: Concept

Disease or Syndrome

Addison's Disease

ADRENAL INSUFFICIENCY (ADDISON'S DISEASE)
ADRENOCORTICAL INSUFFICIENCY, PRIMARY FAILURE
Addison melanoderma
Melasma addisonii
Primary adrenal deficiency
Asthenia pigmentosa
Bronzed disease
Insufficiency, adrenal primary
Primary adrenocortical insufficiency
Addison's, disease

MALADIE D'ADDISON - French
Addison-Krankheit - German
Morbo di Addison - Italian
DOENCA DE ADDISON - Portuguese
ADDISONOVA BOLEZN' - Russian
ENFERMEDAD DE ADDISON - Spanish

SNOMED
MeSH
AOD
Read Codes
…

C0001403

A disease characterized by hypotension, weight loss, anorexia, weakness, and sometimes a bronze-like melanotic hyperpigmentation of the skin. It is due to tuberculosis- or autoimmune-induced disease (hypofunction) of the adrenal glands that results in deficiency of aldosterone and cortisol. In the absence of replacement therapy, it is usually fatal.

Metathesaurus Concepts (2006AD)

- Concept (~ 1.4 M) CUI
  - Set of synonymous concept names
- Term (~ 4.8 M) LUI
  - Set of normalized names
- String (~ 5.4 M) SUI
  - Distinct concept name
- Atom (~ 6.5 M) AUI
  - Concept name in a given source

A0000001 headache (source 1)
A0000002 headache (source 2)
S0000001

A0000003 Headache (source 1)
A0000004 Headache (source 2)
S0000002

L0000001

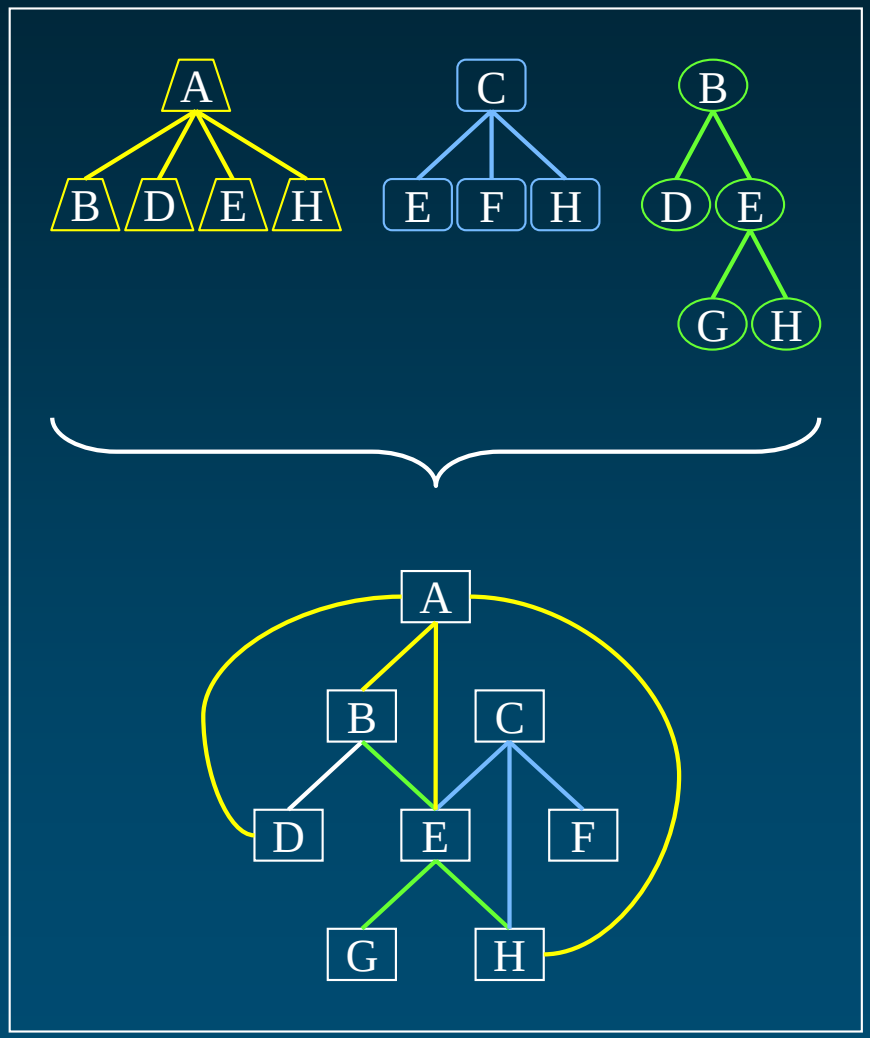
A0000005 Cephalgia (source 1)
S0000003

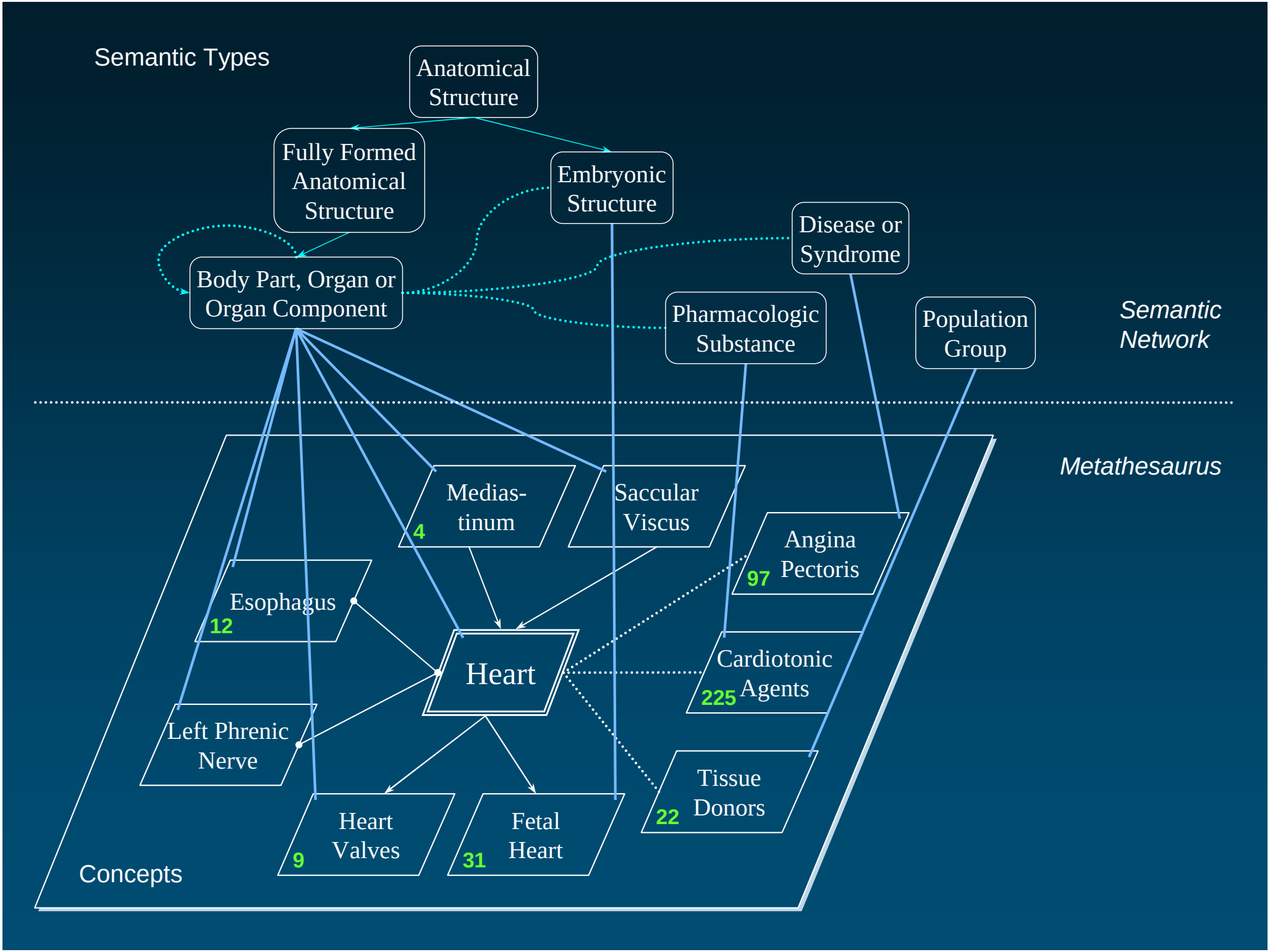
L0000002

C0000001

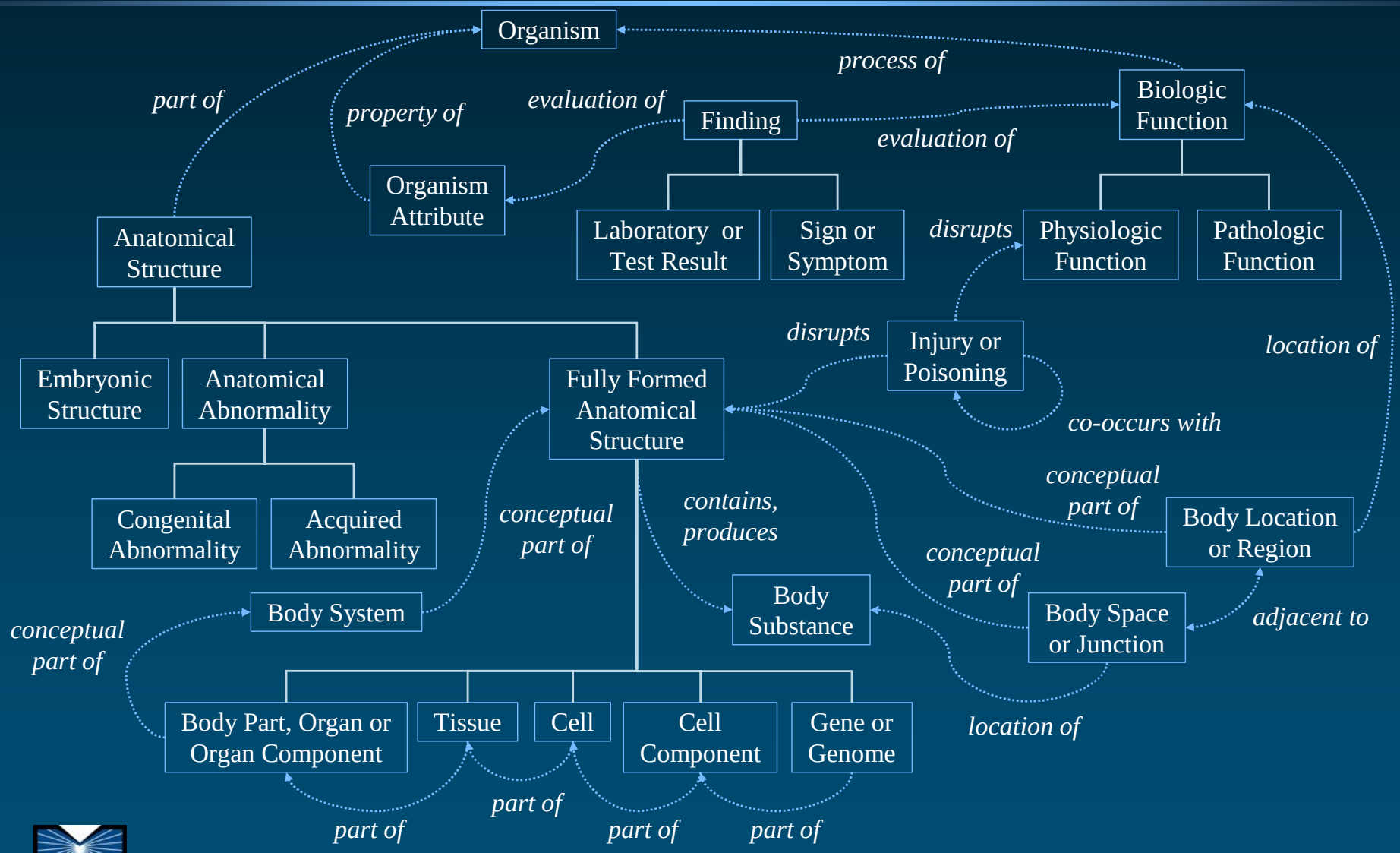
Organize concepts

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

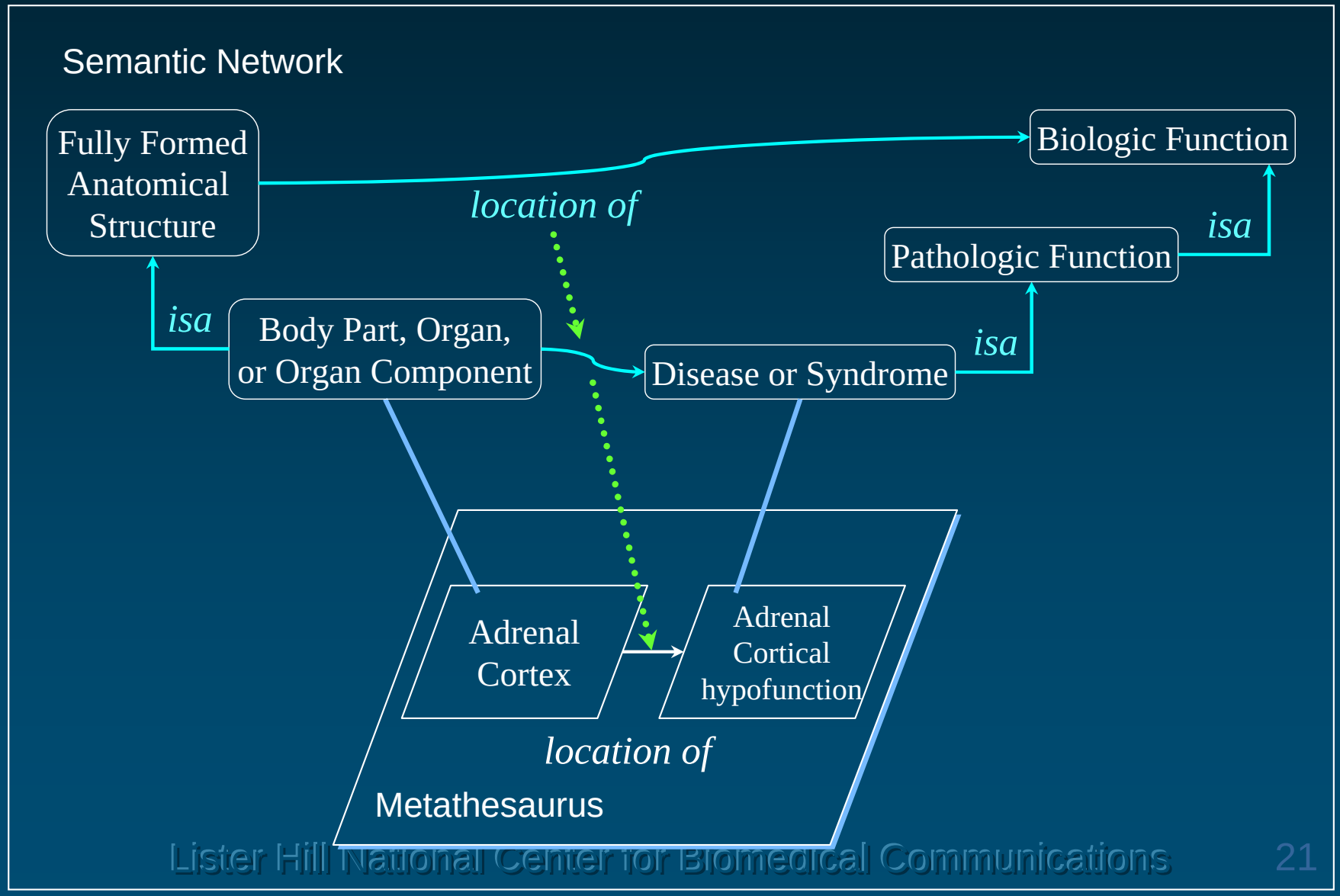



Semantic Types
Anatomical Structure
Fully Formed Anatomical Structure
Embryonic Structure
Disease or Syndrome
Body Part, Organ or Organ Component
Pharmacologic Substance
Population Group
Semantic Network
Metathesaurus
Medias-tinum
4
Saccular Viscus
Angina Pectoris
97
Esophagus
12
Heart
Cardiotonic Agents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

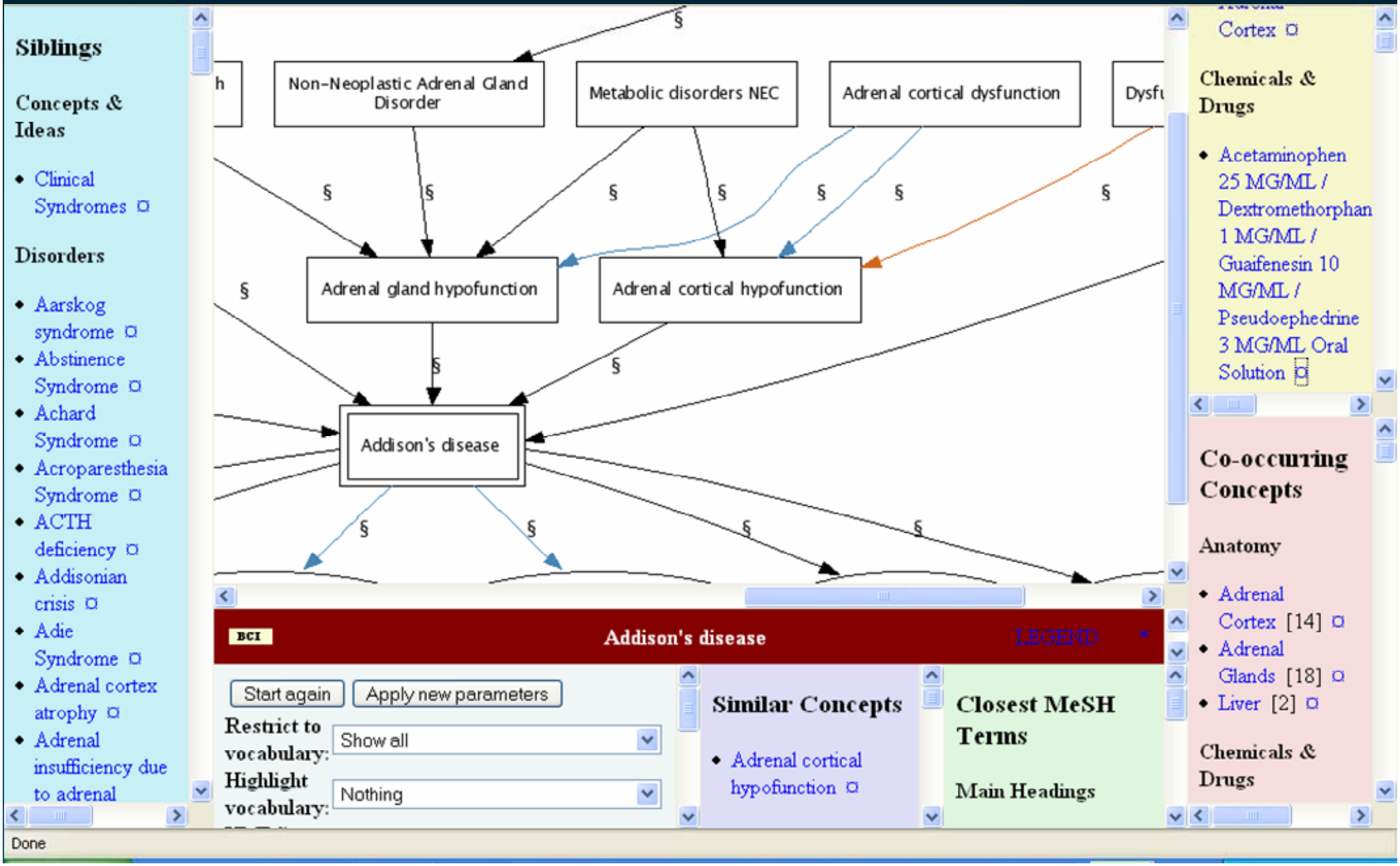
Semantic Network relations



Relationships can inherit semantics



UMLS Semantic Navigator



Terminology vs. ontology

- Terminological resources
  - Collections of terms (e.g., controlled vocabularies)
  - Useful for indexing and annotation
  - MeSH, GO
- Ontological resources
  - Collections of
    - kinds of entities (substances, qualities, processes)
    - relations among them
  - Useful for reasoning
  - UMLS Semantic Network, SNOMED CT

Ontological gradient

Applications

Applications of ontologies

- Information integration
- Support for natural language processing
- Reasoning, automated classification
- Support for knowledge discovery

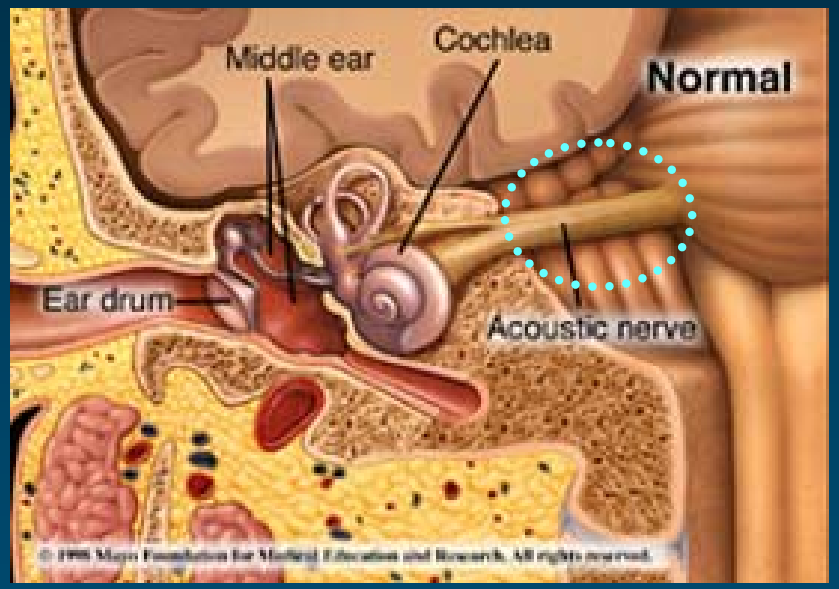
- Support for clinical decision
- …

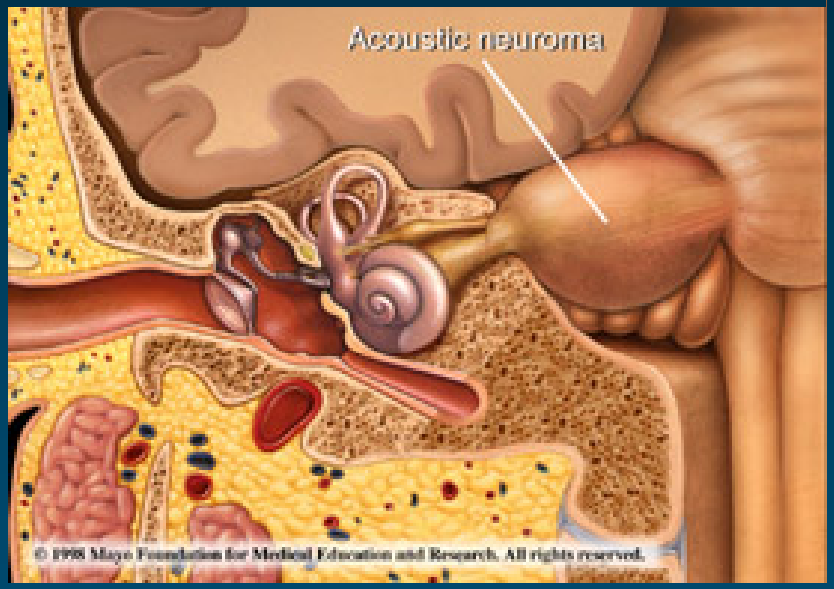
Information integration

NF2 Gene, protein, and disease

Neurofibromatosis 2 is an autosomal dominant disease characterized by tumors called schwannomas involving the acoustic nerve, as well as other features. The disorder is caused by mutations of the *NF2 gene* resulting in absence or inactivation of the protein product. The protein product of NF2 is commonly called *merlin* (but also neurofibromin 2 and schwannomin) and functions as a tumor suppressor.

Schwannoma (acoustic neuroma)





http://www.mayoclinic.com

{UMLS_2003} UMLS® Semantic Navigator ¤ [2.10] - Netscape

{UMLS_2003} UMLS® Semantic Navigator ...

Siblings

Disorders

- Cerebellopontine Angle Acoustic Neuroma ¤
- Diffuse neurofibroma ¤
- Melanocytic Vestibular Schwannoma ¤
- Neurofibromatosis (nonmalignant) ¤
- Neurofibromatosis 1 ¤
- neurofibromatosis 1 and 2 (NF1 and NF2) ¤
- Neurofibromatosis 3 ¤
- Neurofibromatosis type 3 ¤
- NEUROFIBROMATOS TYPE IV, OF RICCARDI ¤
- Neuroma, Acoustic, Unilateral ¤
- Segmental neurofibromatosis ¤

(11 siblings)

[direct children and narrower concepts of direct parents and broader concepts]

Tumor of acoustic vestibular nerve

Benign neoplasm of cranial nerves

Neoplastic Syndromes, Hereditary

Skin tumor of neural c

Neurofibromatosis 2

Neuroma, Acoustic, Bilateral

Schwannoma, Acoustic, Bilateral

Other Related Concepts

Anatomy

- Acoustic Nerve ¤

Chemicals & Drugs

- Neurofibromin 2 ¤

Disorders

- Familial Acoustic Neuromas ¤
- Neoplasm of uncertain behavior NOS ¤
- Neurofibromatoses ¤
- Neurofibromatosis
- Nerve Sheath Tumors [4] ¤
- Nervous System Neoplasms [6] ¤
- Neurilemmoma [35]
- Neurofibromatosis 1 [38] ¤
- Neuroma, Acoustic [26] ¤
- Peripheral Nervous System Diseases [3] ¤
- Peripheral Nervous System Neoplasms [6] ¤
- Postoperative Complications [9] ¤
- Retinal Diseases [6] ¤
- Skin Neoplasms [9] ¤

BCI

Neurofibromatosis 2

LEGEND *

Start again | Apply new parameters

Restrict to vocabulary: Show all

Highlight vocabulary: Nothing

UMLS data: UMLS_2003

Type of hierarchical rel.: All / Parent/Child only / Broader/Narrower only

Similar Concepts

(none)

Allegedly Synonyms

- Neurofibromatosis (nonmalignant) ¤

Closest MeSH Terms

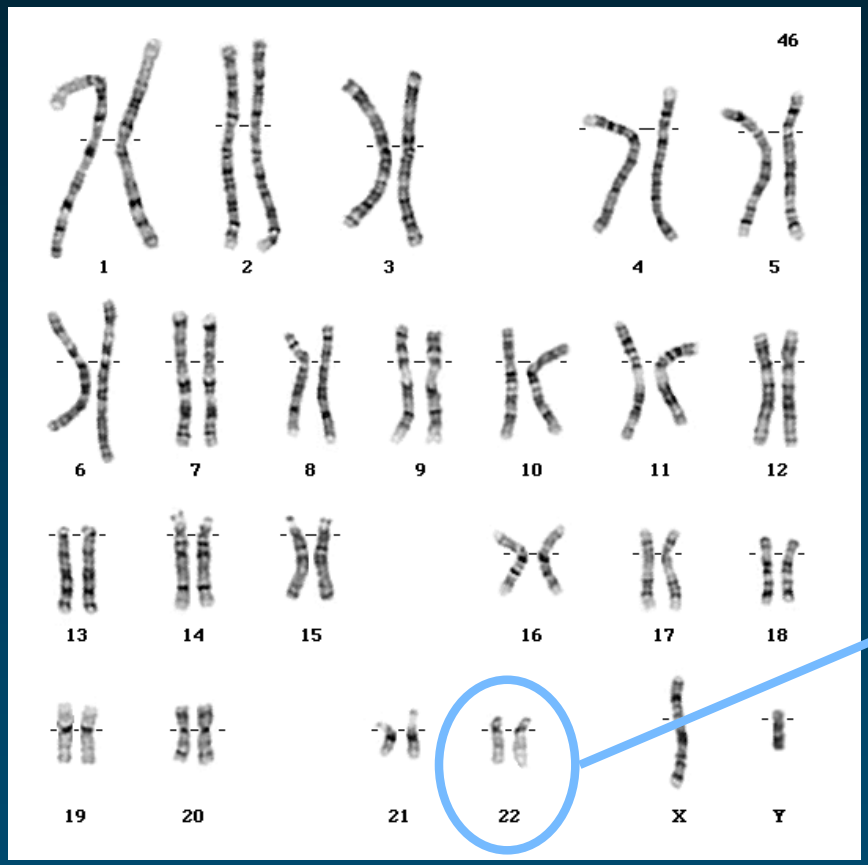
Main Headings

- Neurofibromatosis 2

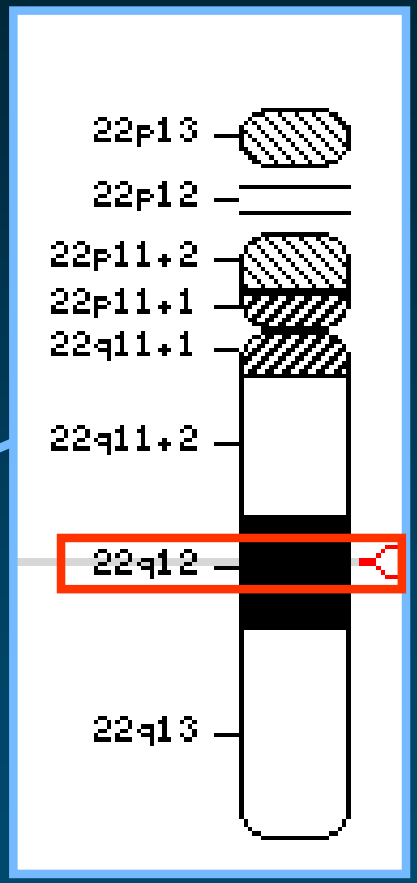
Subheadings

Document: Done (1.328 secs)

NF2 gene



http://staff.washington.edu/timk/cyto/human/



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

{UMLS_2003} UMLS® Semantic Navigator ¤ [2.10] - Netscape

{UMLS_2003} UMLS® Semantic Navigator ...

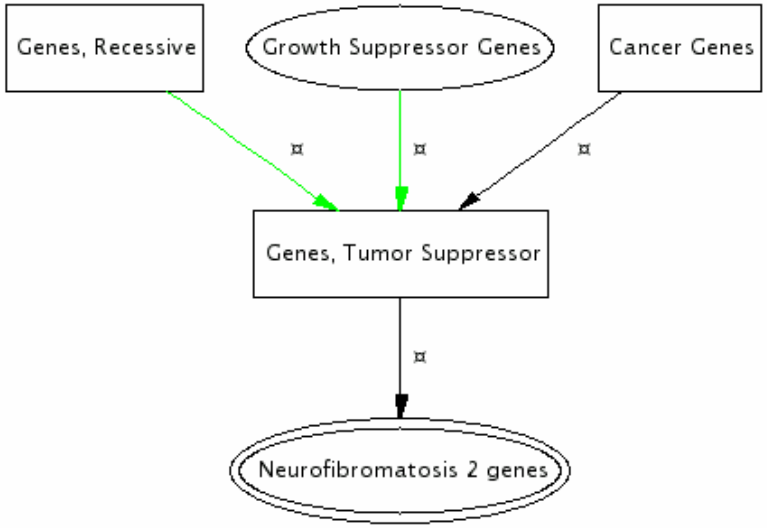
Siblings

Chemicals & Drugs

- ADAM11 protein, human ¤
- DLG5 protein, human ¤
- DPM3 protein, human ¤
- HCCS-1protein, human ¤
- hssh3bp1 protein, human ¤
- HUGL protein, human ¤
- LAPSER1 protein, human ¤
- mitochondria proteolipid-like protein, human ¤
- MRG protein, human ¤
- p53 gene/protein ¤
- PLAGL1 protein, human ¤
- RARRES3 protein, human ¤
- SEZ6L protein, human ¤
- TES protein, human ¤

Genes & Molecular Sequences

- APC Gene ¤
- BAX Gene ¤
- brca gene ¤
- CDH1 gene ¤
- CHES1 Gene ¤
- cyclin-dependent kinase inhibitor 2A



Other Related Concepts

Chemicals & Drugs

- Neurofibromin 2 ¤

Disorders

- Neurofibromatosis 2 ¤

(2 other related concepts)

- Chromosome Deletion [7] ¤
- Ependymoma [4] ¤
- Glioma [4] ¤
- Loss of Heterozygosity [7] ¤
- Meningeal Neoplasms [25] ¤
- Meningioma [30] ¤
- mesothelioma <1> [4] ¤
- Neoplasms [4] ¤
- Neurilemmoma [20]
- Neurofibromatoses
- Neurofibromatosis 2 [64] ¤
- Neuroma, Acoustic [5] ¤
- Spinal Cord Neoplasms [3] ¤

BCI

Neurofibromatosis 2 genes

LEGEND *

Start again | Apply new parameters

Restrict to vocabulary: Show all

Highlight vocabulary: Nothing

UMLS data: UMLS_2003

Type of hierarchical rel.: All / Parent/Child only / Broader/Narrower only

Similar Concepts

(none)

Allegedly Synonyms

(none)

Closest MeSH Terms

Main Headings

- Genes, Neurofibromatosis 2

Subheadings

Document: Done (3.797 secs)

Merlin

- Synonyms
  - Neurofibromin 2
  - Schwannomin
  - Schwannomerlin
  - Neurofibromatosis-2
- 10 isoforms
- Annotations
  - Negative regulation of cell proliferation
  - Cytoskeleton
  - Plasma membrane

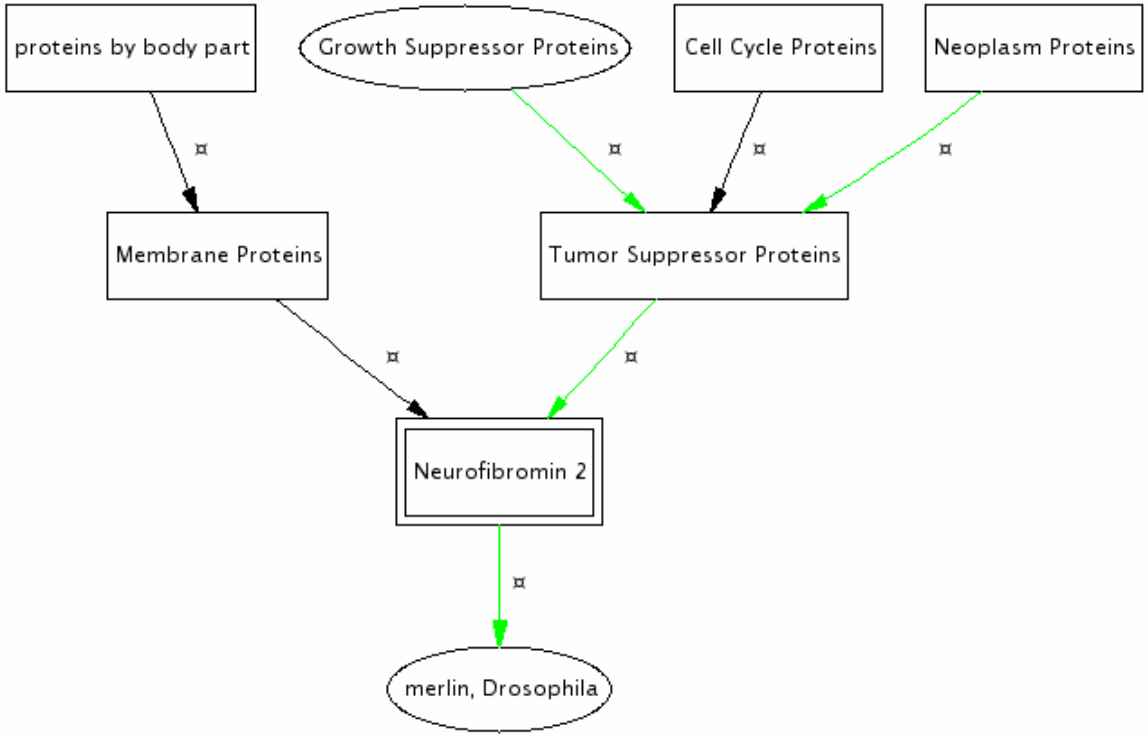
{UMLS_2003} UMLS® Semantic Navigator ¤ [2.10] - Netscape

{UMLS_2003} UMLS® Semantic Navigator ...

Siblings

Chemicals & Drugs

- (LA)12 peptide ¤
- (methyl)ammonium uptake carrier, Corynebacterium ¤
- 120-kDa hemocyte-specific membrane protein, flesh fly ¤
- 15a protein, Aedes aegypti ¤
- 22.6-kDa antigen, Schistosoma japonicum ¤
- 36-kDa vesicular integral membrane protein ¤
- 38L protein ¤
- 5-lipoxygenase-activa protein ¤
- 59 kDa dystrophin-associate protein ¤
- A-1 antigen ¤
- A-kinase anchor protein 149 ¤
- A-kinase anchor protein 15 ¤
- A-kinase anchor protein 200 ¤
- A-kinase anchor protein KL ¤
- A14.5L protein ¤
- A15 protein ¤
- ABC-me protein ¤
- ABU-1 protein, C elegans ¤
- AcfB protein ¤
- ACR3 protein ¤



Other Related Concepts

Disorders

- Neurofibromatosis 2 ¤

Genes & Molecular Sequences

- Neurofibromatosis 2 genes ¤

(2 other related concepts)

Co-occurring Concepts

Anatomy

- Arachnoid [1] ¤
- Cell Membrane [1] ¤
- Cerebellum [1] ¤
- Chromosomes, Human, Pair 22 [1] ¤
- Cytoplasm [1] ¤
- Cytoskeleton [2] ¤
- Microfilaments [1]
- Purkinje Cells [1] ¤
- Schwann Cells [1] ¤
- Stem Cells [1] ¤

BCI

Neurofibromin 2

LEGEND *

Start again | Apply new parameters

Restrict to vocabulary: Show all

Highlight vocabulary: Nothing

UMLS data: UMLS_2003

Type of hierarchical rel.: All / Parent/Child only / Broader/Narrower only

Similar Concepts

(none)

Allegedly Synonyms

(none)

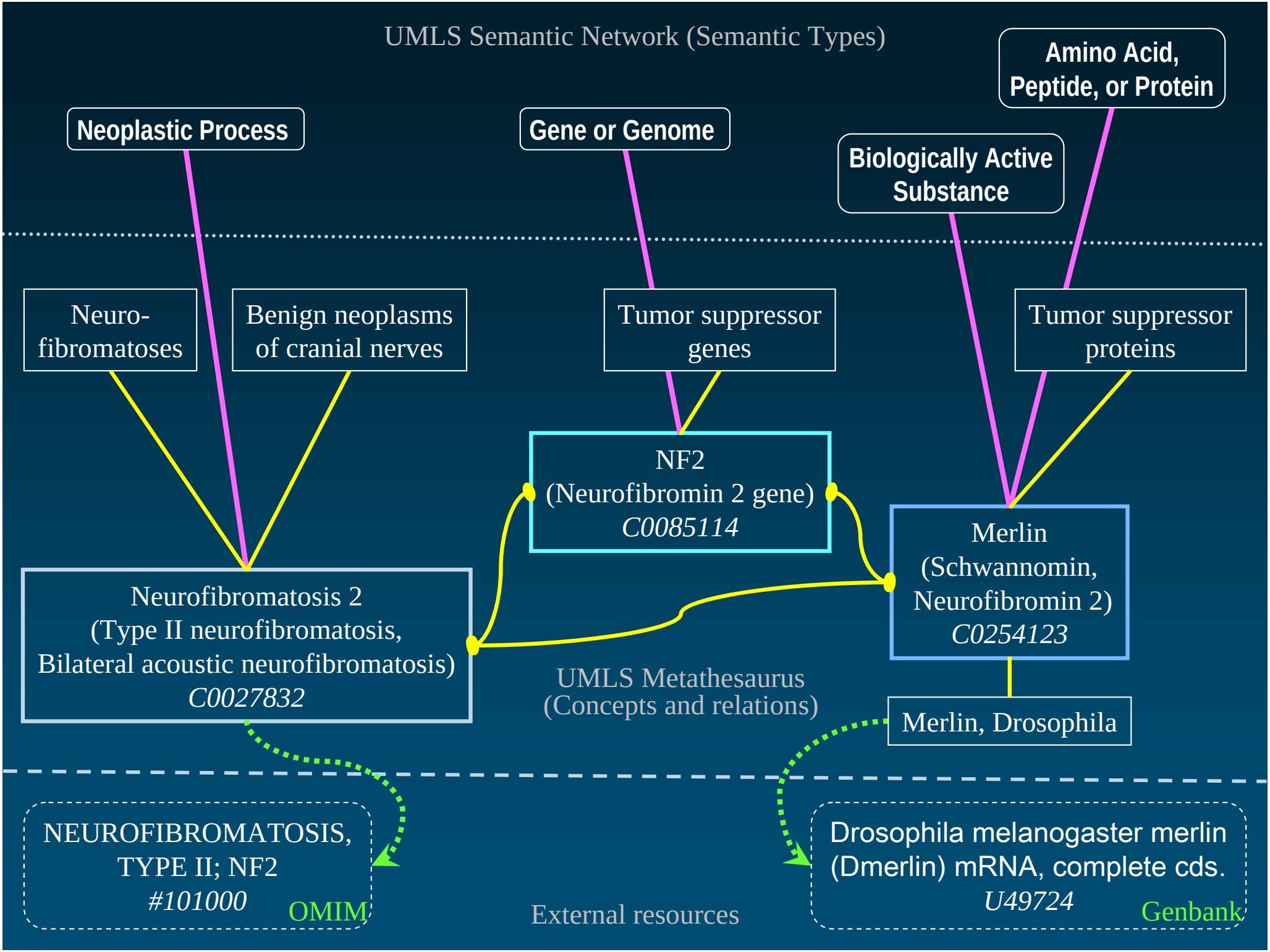
Closest MeSH Terms

Main Headings

- Neurofibromin 2

Subheadings

Document: Done (2.844 secs)


UMLS Semantic Network (Semantic Types)
Neoplastic Process
Gene or Genome
Biologically Active Substance
Amino Acid, Peptide, or Protein
Neuro-fibromatoses
Benign neoplasms of cranial nerves
Tumor suppressor genes
Tumor suppressor proteins
NF2 (Neurofibromin 2 gene) C0085114
Merlin (Schwannomin, Neurofibromin 2) C0254123
Neurofibromatosis 2 (Type II neurofibromatosis, Bilateral acoustic neurofibromatosis) C0027832
UMLS Metathesaurus (Concepts and relations)
Merlin, Drosophila
NEUROFIBROMATOSIS, TYPE II; NF2 #101000
OMIM
External resources
Drosophila melanogaster merlin (Dmerlin) mRNA, complete cds. U49724
Genbank

Limitations

- Genes not systematically represented
  - Most gene products and diseases are
- Gene/Gene product-Disease relations
  - Not systematically represented
  - Not explicitly represented (e.g., co-occurrence)
- Cross-references not systematically represented
- Naming conventions (genes)

Support for natural language processing

Semantic interpretation

- Source: text corpus / terminology
- Correspondence between
  - Linguistic phenomena
  - Semantic relations
- Semantic constraints provided by ontologies

[Navigli & al., TKE, 2002]
[Romacker, AIME, 2001]
[Rindflesch & al., JBI, 2003]

Semantic interpretation

Syntactic analysis

Mapping to UMLS

Semantic rules

Semantic network relationships

Select matching rule

Hemofiltration in digoxin overdose

Hemofiltration | in | digoxin overdose

noun phrase — preposition — noun phrase

Hemofiltration **treats** Digoxin overdose

Semantic interpretation

Therapeutic or Preventive Procedure **treats** Disease or Syndrome

Antibiotic	*treats*	Disease or Syndrome
Medical Device	*treats*	Injury or Poisoning
Pharmacologic Substance	*treats*	Congenital Abnormality
Ther. or Prev. Procedure	*treats*	Disease or Syndrome

[…]

Language & Computing

- Maria van Gurp, Manuel Decoene, Marnix Holvoet, Mariana Casella dos Santos
  LinKBase®, a Philosophically-Inspired Ontology for NLP/NLU Applications
  Proceedings of KR-MED 2006, pp. 67-75
  http://CEUR-WS/Vol-222/

DL reasoning, automated classification

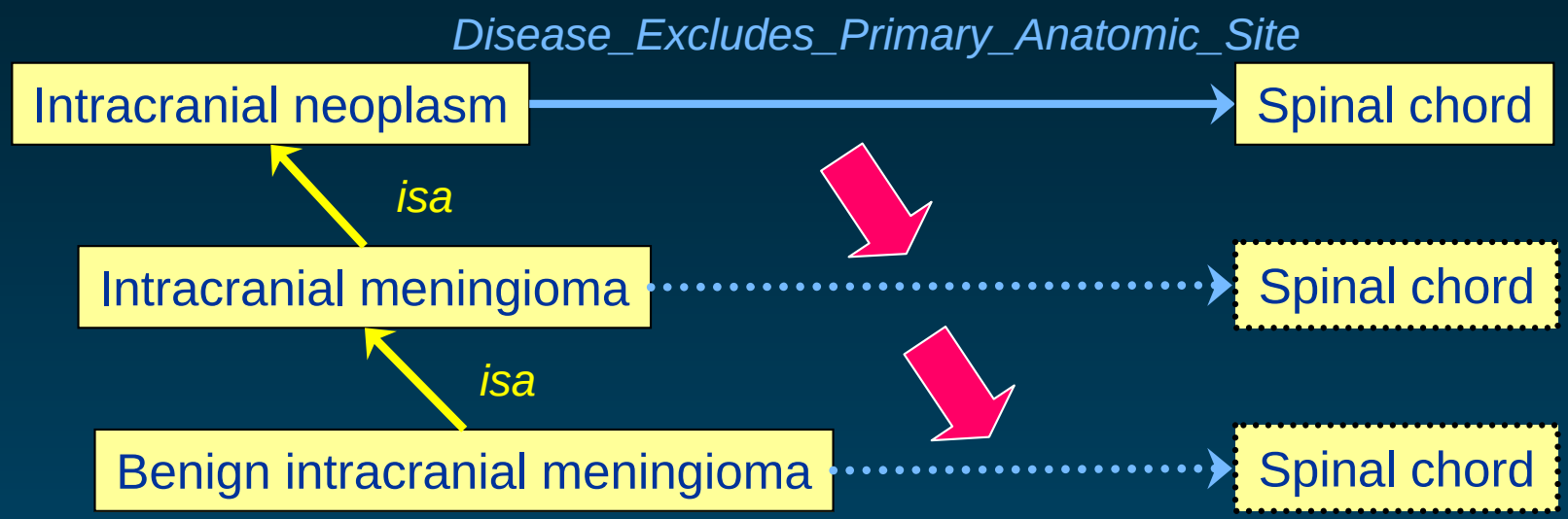
DL reasoning, automated classification

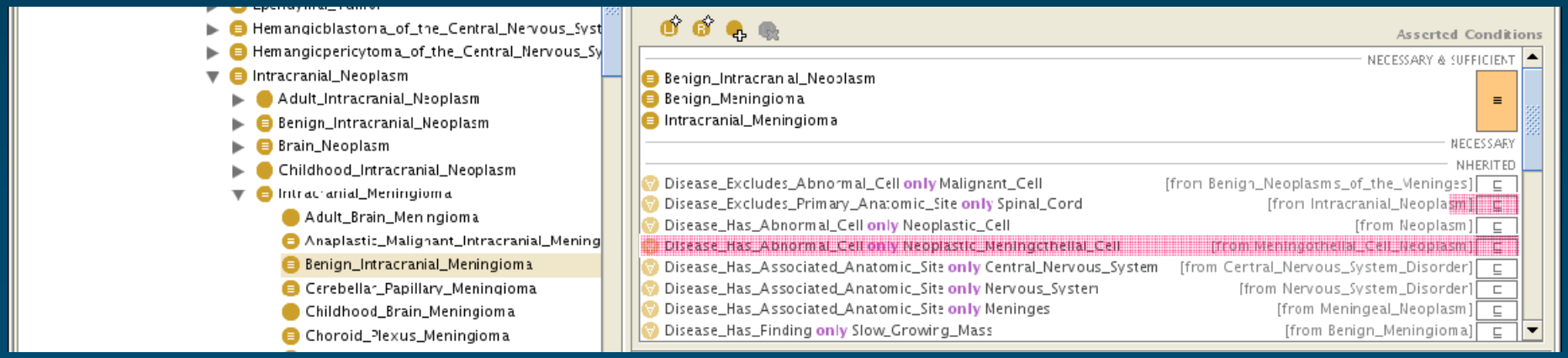
- Ontologies represent knowledge
- Automated reasoners infer conclusions from the given knowledge
  - Make implicit knowledge explicit
  - Help validate the ontology (e.g., consistency checking and automatic classification in DL)
- Need for more expressive logic
  - Inference rules

OWL reasoners

- For OWL DL, not OWL Full
- Reasoners
  - Fact++ http://owl.man.ac.uk/factplusplus/
  - Pellet http://www.mindswap.org/2003/pellet/
  - RacerPro http://www.racer-systems.com/
- Functions
  - Consistency checking
  - Automatic classification

Simple inference





Complex inference

- Clinical decision support
  - If patient is treated by aminoglycosides
    and patient has impaired renal function
    then reduce dose (or frequency of administration) of aminoglycosides
- Not directly supported by DL reasoners
- Require rule languages
  - RuleML http://www.ruleml.org/
  - SWRL (Semantic Web Rule Language)

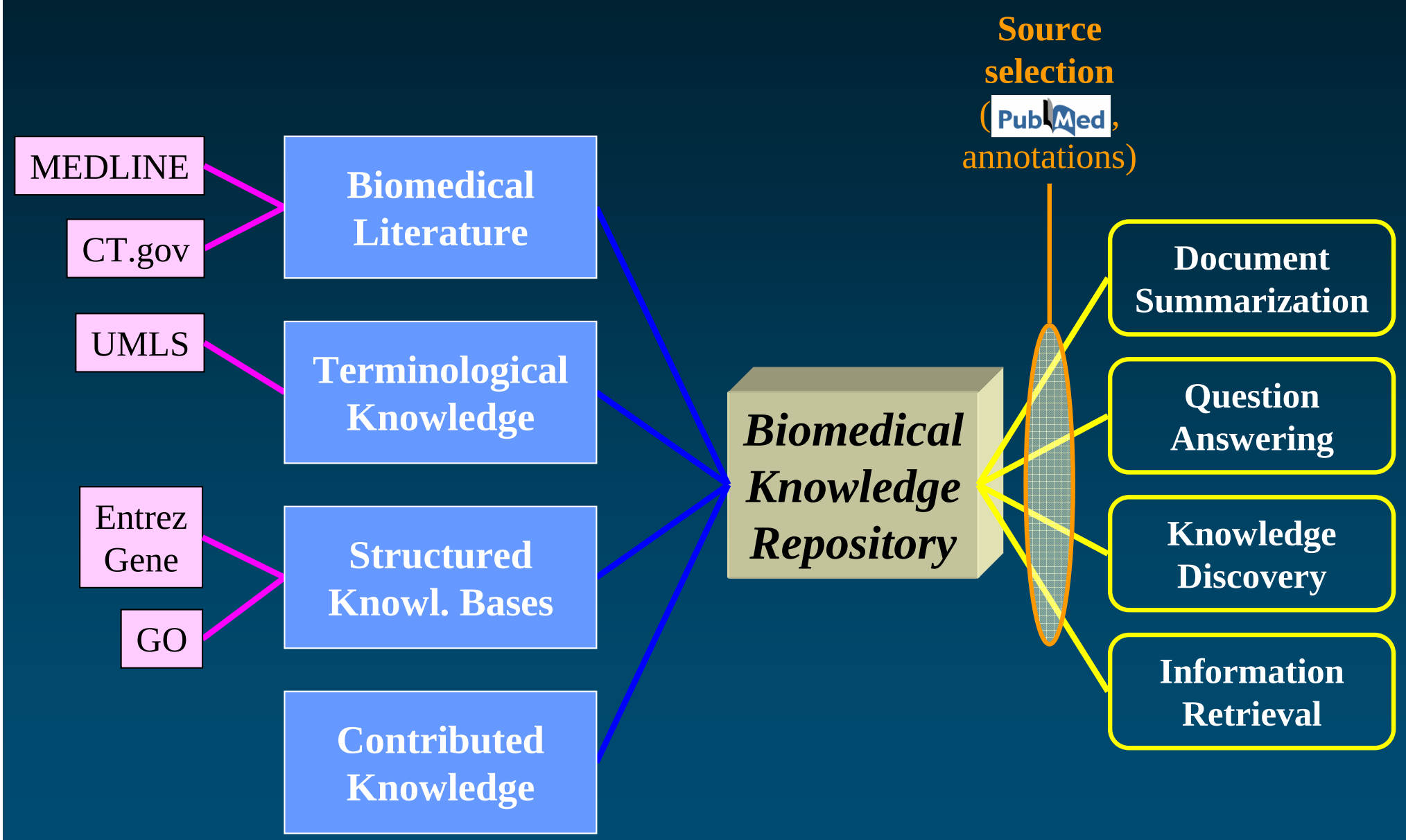
    http://www.w3.org/Submission/SWRL/

Support for knowledge discovery

Motivation

- Biomedical information is growing at an increasingly faster pace
  - High-throughput approach to knowledge processing
- Information retrieval is the starting point, not the end of the journey for the researcher
  - Towards “computable” knowledge
- Integration between literature and other resources is insufficient
  - Adequate for navigation purposes
  - Insufficient for knowledge processing

Advanced Library Services Summary



Some issues for discussion

Issues

- What is the right amount of semantics?
  - "A little semantics goes a long way" – Does it?
- What approach to building ontologies?
  - Top-down vs. bottom-up
- What formalism?
  - XML, RDF/S, OWL Lite/DL/Full, OBO?
- Semantic Web for Health Care and Life Sciences

Current trends and future directions

Briefings in Bioinformatics

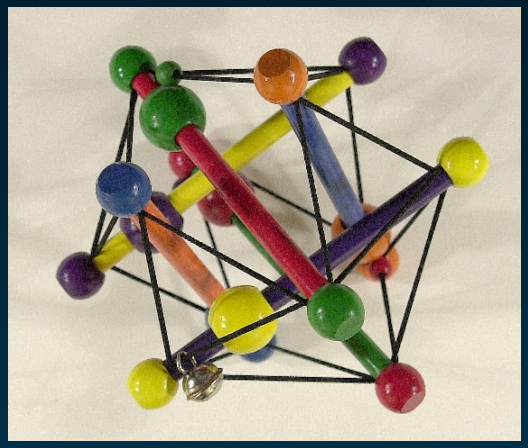
BRIEFINGS IN BIOINFORMATICS. VOL 7. NO 3. 256–274 doi:10.1093/bib/bbl027

Bio-ontologies: current trends and future directions

Olivier Bodenreider and Robert Stevens

Submitted: 23rd June 2006; Received (in revised form): 10th July 2006

http://bib.oxfordjournals.org/cgi/reprint/7/3/256?ijkey=1ejwW7ipyG1ASiI&keytype=ref



Medical Ontology Research

Contact: olivier@nlm.nih.gov

Web: mor.nlm.nih.gov

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.
- Seminal paper
  - Lindberg, D. A., Humphreys, B. L., & McCray, A. T. (1993). The Unified Medical Language System. *Methods Inf Med*, *32*(4), 281-91.

Semantic Web References

- World Wide Web Consortium (W3C)
  - http://www.w3.org/
- W3C Health Care and Life Sciences Interest Group
  - http://www.w3.org/2001/sw/hcls/