

Session 3:

Hands-on session

Make sure you have downloaded the Docker Image for Session 3

- If you haven't already done (following from email):
 1. Install Docker
 2. Download Docker image - <https://hub.docker.com/r/ielabgroup/health-search-tutorial>

Practical session - outline

- In this hands-on session, we will:
 1. Take a collection of clinical trials, annotate them with medical concepts, producing documents with both term and concept representation.
 2. Index these documents in Elasticsearch with multi term/concepts fields.
 3. Search Elasticsearch with either term or concept, demonstrating semantic search capabilities.
 4. Play a bit more
- Instructions: <https://github.com/ielab/health-search-tutorial/tree/master/hands-on>

Auxiliary Details on UMLS

(not covered in tutorial)

Digging into the UMLS

- There UMLS Metathesaurus:
- The heart of UMLS
- Consists of over 1,000,000 concepts

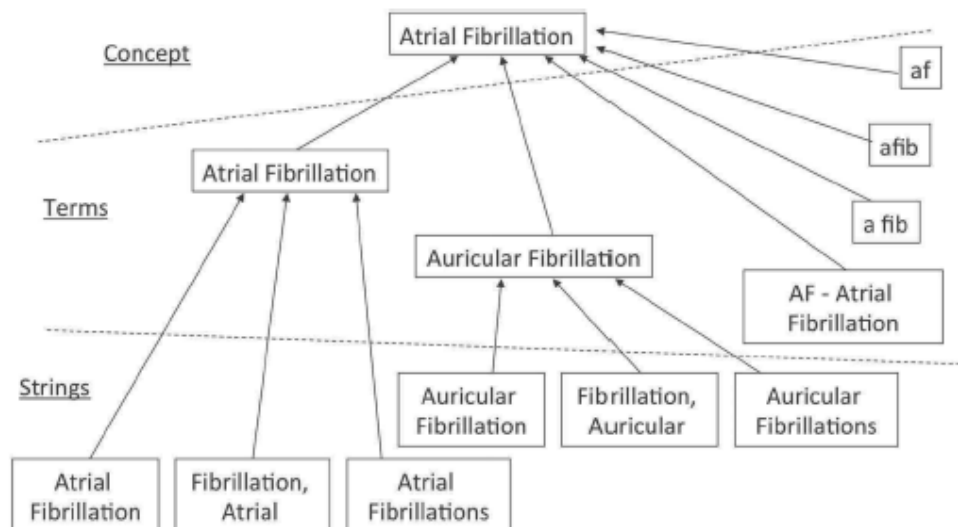


FIGURE 14.3: Unified Medical Language System Metathesaurus concept of atrial fibrillation. (Courtesy of NLM)

Table 1. Concept, Term, Atom, and String Identifiers.

Concept (CUI)	Terms (LUIs)	Strings (SUIs)	Atoms (AUIs) * RRF Only
C0004238 Atrial Fibrillation (preferred) Atrial Fibrillations Auricular Fibrillation Auricular Fibrillations	L0004238 Atrial Fibrillation (preferred) Atrial Fibrillations	S0016668 Atrial Fibrillation (preferred)	A0027665 Atrial Fibrillation (from MSH) A0027667 Atrial Fibrillation (from PSY)
		S0016669 (plural variant) Atrial Fibrillations	A0027668 Atrial Fibrillations (from MSH)
	L0004327 (synonym) Auricular Fibrillation Auricular Fibrillations	S0016899 Auricular Fibrillation (preferred)	A0027930 Auricular Fibrillation (from PSY)
		S0016900 (plural variant) Auricular Fibrillations	A0027932 Auricular Fibrillations (from MSH)

The UMLS Metathesaurus

- UMLS tables - Many of them, but 2 key tables:

1. MRConso.rtf: Links each concept (CUI) to each unique concept in a source vocabulary (AUI)

Sample Records

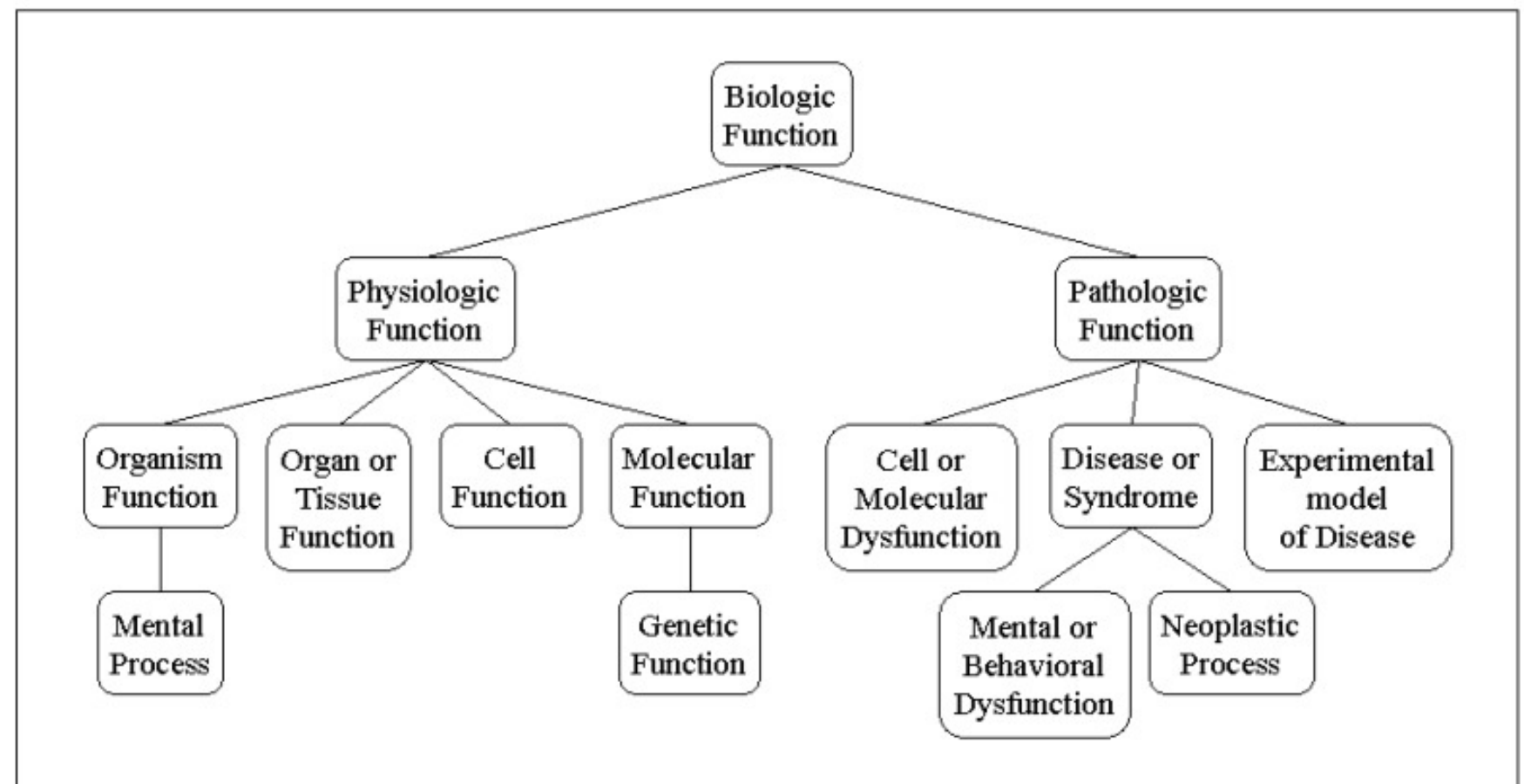
C0001175|ENG|P|L0001175|VO|S0010340|Y|A0019182||M0000245|D000163|MSH|PM|D000163|Acquired Immunodeficiency Syndromes|0|N|1792|
 C0001175|ENG|S|L0001842|PF|S0011877|N|A2878223|103840012|62479008||SNOMEDCT|PT|62479008|AIDS|9|N|3840|
 C0001175|ENG|P|L0001175|VC|S0354232|Y|A2922342|103845019|62479008||SNOMEDCT|SY|62479008|Acquired immunodeficiency syndrome|9|N|3584|
 C0001175|FRE|S|L0162173|PF|S0226654|Y|A7438879||M0000245|D000163|MSHFRE|EN|D000163|SIDA|3|N||
 C0001175|RUS|S|L0904943|PF|S1108760|Y|A13488500||M0000245|D000163|MSHRUS|SY|D000163|SPID|3|N||

2. MRREL.rtf: links 2 concepts with a relationship. Note the relationship is AUI to AUI

REL	Description
AQ	Allowed qualifier
CHD	has child relationship in a Metathesaurus source
DEL	Deleted concept
PAR	has parent relationship in a Metathesaurus source
QB	can be qualified by.
RB	has a broader relationship
RL	the relationship is similar or "alike". the two
RN	has a narrower relationship
RO	has relationship other than synonymous,
RQ	related and possibly synonymous.
RU	Related, unspecified
SIB	has sibling relationship in a Metathesaurus
SUBX	Concept removed from current subset
SY	source asserted synonymy.
XR	Not related, no mapping
	Empty relationship

UMLS Semantic Network

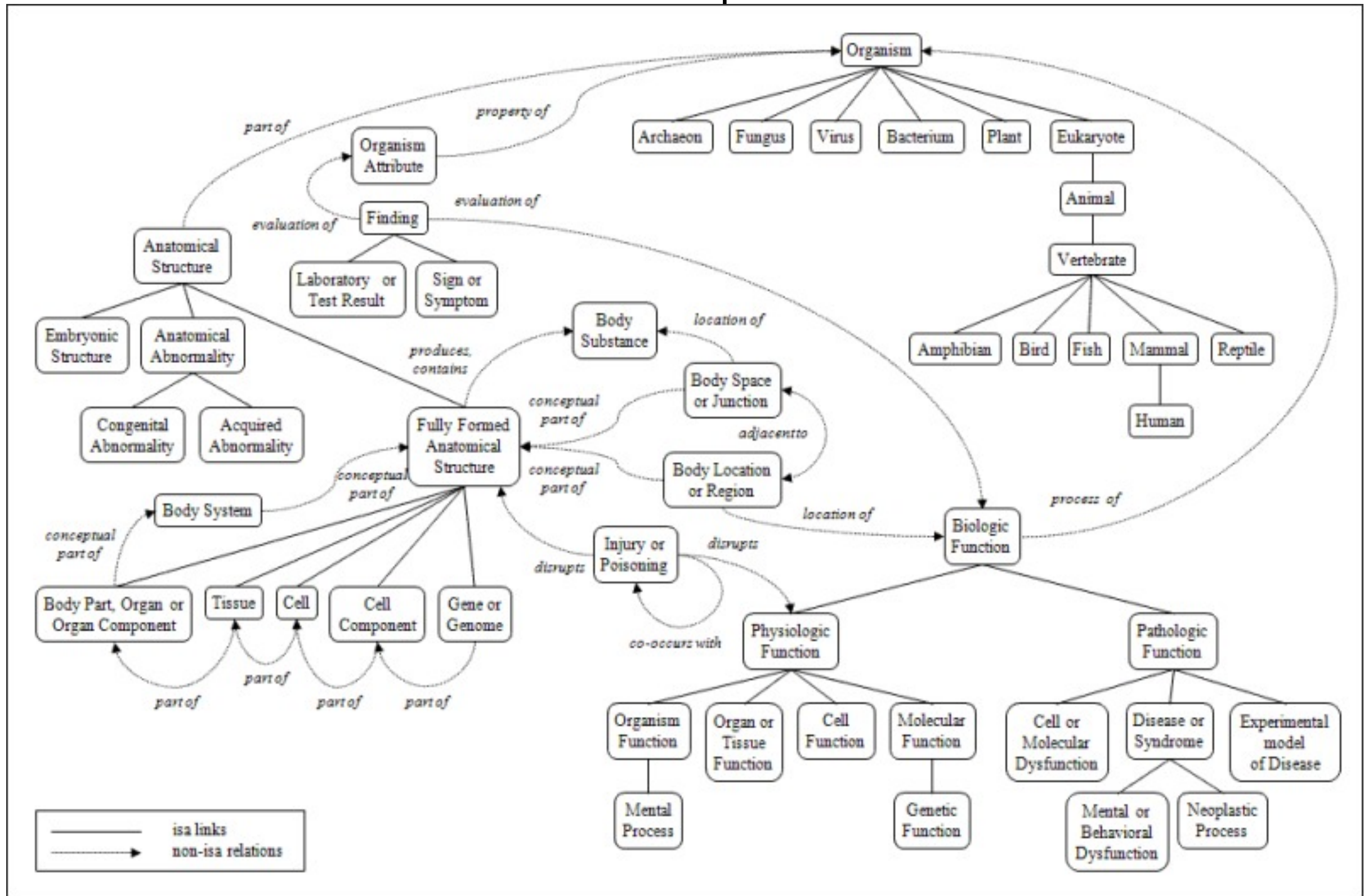
- Categorise concepts and relationships among concepts.
- Each UMLS concept is assigned to at least one semantic type (i.e., defines the concept category).
- Semantic types are related by links: "is a", "part of", "process of", etc.



Part of "is a" semantic hierarchy.

Full hierarchy: https://www.nlm.nih.gov/research/umls/META3_current_semantic_types.html

UMLS Semantic Network Links: Example



UMLS Semantic Type Groups

- To reduce the complexity, semantic types are grouped based on its: semantic validity, parsimony, completeness, exclusivity, naturalness, and utility.

DISO	Disorders	T020	Acquired Abnormality
DISO	Disorders	T190	Anatomical Abnormality
DISO	Disorders	T049	Cell or Molecular Dysfunction
DISO	Disorders	T019	Congenital Abnormality
DISO	Disorders	T047	Disease or Syndrome
DISO	Disorders	T050	Experimental Model of Disease
DISO	Disorders	T033	Finding
DISO	Disorders	T037	Injury or Poisoning
DISO	Disorders	T048	Mental or Behavioral Dysfunction
DISO	Disorders	T191	Neoplastic Process
DISO	Disorders	T046	Pathologic Function
DISO	Disorders	T184	Sign or Symptom

Semantic types within Semantic type group: "Disorders"

UMLS semantic network Detail: SRDEF

- Basic information about the semantic types and relations

Field	Description
RT:	Record Type (STY = Semantic Type or RL = Relation).
UI:	Unique Identifier of the Semantic Type or Relation.
STY/RL:	Name of the Semantic Type or Relation.
STN/ REL:	Tree Number of the Semantic Type or Relation.
DEF:	Definition of the Semantic Type or Relation.
EX:	Examples of Metathesaurus concepts with this Semantic Type (STY
UN:	Usage note for Semantic Type assignment (STY records only).
NH:	The Semantic Type and its descendants allow the non-human flag
ABR:	Abbreviation of the Relation Name or Semantic Type.
RIN:	Inverse of the Relation (RL records only).

SRDEF Sample entries

STY|T047|Disease or Syndrome|B2.2.1.2.1|A condition which alters or interferes with a normal process, state, or activity of an organism. It is usually characterized by the abnormal functioning of one or more of the host's systems, parts, or organs. Included here is a complex of symptoms descriptive of a disorder.||Any specific disease or syndrome that is modified by such modifiers as "acute", "prolonged", etc. will also be assigned to this type. If an anatomic abnormality has a pathologic manifestation, then it will be given this type as well as a type from the 'Anatomical Abnormality' hierarchy, e.g., "Diabetic Cataract" will be double-typed for this reason.||dsyn||S

RL|T186|isa|H|The basic hierarchical link in the Network. If one item "isa" another item then the first item is more specific in meaning than the second item.||||S|inverse_isa|

UMLS semantic network Detail: SRSTR

- Structure of the network

STY/RL:	Argument 1 (Name of a Semantic Type or Relation).
RL:	Relation ("isa" or the name of a non-hierarchical Relation).
STY/RL:	Argument 2 (Name of a Semantic Type or Relation); if this field is blank this means that the Semantic Type or Relation is one of the top
LS:	Link Status (D = Defined for the Arguments and its children; B = Blocked; DNI = Defined but Not Inherited by the children of the

Disease or Syndrome|conceptually_related_to|Experimental Model of Disease|DNI|
Disease or Syndrome|isa|Pathologic Function|D|
Disease or Syndrome|produces|Tissue|D|

Semantic Network Detail: SRSTRE1 OR SRSTRE2

- Fully inherited set of Relations (UI's or Names).

UI/STY:	Argument 1 (UI or name of a Semantic Type).
UI/RL:	Relation (UI or name of a nonhierarchical Relation).
UI/STY:	Argument 2 (UI or name of a Semantic Type). N.B.: The relations expressed in this table are binary relations and the arguments are ordered pairs. All relations have been fully inherited in this table.

SRSTRE1

SRSTRE2

T047 T186 T038	Disease or Syndrome isa Biologic Function
T047 T186 T046	Disease or Syndrome isa Pathologic Function
T047 T186 T051	Disease or Syndrome isa Event
T047 T186 T067	Disease or Syndrome isa Phenomenon or Process
T047 T186 T070	Disease or Syndrome isa Natural Phenomenon or Process