

AMIA 2006 Tutorial T12  
Unified Medical Language System<sup>®</sup>  
**UMLS<sup>®</sup> Overview**

Rachel Kleinsorge

Jan Willis

Allen Browne

Alan Aronson

National Library of Medicine

National Institutes of Health

U.S. Dept. of Health & Human Services



# Schedule

---

- 1:00 - 1:30 UMLS Overview
- 1:30 - 2:00 Metathesaurus
- 2:00 - 2:15 Electronic Health Data Standards
- 2:15 - 2:30 UMLS Knowledge Source Server
- 2:30 - 3:00 MetamorphoSys:  
**Customizing the Metathesaurus**
  
- 3:00 - 3:30 - break -
  
- 3:30 - 4:00 SPECIALIST Lexicon and Lexical Tools
- 4:00 - 4:30 MetaMap Technology Transfer (MMTx)



# Unified Medical Language System

UMLS H

[Home](#) > [Biomedical Research & Informatics](#) > [UMLS](#)

9/30/06: **Draft LOINC to CPT Mappings** now available for download from the UMLSKS. ••• New to the UMLS? [Register now.](#)

## ▶ [About the UMLS Resources](#)

Metathesaurus; Semantic Network; SPECIALIST Lexicon and lexical programs; MetamorphoSys

## ▶ [Accessing UMLS Knowledge Sources](#)

Metathesaurus license; Semantic Network; SPECIALIST Lexicon; DVD

## ▶ [Knowledge Source Server](#)

Download files; searching; additional tools and resources

## ▶ [Documentation](#)

## ▶ [Help](#)

Training; contact us; FAQs; listserv

## About

NLM's Unified Medical Language System (UMLS) project develops and distributes multi-purpose, electronic "Knowledge Sources" and associated lexical programs for system developers. Researchers will find the UMLS products useful in investigating knowledge representation and retrieval questions.

## Metathesaurus Source Vocabularies

- [SNOMED CT](#)
- [LOINC](#)
- [RxNorm](#)
- [MeSH](#)
- [List of Sources](#)
- [Source FAQs](#)
- [Mappings](#)

## More Resources

- [Metathesaurus License](#)
- [Tools](#)
- [Learning Resources](#)
- [MetaMap Transfer \(MMTx\)](#)
  
- [Archives](#)

**Last reviewed:** 03 October 2006

**Last updated:** 03 October 2006

**First published:** 01 January 1999

**Metadata** | [Permanence level](#): Permanence Not Guaranteed

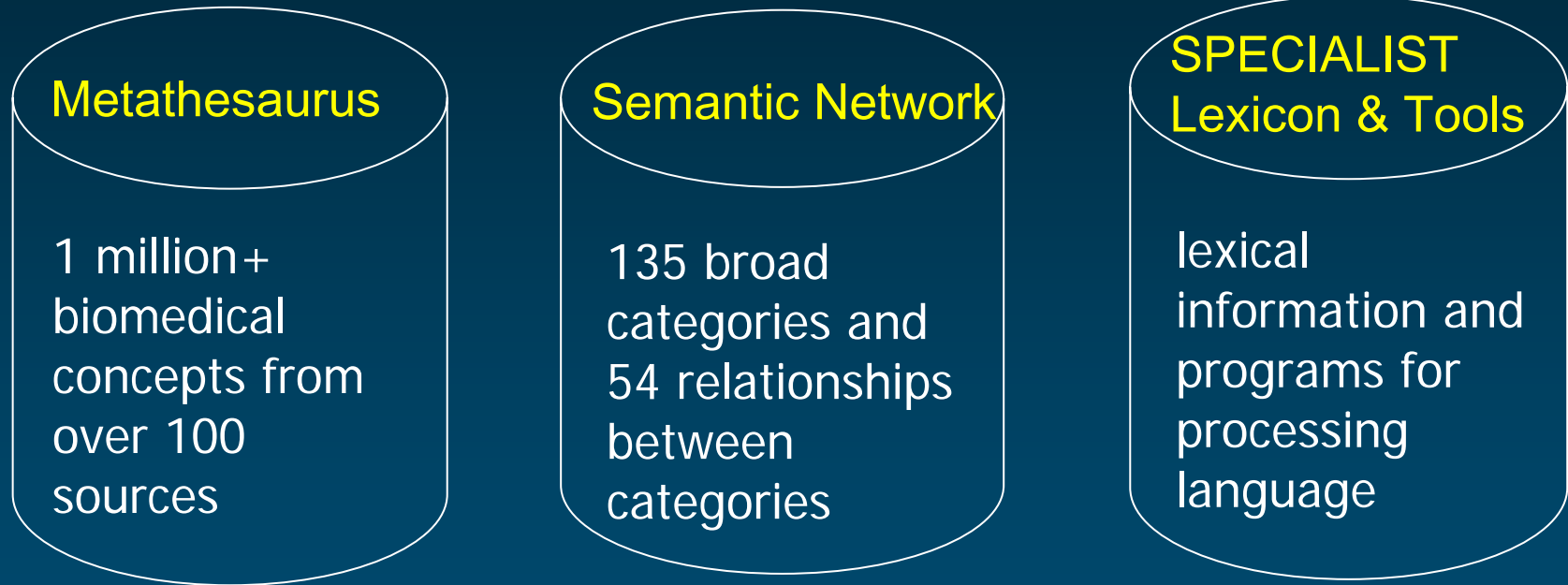
[Copyright](#), [Privacy](#), [Accessibility](#)

U.S. National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894

National Institutes of Health, Health & Human Services

# The UMLS consists of

---



**3 Knowledge Sources  
+ associated tools**

# UMLS Objectives

---

- Intellectual middleware
- A set of multi-purpose tools for system developers
- Knowledge Sources used to overcome:
  - disparities in language and language format  
Ex: atrial fibrillation, auricular fibrillation, af
  - disparities in granularity and perspective  
Ex: Contusions, hematoma, bruise  
Ex: Instruct patient to promptly report nosebleeds and excessive bruising (NIC), Epistaxis (MeSH)

# UMLS made available as:

---

- Data files
  - 3 separate sets of relational files
- Tools
  - MetamorphoSys (installation and customization)
  - RRF Subset Browser
  - lvg (lexical programs)
- Distributed on DVD
- Downloaded from UMLS Knowledge Server (UMLSKS)

**The UMLS is *not* an end-user application**

# UMLS Access

---

## ■ Remote access

### ■ UMLS Knowledge Server (UMLSKS)

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

■ Browsers, Navigators

■ APIs

■ Download files and programs

■ Documentation

## ■ Local access

■ **MetamorphoSys**: install files locally, create customized Metathesaurus subsets

■ **Subset Browser**: search, browse, view customized subsets

# Metathesaurus License Agreement

---

## ■ Online Web-based license:

<http://www.nlm.nih.gov/research/umls/license.html>

- Read license
- Read appendix
- Print a copy for your records
- Complete the Web form

Accept and continue

Accept

 **Printer-friendly Version**

Submit

## ■ Verification and turnaround:

- Receive e-mail from NLM and respond within 72 hours
- NLM official countersigns, license added to database
- Receive 2<sup>nd</sup> e-mail from NLM with new license number



# License Agreement Restrictions

---

- 2. No charges, usage fees or royalties will be paid to NLM.
- 5. Within 30 days of the end of any calendar year ... provide NLM with a brief report
- 11.c. required to include ... identifiers from ... the original source vocabularies
- 12. For material ... from some sources additional restrictions ... may apply.  
See list of current sources in Appendix A.1.

# Metathesaurus

---

- The Metathesaurus is
  - very large
  - multi-purpose
  - multi-lingual
  
- It contains information about
  - biomedical and health related concepts
  - their various names and associated codes
  - the relationships among them

# Metathesaurus: clusters terms by meaning

- Concepts contain synonymous terms
- Preferred term is chosen (default can be changed)
- Unique identifier (CUI) is assigned

term	source	term type	source ID
Addison's disease	Metathesaurus	PN	
<b>Addison's disease</b>	<b>SNOMED CT</b>	<b>PT</b>	<b>363732003</b>
Addison's Disease	MedlinePlus	PT	T1233
<b>Addison Disease</b>	<b>MeSH</b>	<b>PT</b>	<b>D000224</b>
Bronzed disease	SNOMED Intl	SY	DB-70620
<b>Primary Adrenal Insufficiency</b>	<b>MeSH</b>	<b>EN</b>	<b>D000224</b>
Primary hypoadrenalism syndrome, Addison	MedDRA	LT	10036696

C0001403

Addison's disease

# Organization of Concepts

Concept  
C0001621

Term  
L0001621

**S0011231 Adrenal Gland Disease**

A0020266 MeSH

A7568579 NCI Thesaurus

**S0000441 Disease of adrenal gland**

A0001264 SNOMED 1982

A6917004 SNOMED Clinical Terms

**S0481705 Diseases of Adrenal Gland**

A0014499 SNOMED 1982

**S0220090 Diseases, adrenal gland**

A0049924 MeSH

Term  
L0181041

**S0632950 Disorder of adrenal gland**

A0688820 Read Codes

A4778687 SNOMED Clinical Terms

**S0354509 Adrenal Gland Disorders**

A6996540 MedlinePlus

A7576253 NCI Thesaurus

A7561794 Psychological Index Terms

Term  
L1279026

**S1520972 Nebennierenkrankheiten**

A7500884

# Semantic Network

---

## ■ 135 Semantic Types

- Broad subject categories (Clinical Drug, Virus)

■ Ex:

- Addison's Disease

- Semantic Type: Disease or Syndrome

## ■ 54 Semantic Relationships

- Links between categories (isa, causes, treats)

■ Ex:

- Virus causes Disease or Syndrome

## ■ Types + Relationships

- Form the structure of the semantic network
- Broadly categorize the biomedical domain

# 54 Semantic Relationships

---

## ■ Hierarchical (isa = is a kind of)

### ■ among types

■ Animal *isa* Organism

■ Enzyme *isa* Biologically Active Substance

### ■ among relationships

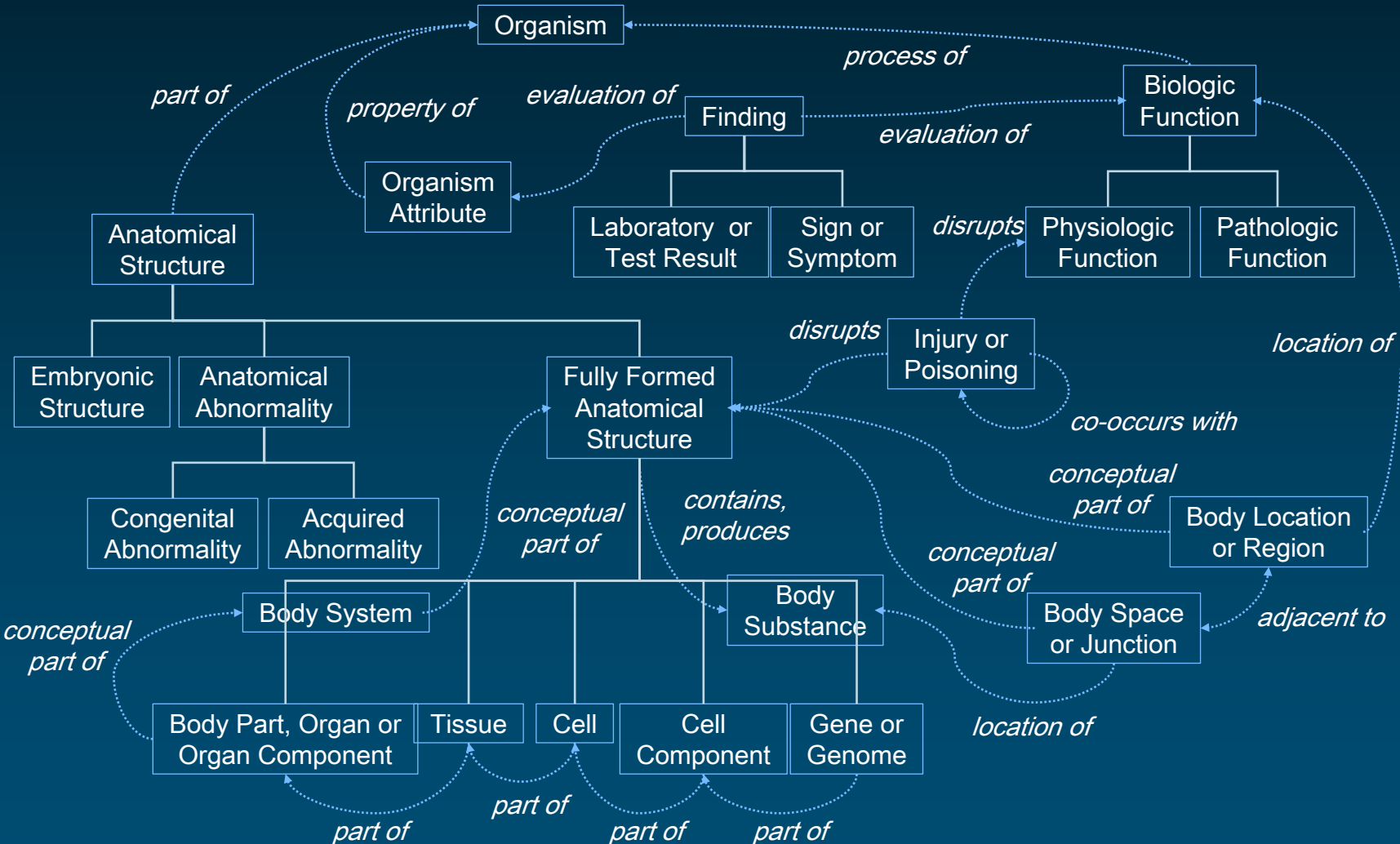
■ prevents *isa* affects

## ■ Non-hierarchical

■ Sign or Symptom *diagnoses* Pathologic Function

■ Pharmacologic Substance *treats* Pathologic Function

# Semantic Network Relationships



# Why have a Semantic Network?

---

## ■ Semantic Types

- **High level categories** assigned to Metathesaurus concepts
- Independent of position in source hierarchies

## ■ Semantic Relations

- **Useful links between** Semantic Types
- Relationships may hold at the concept level
- Other relationships may apply at the concept level



# SPECIALIST Lexicon

---

- Syntactic English lexicon of common words, biomedical terms (250K+ words, 400K+ variants)
  - Word properties
    - Syntax (how words are put together)
    - Morphology (inflection, derivation, and compounding)
    - Orthography (spelling)
  - Used by SPECIALIST Natural Language Processing System to process text and terms
    - Customizable
    - Used to maintain Metathesaurus, indexes
- Adrenal gland diseases  
Diseases of the adrenal glands  
Adrenal disorder  
Disorder of adrenal gland  
C0001621

17

# SPECIALIST Lexicon

# lexical records

```
{base=Kaposi's sarcoma
spelling_variant=Kaposi
sarcoma
entry=E0003576
  cat=noun
  variants=uncount
  variants=reg
  variants=glreg
}
```

```
{base=aspirate
entry=E0010803
  cat=verb
  variants=reg
  tran=np
  nominalization=aspiration|noun|E0010804
}
```

```
{base=chronic
entry=E0016869
  cat=adj
  variants=inv
  position=attrib(1)
  position=pred
  stative
}
```

```
{base=in
entry=E0033870
  cat=prep
}
```

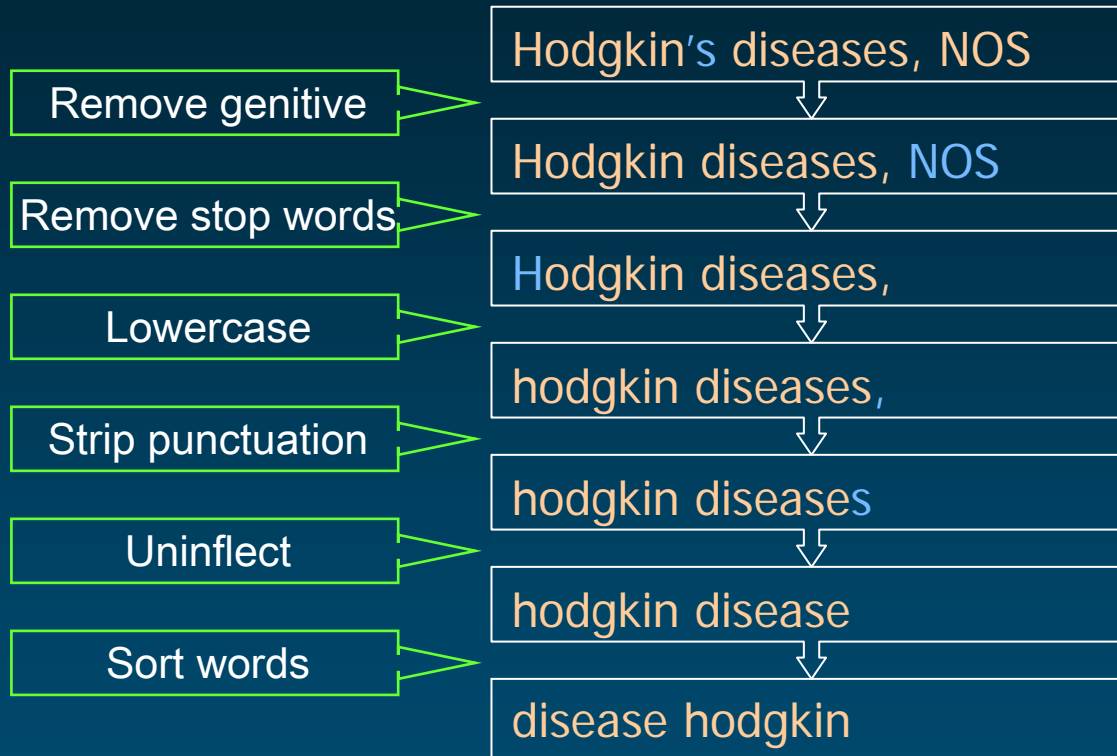
# Lexical Tools

---

- **JAVA programs to manage lexical variation, indexing, normalization in biomedical text**
- **Wordind**
  - breaks strings into words
  - produces the Metathesaurus word indexes (MRXW)
- **lvg**
  - performs various lexical transformations
  - 58 flow components and 38 options in 2006
- **NORM**
  - a selection of LVG transformations
  - produces Metathesaurus normalized word and string indexes (MRXNW & MRXNS)
  - used to access those indexes

# Normalization

---



# Normalization: Example

---

Hodgkin Disease  
HODGKINS DISEASE  
Hodgkin's Disease  
Disease, Hodgkin's  
Hodgkin's, disease  
HODGKIN'S DISEASE  
Hodgkin's disease  
Hodgkins Disease  
Hodgkin's disease NOS  
Hodgkin's disease, NOS  
Disease, Hodgkins  
Diseases, Hodgkins  
Hodgkins Diseases  
Hodgkins disease  
hodgkin's disease  
Disease, Hodgkin

normalize

disease hodgkin

Normalized  
term is not  
necessarily  
readable

- Metathesaurus:
  - 120 sources
  - 1,352,403 concepts
  - 17 languages
  
- Semantic Network:
  - 135 Semantic Types
  - 54 Semantic Relationships
  
- SPECIALIST Lexicon:
  - Over 297K records (over 482K inflectional forms)

Metathesaurus

# Metathesaurus Sources

---

- Wide range of general and specialized biomedical terminologies
- Used in variety of settings and purposes:
  - Clinical
  - Administrative
  - Public Health Reporting
  - Research



# Metathesaurus Sources

---

- 120 vocabularies in 17 languages
- Sets of valid values
  - Thesauri, e.g., MeSH, CRISP, NCI
  - Statistical classifications, e.g., ICD-9-CM
  - Billing codes, e.g., CPT
  - Clinical coding systems, e.g., SNOMED CT
- See License Appendix, documentation

# Metathesaurus: not a single vocabulary

---

- One size does not fit all
  - NLM supports coordination when possible
  - Growing awareness of benefits of standardization

"The UMLS approach assumes continuing diversity in the formats and vocabularies of different information sources and in the language employed by different elements of the biomedical community. It is not an attempt to build a single standard biomedical vocabulary."



**Humphreys, BL and PL Schuyler.** In: Broering NC, ed. High- Performance Medical Libraries: advanced information management for the virtual era. Westport (CT): Meckler; 1993, p. 33.

# Metathesaurus highlights

---

- Concept based
- Represents the meaning in each source
- Represents and delivers data in common format
- Adheres to principle of “source transparency”
  
- Source information tagged
- Context-free unique identifiers added
  
- Normalized word and string indexes included

# Source Data to Metathesaurus Files

---

Names, Synonyms

Terms, Codes → MRCONSO

Relationships → MRREL

**Hierarchies** → MRHIER

Attributes → MRSAT

Definitions → MRDEF

Mappings → MRMAP, MRSMAP,  
(MRREL, MRCONSO)

# MRCONSO (sample rows 1..5)

	1	2	3	4	5	6	7	8	9	10	11
	CUI	LAT	TS	LUI	STT	SUI	ISPREF	AUI	SAUI	SCUI	SDUI
1	C0001403	ENG	P	L0001403	PF	S0354372	Y	A4367951			
2	C0001403	ENG	P	L0001403	PF	S0354372	N	A2922421	485624014	363732003	
3	C0001403	ENG	P	L0001403	VC	S0010794	Y	A0019740		M0000346	D000224
4	C0001403	ENG	S	L0494851	PF	S2164152	N	A2018589			
5	C0001403	FRE	P	L3246333	PF	S3773545	Y	A3996251			D000224

	12	13	14	15	16	17	18
	SAB	TTY	CODE	STR	SRL	SUPPRESS	CVF
1	MTH	PN	NOCODE	Addison's disease	0	N	
2	SNOMEDCT	PT	363732003	Addison's disease	4	N	
3	MSH	MH	D000224	Addison's Disease	0	N	
4	MDR	LT	10052381	Primary adrenal insufficiency	3	N	
5	MSHFRE	MH	D000224	Addison, maladie	3	N	

# MRCONSO (sample rows 6..10)

	1	2	3	4	5	6	7	8	9	10	11
	CUI	LAT	TS	LUI	STT	SUI	ISPREF	AUI	SAUI	SCUI	SDUI
6	C0001403	FRE	S	L1272481	PF	S1514427	Y	A1464383			
7	C0001403	GER	P	L1229627	PF	S1471573	Y	A4030156			D000224
8	C0001403	GER	S	L1239271	PF	S1481217	Y	A4034094			D000224
9	C0001403	JPN	P	L3437833	PF	S3965327	Y	A4264008			D000224
10	C0001403	JPN	S	L3465347	PF	S3992841	Y	A4291522			D000224

	12	13	14	15	16	17	18
	SAB	TTY	CODE	STR	SRL	SUPPRESS	CVF
6	WHOFRE	IT	0410	MALADIE D'ADDISON	2	N	
7	MSHGER	MH	D000224	Addison-Krankheit	3	N	
8	MSHGER	SY	D000224	Bronzehautkrankheit	3	N	
9	MSHJPN	MH	D000224	Addison病	3	N	
10	MSHPJN	SY	D000224	副腎性黒皮症	3	N	

# MRREL (sample rows)

	1	2	3	4	5	6				
	CUI1	AUI1	STYPE1	REL	CUI2	AUI2	STYPE2			
1	C0001403		CUI	RB	C0001621		CUI			
2	C0001403	A0019738	AUI	SY	C0001403	A0049628	AUI			
3	C0001403	A2922421	SCUI	CHD	C0085859	A2977940	SCUI			
4	C0001403	A6326321	SCUI	RO	C0688490	A6339383	SCUI			
5	C0001403	A0019743	AUI	PAR	C0935495	A1988358	AUI			

	7	8	9	10	11	12	13	14	15
	RELA	RUI	SRUI	SAB	SL	RG	DIR	SUPPRESS	CVF
1		R02837989		MTH			N	N	
2		R18849683		MSH	MSH			N	
3	isa	R19859511	1658795027	SNOMEDCT	SNOMEDCT	0	Y	N	
4	may_treat	R27600039		NDFRT	NDFRT			N	
5	has_member	R08110401		PSY	PSY			N	

# MRHIER (sample rows)

	1	2	3	4	5	6
	CUI	AUI	CXN	PAUI	SAB	RELA
1	C0001403	A0019740	1	A0020270	MSH	
2	C0001403	A0019740	2	A0028022	MSH	
3	C0001403	A0019743	3	A1988358	PSY	member_of_cluster
4	C0001403	A2922421	1	A3307650	SNOMEDCT	isa
5	C0001403	A2922421	2	A3307650	SNOMEDCT	isa

	7	8	9
	PTR	HCD	CVF
1	A0434168.A2367943.A2366890.A0135391.A0054194.A0020267.A0020270	C19.053.264.263	
2	A0434168.A2367943.A2366890.A0135391.A0072566.A0028022	C20.111.163	
3	A0449751.A1988279.A1988358		
4	A3684559.A3886745.A2880798.A3398606.A3399335.A3398961.A2872359. A2872360.A3307650		
5	A3684559.A3886745.A2880798.A3398606.A3399335.A3398961.A2872359. A2933400.A2989549.A3307650		



# MRSAT (sample rows)

	1	2	3	4	5	
	CUI	LUI	SUI	METAUI	STYPE	CODE
1	C0001403	L0001403	S0010792	A0019738	AUI	D000224
2	C0001403	L0001403	S0010794	A6326321	SCUI	C712
3	C0001403	L0001403	S0354372	A2922421	SAUI	363732003
4	C0001403			R15742591	SRUI	
5	C0001403				CUI	

	7	8	9	10	11	14	15
	ATUI	SATUI	ATN	SAB	ATV	SUPPRESS	CVF
1	AT15321482		DID	MSH	D000224	N	
2	AT33411754		MESH_UI	NDFRT	D000224	N	
3	AT24166602		DESCRIPTIONS TATUS	SNOMEDCT	0	N	
4	AT27438950		REFINABILITY	SNOMEDCT	0	N	
5	AT02925340		ST	MTH	R	N	

# MRDEF

CUI	AUI	ATUI	SATUI	SAB	DEF	SUPPRESS	CVF
C0001403	A0019740	AT15061584		MSH	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.	N	

# Metathesaurus Metadata Files

---

- MRFILES: files authority
  - MRCOLS: data elements
  - MRDOC: data element valid values
  - MRSAB: source vocabularies
  - MRRANK: source/term types ranks
- 
- Generate UMLS release documentation
  - Full set of metadata files in subset

# Metathesaurus Indexes

---

- Word: each word in each unique string in each language
  - Heart disease yields
    - ENG| disease |C0024117|L0024117|S0058458|
    - ENG| heart |C0018787|L0018787|S0047194|
- Normalized word: each normalized word in each unique English string
  - Disease, diseases, diseased yields
    - ENG| disease |C0024117|L0024117|S0058458|
- Normalized string obstructive lung disease
  - ENG| disease lung obstructive |C0024117| L0024117| ... S0058458|

# Different Formats for Different Purposes

---

- Original Release Format (ORF)
- **Lexical** View
- Natural language processing
- Rich Release Format (RRF)
- Atomic View
- Greater specificity
  - Facilitates maintenance
  - Enables other types of changes and applications
- RRF Browser in MetamorphoSys

# How do I use the Metathesaurus?

---

- Identify useful sources
- Identify useful content from specific sources
- Create customized Metathesaurus using MetamorphoSys
  
- Use UMLS Release Documentation to understand file content and structure

# Metathesaurus Use Cases

---

- **Information retrieval**
- **Thesaurus construction**
- **Natural language processing**
- **Automated indexing**
- **Electronic patient records**
  
- **Distribution of health data standard vocabularies**

# UMLS – MeSH mapping file

---

Used in MEDLINE/PubMed searching

Based on synonymy

...

myocardial infarction|attack coronary

myocardial infarction|attack heart (nos)

myocardial infarction|cardiac infarction

myocardial infarction|cardiac infarction, nos

myocardial infarction|cardiac; infarction

myocardial infarction|heart attack

myocardial infarction|heart attack, nos

myocardial infarction|heart attacks

myocardial infarction|heart infarction



# Electronic Health Data Standards



## Health IT Home

### General Information

### Federal Efforts

### Consumer Awareness

### Focus Areas

## Health IT

Health IT initiatives harness current and emerging information technologies to improve patient safety and convenience while reducing the cost of providing care. Some of the numerous benefits of health IT initiatives will include: a reduction in medical errors, avoidance of costly duplicate testing, and elimination of unnecessary hospitalizations.

The benefits of health IT initiatives range from consumer convenience, as patients will not have to fill out repetitive paper work, to life-saving early detection of an infectious disease outbreak, as anonymous data from emergency rooms is sent to public health systems instantly. These initiatives aid in fulfilling the President's goal for most Americans to have electronic health records by the year 2014.

### American Health Information Community (AHIC) Topics

The American Health Information Community (AHIC) is a federal advisory body, chartered to make recommendations to the Secretary of HHS on how to accelerate the development and adoption of Health IT. The Community will make recommendations to the Secretary of HHS to enable advancement in four areas of focus by the end of 2006.

- [Consumer Empowerment](#) - Make available a consumer-directed and secure electronic record of health care registration information and a medication history for patients.
- [Chronic Care](#) - Allow the widespread use of secure messaging, as appropriate, as a means of communication between doctors and patients about care delivery.
- [Biosurveillance](#) - Enable the transfer of standardized and anonymized health data from the point of health care delivery to authorized public health agencies within 24 hours of its collection.
- [Electronic Health Records](#) - Create an electronic health record that includes laboratory results and interpretations, that is standardized, widely available and secure.

## Highlighted Federal Efforts

### Health & Human Services

[Health & Human Services \(HHS\)](#) - HHS is facilitating the development of standards for Health IT systems that will improve patient care and increase efficiency across the health care system. HHS, through several of its agencies, also provides funding to organizations engaged in building and testing Health IT systems, standards and projects.

### Department of Defense

[Department of Defense \(DoD\)](#) - Currently, thousands of military medical providers use the DoD's electronic health record system, AHLTA, and nearly 300,000 outpatient visits are captured digitally every week. DoD's vision is to provide each patient with a continuously updated digital medical record from the

<http://www.hhs.gov/healthinformationtechnology/>

[Veterans Health Administration \(VHA\)](#) - The VHA is a division of the U.S. Department of Veteran's Affairs, and provides care for over five million veterans of the United States Armed Services. The VHA's

## Mission Statement

"We will make wider use of electronic records and other health information technology to help control costs and reduce dangerous medical errors."

-- President Bush, January 31, 2006

".to link all health records through an interoperable system that protects privacy as it connects patients, providers and payers. Resulting in fewer medical mistakes, less hassle, lower costs and better health."

-- HHS Secretary Mike Leavitt

## News

**08/22/2006** - [President Directs Federal Agencies to Provide Health Care Quality and Price Information for Consumers](#)

**05/17/2006** - [American Health Information Community Approves First Set of Recommendations](#)

**01/17/2006** - [HHS Announces Pilot Project Launched to Expand Electronic Prescribing](#)

## Speeches

**10/07/2005** - [Remarks at the First Meeting of the American Health Information Community](#)

- HHS Secretary Mike Leavitt

## Testimony

**09/28/2005** - [Full Committee Hearing on Healthcare and the IT Revolution](#)

-- Dr. David J. Brailer, HHS National Coordinator for Health IT

## Interactive

**RSS** [News Releases](#)

News and information sent as issued.

[Help with RSS](#)

## Personal Experiences



## Health Information Technology at NLM

[Printer-friendly Version](#)

NLM is the central coordinating body for clinical terminology standards within the Department of Health and Human Services (HHS). NLM works closely with the Office of the National Coordinator for Health Information Technology (ONC) to ensure NLM's efforts are aligned with HHS Secretary Mike Leavitt and President Bush's goal for the nationwide implementation of an interoperable health information technology infrastructure to improve the quality and efficiency of health care.

### Health Data Standards

NLM supports the development, enhancement, and distribution of clinically specific vocabularies to facilitate the exchange of clinical data and improve retrieval of health information.

- Clinical Vocabularies supported, licensed, or developed by NLM:
  - [SNOMED CT](#) | [LOINC](#) | [RxNorm](#)
- Uniform distribution mechanism for HIPAA and clinical vocabulary standards through the [UMLS Metathesaurus](#)

NLM coordinates efforts to develop mappings between HIPAA code sets and standard clinical vocabularies.

- SNOMED CT to ICD-9-CM | SNOMED CT to CPT | [LOINC to CPT](#) ~~XXXX~~

NLM promotes harmonization between standards.

- [Contract with HL7](#) to align HL7 message standard with CHI standard vocabularies and create implementation guides for exchange of entire EHRs.

### System Development Tools

- [UMLS Resources](#)
- [Language and Knowledge Processing](#)
- [Image Processing](#)

### Grants and Funding

### Education and Training

- [Funded Support for Academic Training in Biomedical Informatics & Bioinformatics](#) - Formal programs and individual fellowships are offered to assist medical informaticians in pursuit of a degree.
- [Medical Informatics Training Program](#) - Support for visiting scientists and students to participate in research at the Lister Hill National Center for Biomedical Communications. [Medical Informatics Course at Woods Hole](#) sponsored by NLM and the Woods Hole Marine Biology Laboratory.

### Key Reports Supported or Produced by NLM

- [Ending the Document Game](#) (2005)
- [Networking Health: Prescriptions for the Internet](#) (2000)
- [Evaluating the Coverage of Controlled Health Data Terminologies: Report on the Results of the NLM/AHCPR Large Scale Vocabulary Test](#) (1997)
- [For the Record: Protecting Electronic Health Information](#) (1997)
- [Making a Powerful Connection](#) (1995)

[Back to Top](#)

<http://www.nlm.nih.gov/healthit.html>

### Related Sites

[HealthIT](#) (U.S. Health Information Technology Web Site)

[Office of the National Coordinator for Health Information Technology \(ONC\)](#)

### NIH Initiatives

[NIH Re-engineering the Clinical Research Enterprise](#)

[CaBIG \(Cancer Biomedical Informatics Grid\)](#)

### NLM Participation in Standards Development

[Federal Health Architecture/Consolidated Health Informatics \(FHA/CHI\)](#)  
(committee member)

[National Committee on Vital and Health Statistics \(NCVHS\)](#)  
(staff to Standards and Security subcommittee)

[Healthcare Information Technology Standards Panel \(HITSP\)](#) (board member)

[Public Health Data Standards Consortium \(PHDSC\)](#)  
(founding member)

# Electronic Health Data Standards

---

- Include standard vocabularies
- Key element of health information technology infrastructure for:
  - Effective decision support
  - Safe, evidence-based, and coordinated health care
  - Cost-effective care
  - Increased/informed choice
  - More efficient clinical, public health, and health services research
  - Timely public health and bioterrorism surveillance

# NLM-led Support for Development and Maintenance

---

- **1999** – LOINC (lab tests/instrument observations) - contract support
- **2002** – RxNorm (clinical drugs) - direct development
- **2003** – SNOMED CT contract & license for U.S-wide use *(as distributed by NLM in UMLS)*
- **2004** – NLM designated central coordinating body for clinical terminology standards within U.S. Dept. of Health and Human Services (HHS)
- **2006** – Draft LOINC to CPT mapping distributed

# SNOMED CT

---

- CHI recommended standard
- Comprehensive clinical vocabulary
- Substantially increased Metathesaurus content
  - Concepts: ~300K (+37K to Metathesaurus)
  - Descriptions: ~737K (+350K to Metathesaurus)
  - Relationships: 1.3M
  - Generic drugs, History table
- Updates
  - January SNOMED CT update → Spring UMLS
  - July SNOMED CT update → Fall UMLS

# RxNorm

---

- CHI recommended standard
- Developed to address **missed synonymy** in UMLS clinical drugs
- Provides standardized (normalized) forms of U.S. clinical drugs
- Supports
  - **effective sharing of drug data across systems**
  - **electronic health record (EHR)**
  - **computerized physician order entry (CPOE)**

# Mappings

---

- Link specific terms or codes between two sources
- NLM given the responsibility for funding, coordinating, and/or performing official mappings between **standard clinical terminologies** and **HIPAA code sets**
- Represented in Metathesaurus files MRMAP, MRSMAP



# Mapping projects planned/underway

---

- Clinical standards → HIPAA code sets:
  - LOINC → CPT (**Draft now available for testing!**)
  - SNOMED CT → ICD-9-CM, ICD-10-CM
  - SNOMED CT → CPT
- SNOMED CT → “other” vocabularies:
  - International Classification of Functioning, Disability and Health (ICF)
  - International Classification of Primary Care (ICPC)
  - Medcin
  - Medical Dictionary for Regulatory Affairs (MedDRA)
  - Medical Subject Headings (MeSH®)
  - Nursing Vocabularies (NIC, NOC, NANDA)

# Key NLM Assumptions about Mapping

---

- Mapping is an R & D problem
- Iteration required to build highly functional maps
- Testing, validation, and use in real world settings
- Participants include **producers of vocabularies** on both ends, and prospective **users, recipients** of maps
- Mapping will prompt changes/corrections to vocabulary content and update schedules
- Mappings must be updated when either end is updated
- Mappings will be distributed in the UMLS (not exclusively)

# Draft LNC215 to CPT2005 Mappings

---

- Use case assumes
  - LOINC codes used to order lab tests, observations
  - CPT codes used for billing (one-way)
- 2000+ most common mappings contributed
- Created by Intermountain Healthcare
- Reviewed by Regenstrief Institute and AMA
- Future developments to include:
  - Radiology tests
  - Document names (consults, progress notes)
  - More Laboratory LOINC mappings
  - Clinical findings (vital signs, height, weight)

# Mapping examples – LOINC → CPT

---

LOINC Code

CPT Code

1795-4 AMYLASE:CCNC:PT:FLU:QN →

82150 Amylase

1798-8 AMYLASE:CCNC:PT:SER:QN →

(blood or urine)

1799-6 AMYLASE:CCNC:PT:UR:QN →

LNC215\_TO\_CPT2005

MRSMAP

---

C1704201|LNC|ATX104051417|  
|**1795-4**|CODE|RN||**82150**|CODE||

C1704201|LNC|ATX104051418|  
|**1798-8**|CODE|RN||**82150**|CODE||

C1704201|LNC|ATX104051419|  
|**1799-6**|CODE|RN||**82150**|CODE||

# LNC215\_TO\_CPT2005 MRCONSO

---

■ **C1704201** | ENG | P | L6107611 | PF | S7015223 | Y  
| A10924448 | | | LNC | XM | NOCODE | **LNC21**  
**5 to CPT2005 Mappings** | 0 | N | |

# LNC215\_TO\_CPT2005 MRSMAP

---

C1704201|L6107611|S7015223|A10924448|AUI|NOCODE|...

ATX104051002||MAPSETNAME|LNC|LNC215 to CPT2005  
Mappings|N||

ATX104051003||MAPSETTYPE|LNC|LOINC has associated CPT  
code|N||

ATX104051007||FROMVSAB|LNC|LNC215|N||

ATX104051008||TOVSAB|LNC|CPT2005|N||

...

ATX104051018||MAPSETVERSION|LNC|Sept 2006|N||



## Unified Medical Language System

[UMLS Home](#)

[Home](#) > [Biomedical Research & Informatics](#) > [UMLS](#)

9/30/06: **Draft LOINC to CPT Mappings** now available for download from the UMLSKS. ●●● New to the UMLS? [Register now.](#)

### ▶ [About the UMLS Resources](#)

Metathesaurus; Semantic Network; SPECIALIST Lexicon and lexical programs; MetamorphoSys

### ▶ [Accessing UMLS Knowledge Sources](#)

Metathesaurus license; Semantic Network; SPECIALIST Lexicon; DVD

### ▶ [Knowledge Source Server](#)

Download files; searching; additional tools and resources

### ▶ [Documentation](#)

### ▶ [Help](#)

Training; contact us; FAQs; listserv

### About

NLM's Unified Medical Language System (UMLS) project develops and distributes multi-purpose, electronic "Knowledge Sources" and associated lexical programs for system developers. Researchers will find the UMLS products useful in investigating knowledge representation and retrieval questions.

### Metathesaurus Source Vocabularies

- [SNOMED CT](#)
- [LOINC](#)
- [RxNorm](#)
- [MeSH](#)
- [List of Sources](#)
- [Source FAQs](#)
- [Mappings](#)

### More Resources

- [Metathesaurus License](#)
- [Tools](#)
- [Learning Resources](#)
- [MetaMap Transfer \(MMTx\)](#)
  
- [Archives](#)



# UMLS Knowledge Source Server (UMLSKS)

# UMLS Knowledge Source Server

---

- Licensed users access online:  
<http://umlsks.nlm.nih.gov>
- Web search interface  
for the three Knowledge Sources
- Application Programming Interface (API)

# Using the UMLSKS

- Create login ID and password using the form on the right
- Each UMLS license can have multiple login IDs
- NLM does not maintain a copy of the password or Login ID
- Passwords may be reset, not login IDs

**UMLS Knowledge Source Server (UMLSKS)**  
UMLSKS Version 2.0  
UMLS Release: 2001 2002AC 2002AC 2002AC 2002AA 2002AD 2002AC 2004AA 2004AD 2002AC 2004AA 2004AD 2002AC 2005AA  
U.S. National Library of Medicine  
Lester Hill National Center for Biomedical Communications (LHNC/BC)

**About the UMLSKS**

- ▶ [Login](#)
- ▶ [Overview](#)
- ▶ [Frequently Asked Questions](#)
- ▶ [UMLS Metathesaurus®](#)
- ▶ [Semantic Network](#)
- ▶ [SPECIALIST Lexicon](#)

**Documentation**

- ▶ [User's Guide](#)
- ▶ [Developer's Guide](#)
- ▶ [UMLS Documentation Set](#)

**UMLSKS Registered Users**

Login ID:   
Password:   
**Login**  
[Forgot your password?](#)

**What's New**

- ▶ The primary database server for the UMLSKS has been restored and users should now have access to all available UMLS releases. We're sorry for any inconvenience this may have cause. (September 22, 2006 10:00 ET)
  - ▶ 2006AC UMLS release is available for searching and for download within the UMLSKS. For more information, click [here](#). (July 31, 2006 13:45PM ET)
  - ▶ On June 14, 2006 licensees who have not submitted their annual report (as

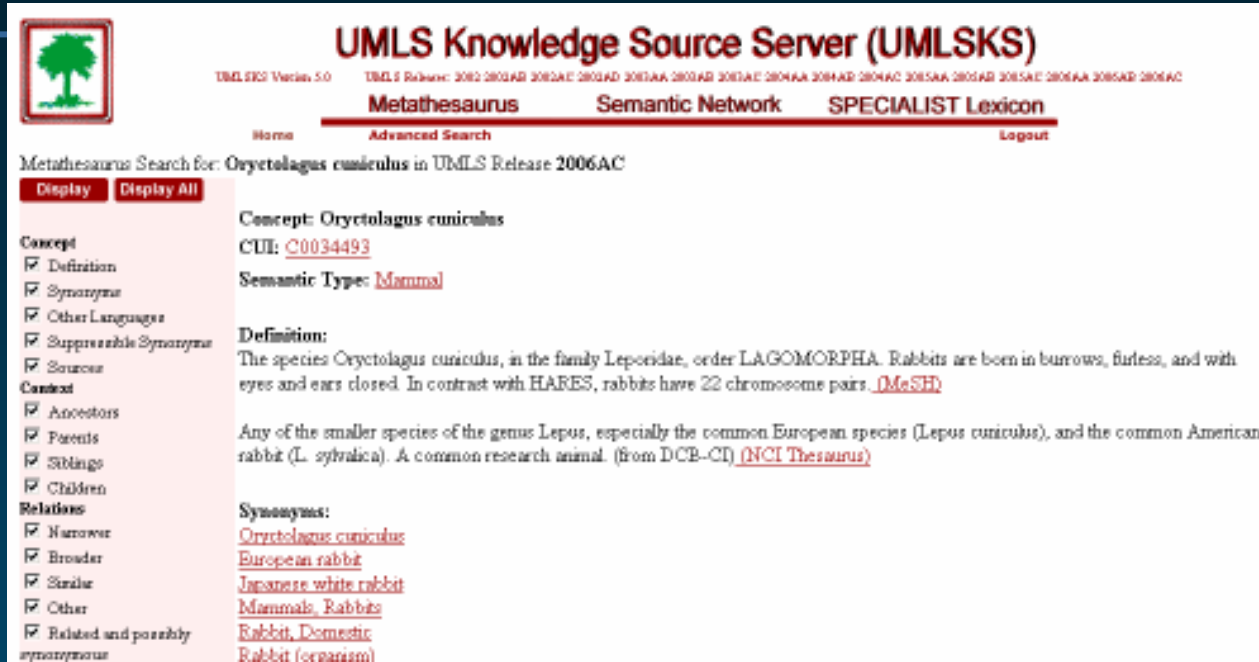
**New UMLSKS Users**

The UMLSKS is restricted to registered users. If you are not a registered user and would like to use the resources of the UMLS, please read the [UMLS license agreement](#), obtain a UMLS license number, and fill in the form below to request a UMLSKS user account.

Login ID:   
Password:   
Confirm Password:   
First Name:   
Last Name:   
Email:   
UMLS License #:

**Request Account**

# UMLS Knowledge Source Server (UMLSKS) Basic Concept Report



The screenshot displays the UMLS Knowledge Source Server (UMLSKS) interface. At the top, there is a logo of a tree and the title "UMLS Knowledge Source Server (UMLSKS)". Below the title, there are navigation links: "Metathesaurus", "Semantic Network", and "SPECIALIST Lexicon". The main content area shows a search result for "Oryctolagus cuniculus" in UMLS Release 2006AC. The result includes a "Concept" section with a CUI of "C0034493" and a "Semantic Type" of "Mammal". A "Definition" section provides a detailed description of the species. A "Synonyms" section lists various terms associated with the concept. On the left side, there is a sidebar with a list of categories and checkboxes, including "Concept", "Definition", "Synonyms", "Other Languages", "Suppressible Synonyms", "Sources", "Context", "Ancestors", "Parents", "Siblings", "Children", "Relations", "Narrower", "Broader", "Strike", "Other", and "Related and possibly synonymous".

Includes:

- Definitions
- Synonyms
- Other Languages
- Sources
- Context Information
- Relations
- Co-occurring Concepts

Can be used to find:

- Source IDs
- Other Language
- Synonyms
- Term Type Information
- Term Variants

# Metathesaurus Advanced Search Options

---

## ■ Focused Search

Search by release, source vocabulary or language

Partial string matching with right or left truncation

## ■ XML Query

Send Standard API commands or specialized requests

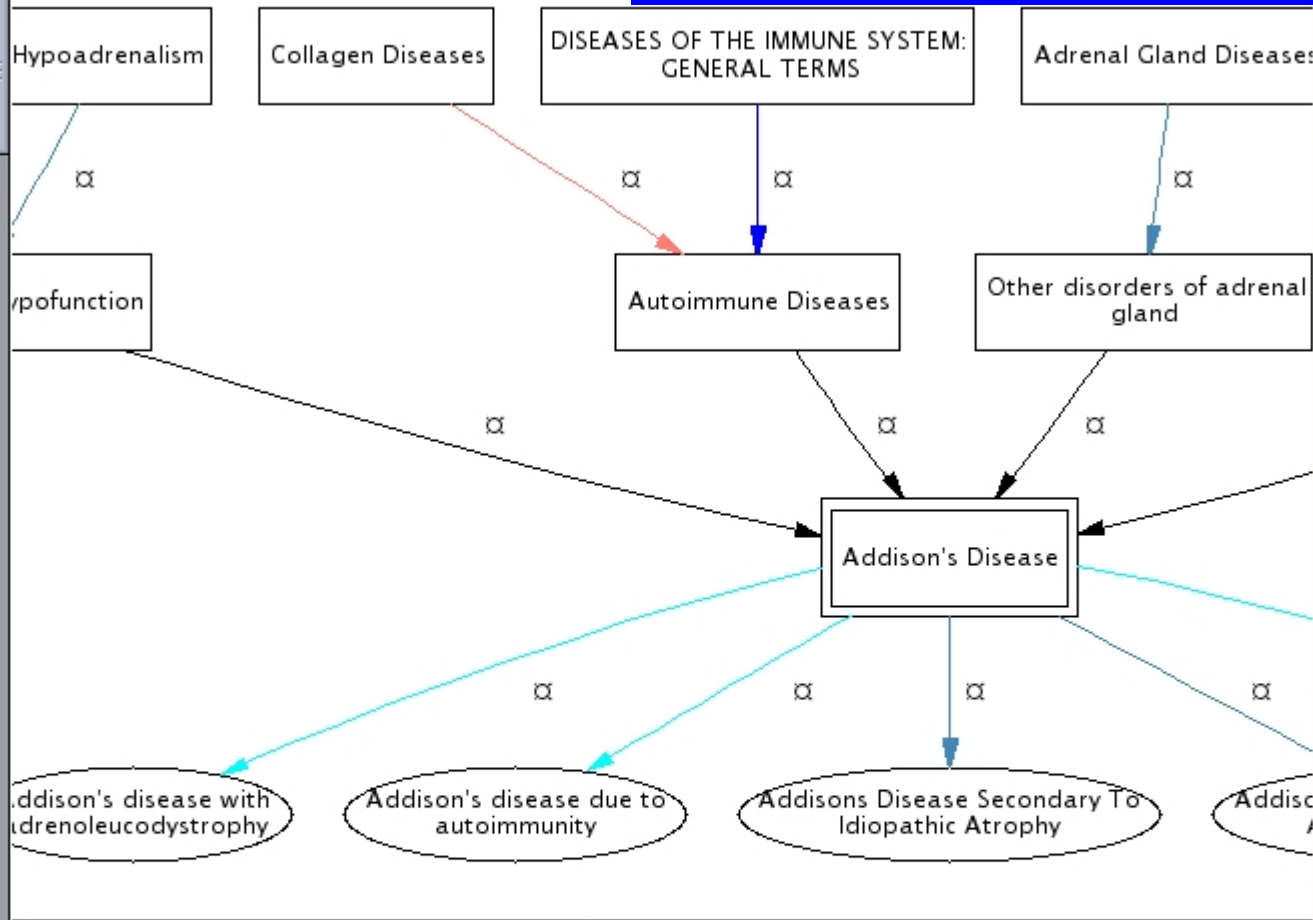
Run against the UMLSKS, results returned in tab delimited text

## ■ Request relational records

View records by CUI for the files MRCONSO, MRDEF, MRSTY, MRCXT, MRREL, MRSAT, & MRCOC as tab delimited text

# Semantic Navigator

- Siblings**
- Concepts & Ideas**
- Clinical Syndromes
- Disorders**
- Acquired Immunodeficiency Syndrome
  - Acute adrenal insufficiency
  - Addisonian crisis
  - Adrenal atrophy
  - Adrenal calcification
  - Adrenal hemorrhage
  - Adrenal infarction
  - Adrenal insufficiency due to adrenal metastasis
  - Adrenogenital Syndrome
  - Allergic arthritis
  - Angelman Syndrome
  - Asperger syndrome
  - Autoerythrocyte sensitivity



- Other Related Concepts**
- Disorders**
- Addisonian crisis
  - Autoimmune Syndrome Type II, Polyglandular
  - Tuberculosis
  - Tuberculosis of adrenal glands
  - Tuberculous Addison's disease
- (5 other related)
- Co-occurring Concepts**
- Anatomy**
- Adrenal Cortex [14]
  - Adrenal Glands [17]
  - Liver [2]
  - Tears body substance [2]
  - X Chromosome [3]
- Chemicals & Drugs**

**BCI** **Addison's Disease** **LEGEND \***

Start again Apply new parameters

Restrict to vocabulary: Show all

Highlight vocabulary: Nothing

UMLS data: UMLS\_2002

Type of hierarchical:  All  Parent/Child only

**Similar Concepts**

- Adrenal cortical hypofunction

(1 concept)

**Closest MeSH Terms**

**Main Headings**

- Addison's Disease

# Semantic Network search

- Enter search term or select from drop down box
- View Type, TUI, and Definition
- Select checkboxes to see Meta Concepts, Relations And Raw Records

The screenshot shows the UMLS Knowledge Source Server (UMLSKS) Semantic Network search interface. At the top, there is a logo of a tree and the text "UMLS Knowledge Source Server (UMLSKS)". Below this, there are navigation links: "Home", "Logout", "Metathesaurus", "Semantic Network", and "SPECIALIST Lexicon". A note indicates that the site is best viewed with Netscape 6.x and above or IE 5.x and above. The main section is titled "Semantic Network" and contains a search form. The form has a text input field, a "Semantic Types" dropdown menu (currently set to "Alga"), and a "Semantic Relations" dropdown menu (currently set to "adjacent to"). A "Find" button is located below the form. To the right of the form, there are checkboxes for "Meta Concepts", "Relations", and "Raw Records". Below the search form, the results are displayed in a table. The first row shows the "Entity" column with a list of semantic types: "Physical Object", "Organism", "Plant", "Alga", "Fungus", "Virus", "Rickettsia or Chlamydia", and "Bacterium". The second row shows the "Semantic Type: Alga" with its TUI (T003) and a definition: "Definition: A chiefly aquatic plant that contains chlorophyll, but does not form embryos during development and lacks vascular tissue."

# SPECIALIST Lexicon Search

- Search by entering the complete term
- View record in relational or unit record format

**UMLS Knowledge Source Server (UMLSKS)**  
v3.8 UMLS Release: 2803 2082AD 2803AC 2082AD 2803AA 2085AD 2803AC 2081AA 2804AD 2081AC 2803AA 2085AD 2803AC 2085AA 2804AD

[Metathesaurus](#)   [Semantic Network](#)   [SPECIALIST Lexicon](#)   [Logout](#)

---

## Relational Records for "heart attack"

**LRWD**

```
heart|E0431577|
attack|E0431577|
attacks|E0431577|
```

**LRAGR**

```
E0431577|heart attack|noun|count(the_sing)|heart attack|heart atta
E0431577|heart attacks|noun|count(the_plus)|heart attack|heart att
```

**LRTYP**

```
E0431577|heart attack|noun|reg|
```

---

View the unit record for "[heart attack](#)"

**UMLS Knowledge Source Server (UMLSKS)**  
3.8 UMLS Release: 2803 2082AD 2803AC 2082AD 2803AA 2085AD 2803AC 2081AA 2804AD 2081AC 2803AA 2085AD 2803AC 2085AA

[Metathesaurus](#)   [Semantic Network](#)   [SPECIALIST Lexicon](#)   [Logout](#)

---

## Specialist Lexical Record

```
{base=heart attack
entry=E0431577
cat=noun
variants=reg
}
```

View [heart attack](#) in relational format.



# Application Programming Interfaces

---

- Remote server at NLM
- Local application connected through

## Java RMI

- Java-based applications
- Developer's Guide: Chapter 3
- Set of Java classes (part of the UMLSKS API download)
- Detailed *Javadoc* documentation online and with API download

## TCP/IP socket

- XML-based queries
- Developer's Guide: Chapter 5
- XML schema
- Socket server
  - Host: [umlsks.nlm.nih.gov](http://umlsks.nlm.nih.gov)
  - Port: 8042

You MUST register your IP Address!

# New UMLSKS expected Spring 2007

---

- Web services based
- Create your own tabs
- Forms auto-complete, learning as you go
- Looking for beta testers

MetamorphoSys

# MetamorphoSys

---

- Multi-platform Java software
- Included in each UMLS release
- Unzips native Metathesaurus compressed files
- Installs Knowledge Sources to local storage
- Customizes a local Metathesaurus

# Download from UMLSKS ...

---

- High speed Internet connection required
- Files must be stored in the same folder
- 2006AD UMLS Files
  - **mmsys.zip (zipped MetamorphoSys application)**
  - 2006AD-1-meta.nlm (compressed Metathesaurus data)
  - 2006AD-2-meta.nlm (compressed Metathesaurus data)
  - 2006AD-otherks.nlm (compressed Semantic Network and SPECIALIST Lexicon)
  - 2006AD.CHK
  - 2006AD.MD5
  - Copyright\_Notice.txt
  - **README.txt**

- ... **or request DVD**
- [umls\\_support@nlm.nih.gov](mailto:umls_support@nlm.nih.gov)
- Include your license number
- Run MetamorphoSys from DVD

# Machine Requirements

---

- A fast CPU – 1 GHz or higher
- 1 GB RAM recommended (512 MB min.)
- 6x (or better) DVD drive
- **13** GB minimum free disk space
  
- Runs on Sun Solaris 8 & 9, Windows XP, NT, 2000, Linux, and Mac
- 1-10 hours run time on platforms tested

# Customize the Metathesaurus

---

- Use MetamorphoSys
  - To comply with terms of license agreement
  - To remove unhelpful or harmful content
  - To change default settings (precedence, output)
- Customization is critical and requires understanding of:
  - Selected vocabularies
  - Functional requirements, purpose and perspective
- Technical expertise requires multidisciplinary team

# How MetamorphoSys Works

---

- Removes all information from all relational files from excluded vocabularies
  - atoms, strings, relations, hips, attributes, mappings
- Applies additional options selected by user
  - Changes to suppressibility or precedence
- Produces custom set of Metathesaurus relational files reflecting selected criteria
- Log file records subset details
- Output directory: set of Metathesaurus files



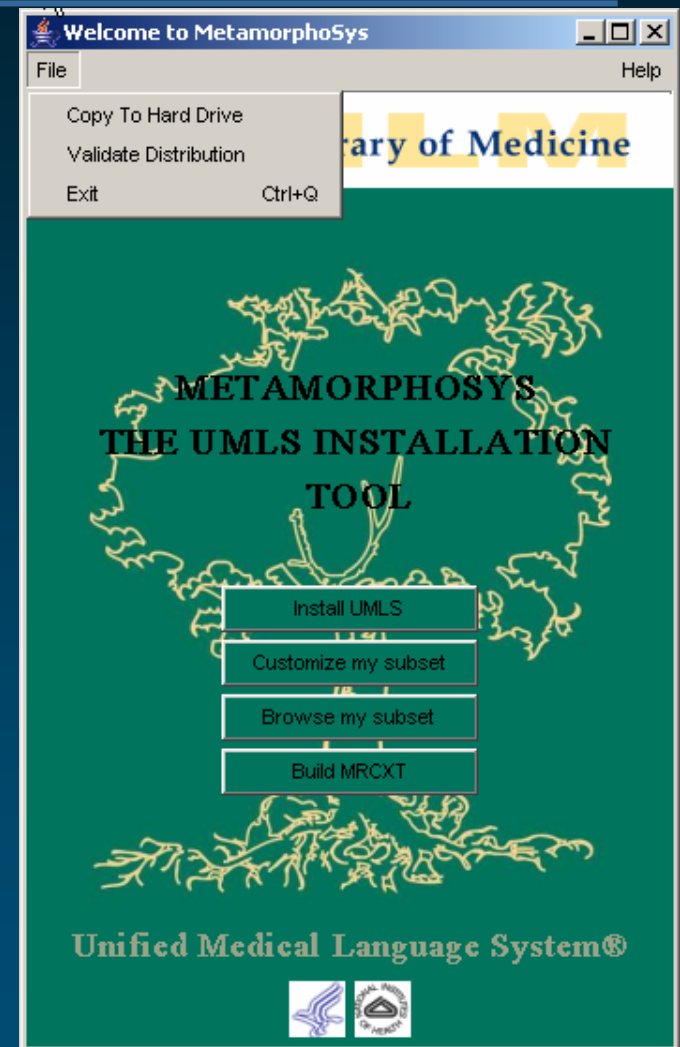
# MetamorphoSys Welcome Screen

## ■ Validate Distribution

- ensures that all files were downloaded
- process takes @ 30 minutes
- writes validation.log file

## ■ Copy DVD to hard drive

- Copies all files to local storage
- Allows multiple people to use one DVD
- May improve run time



How do I?

# Specify sources for a customized subset?

Input Options | Output Options | **Source List** | Precedence | Suppressibility

Indicate below to INCLUDE or EXCLUDE selected sources.  
SELECTED sources appear on a dark background, e.g., AIRRHEUM, 1993.

To undo selections and return to default source list, select "Reset Source List" from Reset menu at top.

Select sources to EXCLUDE from subset  
 Select sources to INCLUDE in subset

Sources to Exclude

Exclude or Include

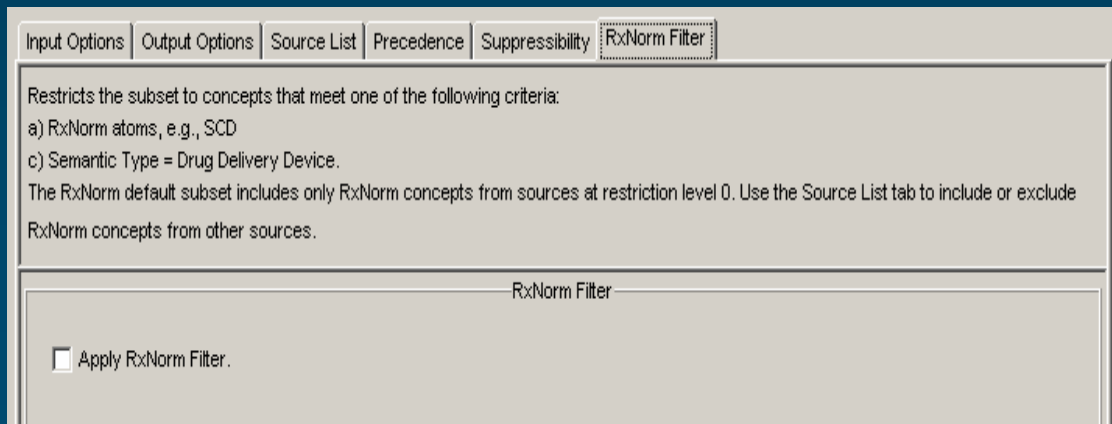
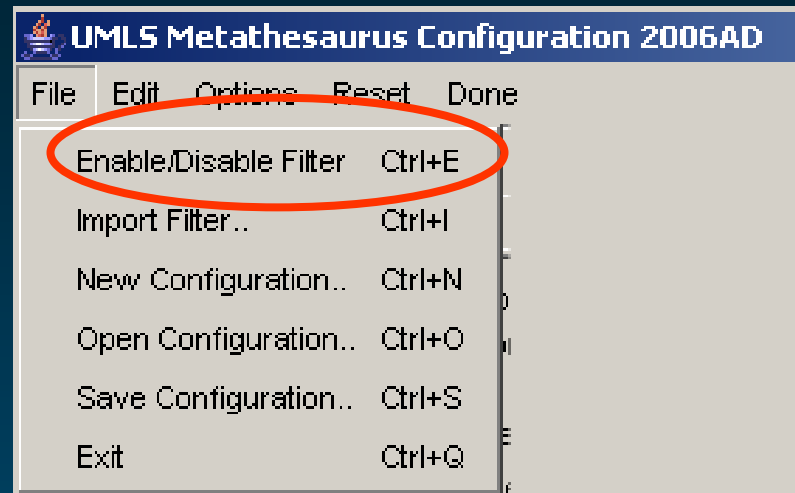
Sources to Exclude

Full Source Name	Source Abbreviation	Source Family	Language	Level
AIRRHEUM, 1993	AIR93	AIR	ENG	0
Alternative Billing Concepts, 2004	ALT2004	ALT	ENG	3
Alcohol and Other Drug Thesaurus, 2000	AOD2000	AOD	ENG	0
Beth Israel Vocabulary, 1.0	BI98	BI	ENG	2
Canonical Clinical Problem Statement System, 1999	CCPSS99	CCPSS	ENG	3
Clinical Classifications Software, 2003	CCS2003	CCS	ENG	0
Current Dental Terminology 2005 (CDT-5), 5	CDT5	CDT	ENG	3
COSTAR, 1989-1995	COSTAR_89-95	COSTAR	ENG	0
Medical Entities Dictionary, 2003	CPM2003	CPM	ENG	2
Physicians' Current Procedural Terminology, Spanish Translation, 2001	CPT01SP	CPT	SPA	3
Physicians' Current Procedural Terminology, 2005	CPT2005	CPT	ENG	3
CRISP Thesaurus, 2004	CSP2004	CSP	ENG	0
COSTART, 1995	CST95	CST	ENG	0
Diseases Database, 2000	DB00	DB	ENG	3
German translation of ICD10, 1995	DMDICD10_1995	ICD10	GER	1
German translation of UMDNS, 1996	DMDUMD_1996	UMD	GER	1
DSM-III-R, 1987	DSM3R_1987	DSM3R	ENG	3
DSM-IV, 1994	DSM4_1994	DSM4	ENG	3
DXplain, 1994	DXP94	DXP	ENG	0
Gene Ontology, 2004_12_20	GO2004_12_20	GO	ENG	0
HCPCS Version of Current Dental Terminology 2005 (CDT-5), 5	HCDT5	CDT	ENG	3

Highlighted rows are excluded from the subset.

# Extract an RxNorm subset?

- On the File menu select “Enable/Disable Filter”  
*Selecting any filter opens a new tab for that filter.*
- Make your selections on the tab for that filter

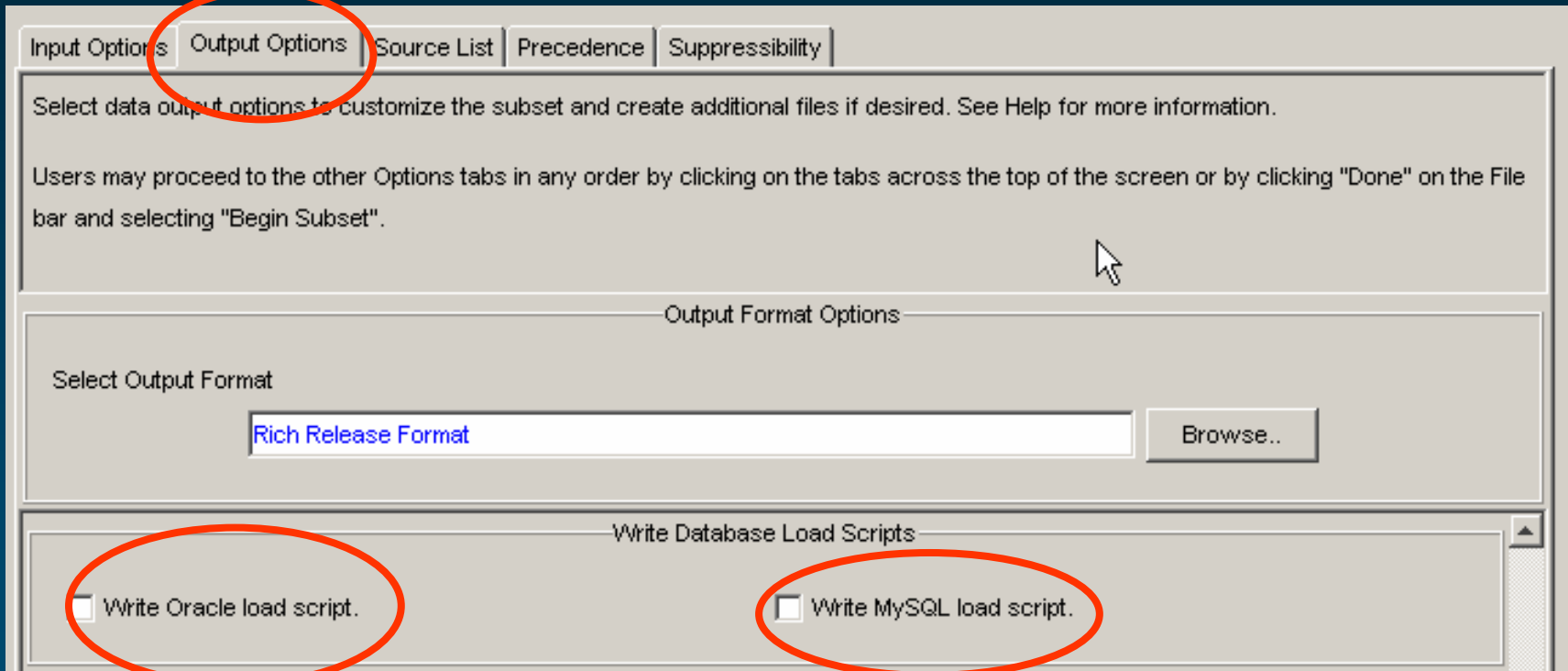


- Other filters allow users to remove terms:

Attribute Type List  
Relationship Type List  
Semantic Types List  
Source Term Types

# Create a custom database load script?

- Select the Output Options tab
- Check the box next to the type of load script you require



- Other Options on this tab

# Change how preferred term is set?

- Select the Precedence tab
- Cut and paste or drag and drop source and term types to reflect your preferred ranking order

Input Options | Output Options | Source List | **Precedence** | Suppressibility

Change the ranking of sources and their associated term types to create concept names that are more useful in your local application. Concept names are determined by the term with highest ranking source/term type.

To move rows, either cut and paste rows, or drag and drop.

When you have made your selections on this tab, you may proceed to the other Options tabs in any order. If you are finished customizing

Precedence

Full Source Name	Source Abbreviation	Term Type
UMLS Metathesaurus	MTH	PN
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	MH
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	TQ
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	EP
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	EN
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	XQ
Medical Subject Headings, MSH2004_2003_12_12	MSH2004_2003_12_12	NM
RXNORM Project, META2004AB	RXNORM_04AB	SCD

# Remove specific term types from subset?

This screen contains the Suppressibility Filter, which specifies source/term type combinations to be suppressed. Users can customize the subset by selecting and deselecting source/term type combinations. See Help for more information.

Users may proceed to the other Options tabs in any order by clicking on the tabs across the top of the screen or by clicking "Done" on the File bar and selecting "Begin Subset".

Select One or More Suppressible Term Types

Source	Source Abbreviation	Term Type
AIRRHEUM, 1993	AIR93	DI
AIRRHEUM, 1993	AIR93	FI
AIRRHEUM, 1993	AIR93	HT
AIRRHEUM, 1993	AIR93	SY

- Select Suppressibility tab
- Highlight rows to exclude

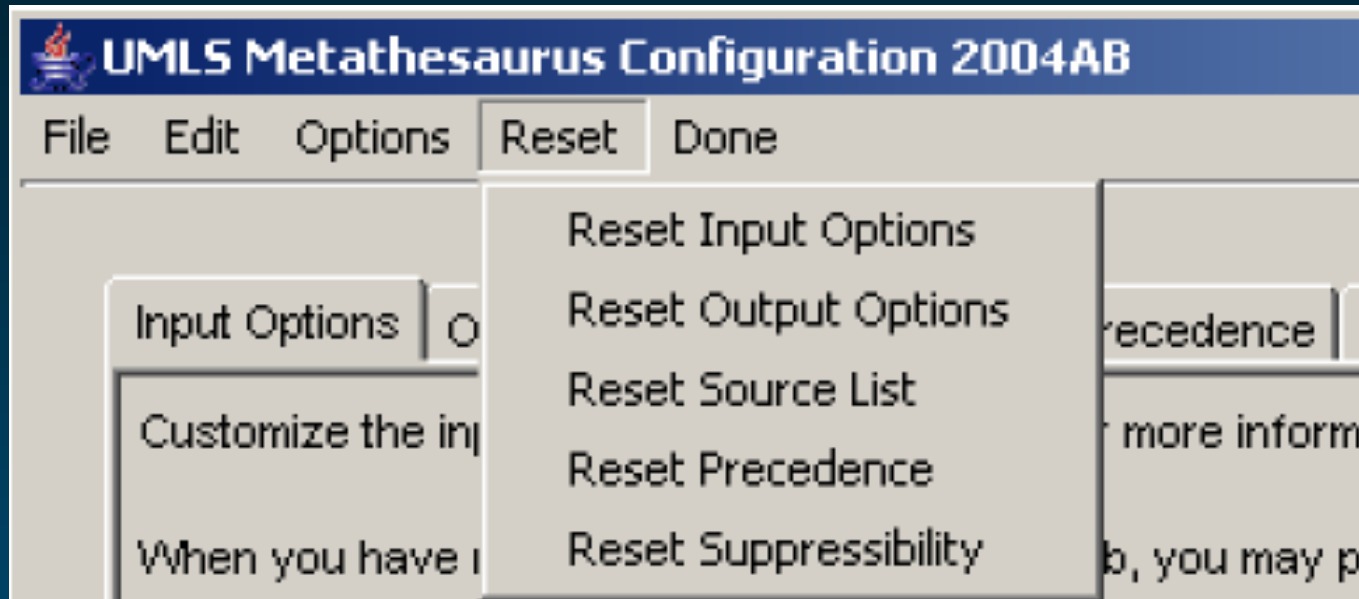
- Select Options → Advanced Suppressibility Options
- Select the term types to remove

Advanced Suppressibility Options

- Remove Source Term Type Suppressible Data.
- Remove Editor Assigned Suppressible Data.
- Remove Obsolete Data.

Done

# Reset default MMSYS Options?

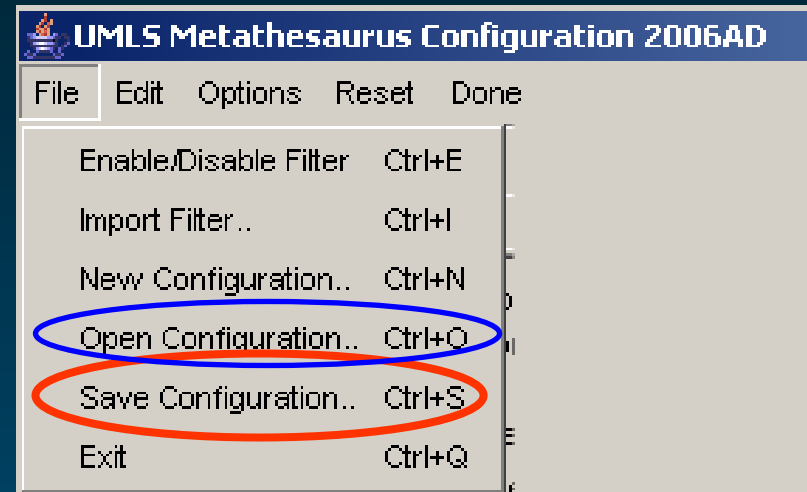


- Returns all filters to default selections
- Default selections in "mmsys.prop.default file" in config folder
- mmsys.prop.default contains properties in last run



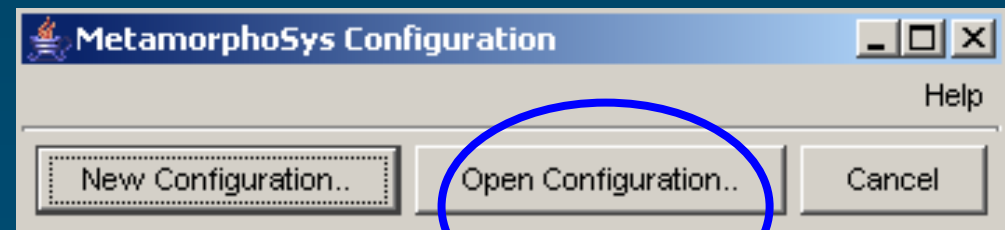
# Ensure all team members have the same subset?

- On File menu "Save Configuration"
- Share configuration file with team members
- Have team members select "Open Configuration" from File Menu



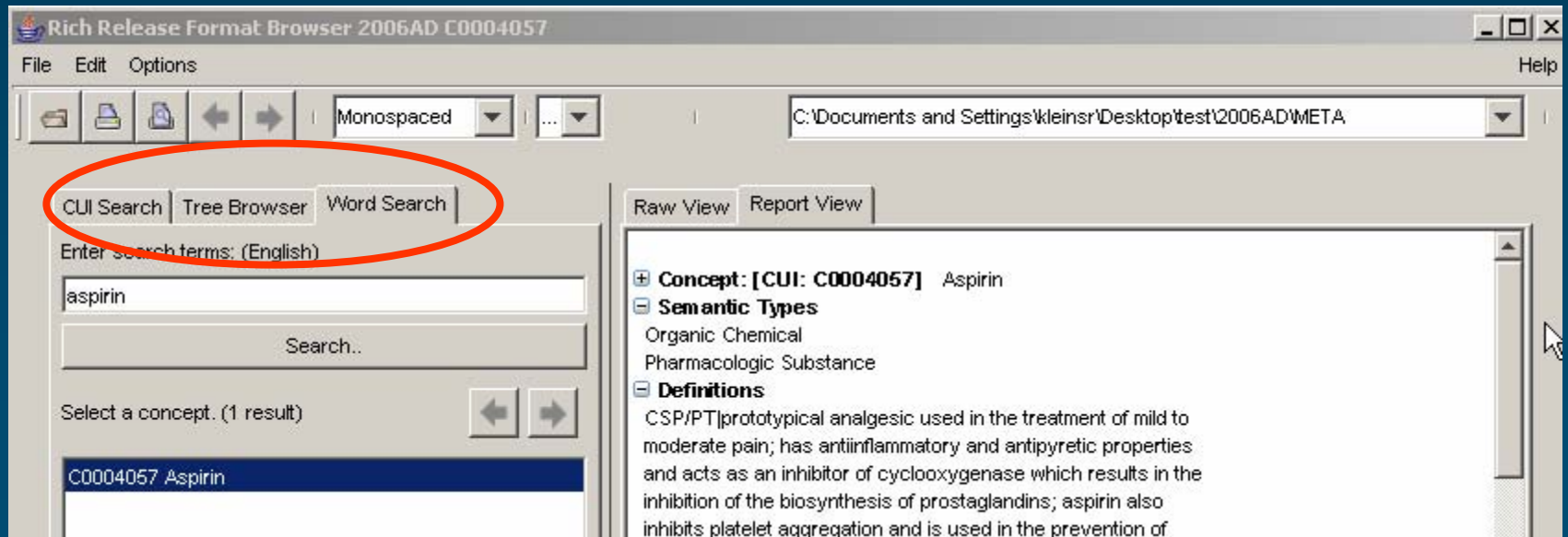
/or/

Select "Open Configuration" from Configuration Screen



# Search for a term in my RRF subset?

- Select "Browse my Subset" from welcome screen
- Browse to your subset location
- Search by term, string or CUI
- Reports include:
  - Hyperlinked concepts
  - Raw data view
  - Attributes and Relations



Rich Release Format Browser 2006AD C0004057

File Edit Options Help

Monospaced

C:\Documents and Settings\kleinsr\Desktop\test\2006AD\META

CUI Search Tree Browser Word Search

Enter search terms: (English)

aspirin

Search..

Select a concept. (1 result)

C0004057 Aspirin

Raw View Report View

⊕ **Concept: [CUI: C0004057] Aspirin**

⊖ **Semantic Types**

- Organic Chemical
- Pharmacologic Substance

⊖ **Definitions**

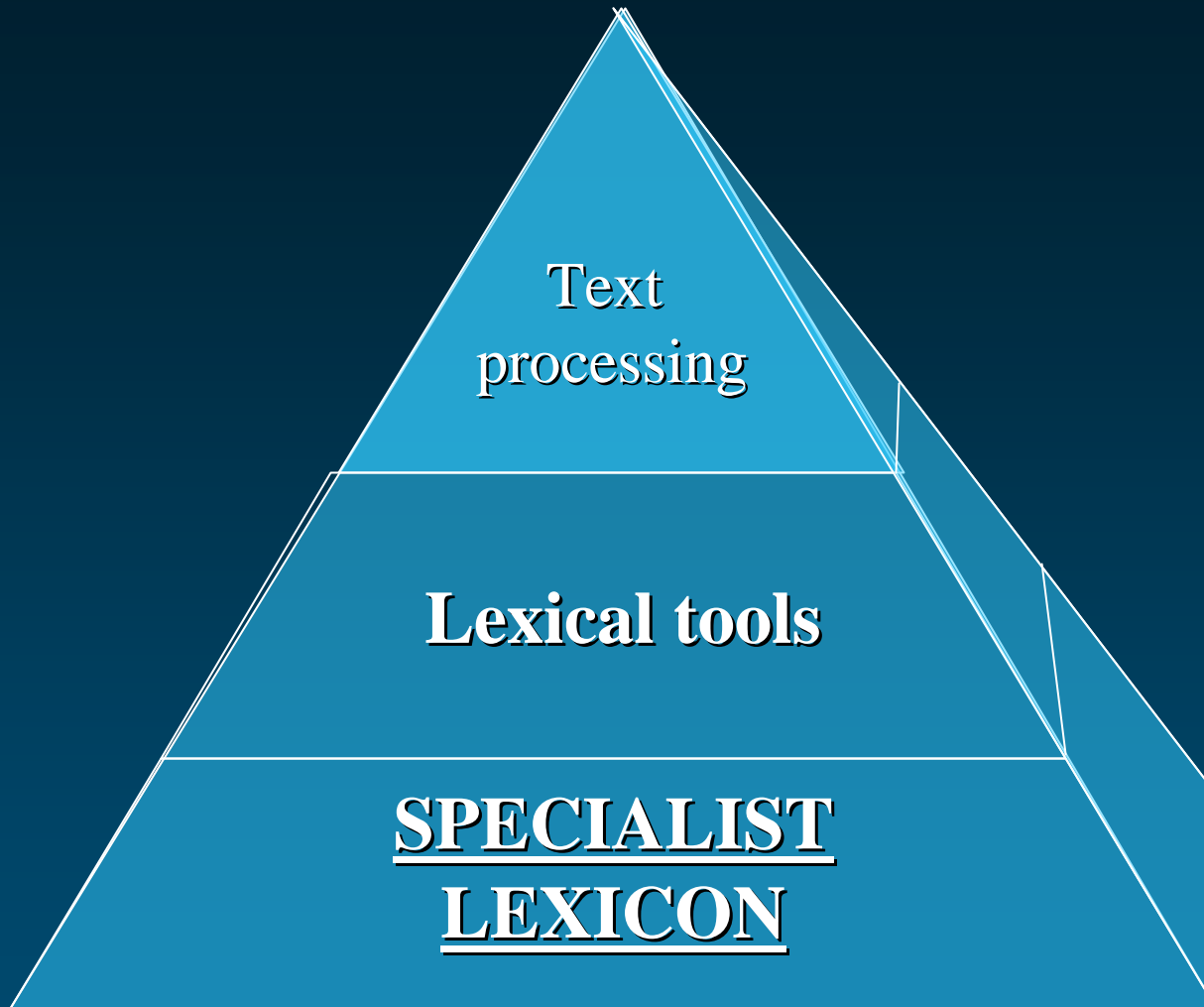
CSP/PT]prototypical analgesic used in the treatment of mild to moderate pain; has antiinflammatory and antipyretic properties and acts as an inhibitor of cyclooxygenase which results in the inhibition of the biosynthesis of prostaglandins; aspirin also inhibits platelet aggregation and is used in the prevention of

# More information and help

---

- MetamorphoSys Documentation at:  
<http://www.nlm.nih.gov/research/umls/meta6.html>
- Readme file on the DVD or downloaded from the UMLSKS
- Help Menu from any page in MetamorphoSys

# SPECIALIST Lexicon and Lexical Tools



---

# MetaMap and MetaMap Transfer (MMTx)

# Outline

---

- Purpose of MetaMap/MMTx
- The MetaMap/MMTx Algorithm
- Availability
- Demo

# Purpose of MetaMap/MMTx

---

- To map biomedical text to concepts in the UMLS Metathesaurus
- Or, equivalently, to find Metathesaurus concepts in text
  
- MMTx was created to provide a distributable version of MetaMap



# MetaMap/MMTx Example

■ Text: Termination of clinical trials: the beta-blocker heart attack trial ...

## ■ Concepts

- Termination
- Clinical Trials
- Adrenergic beta-Antagonists
- Myocardial Infarction

Heart attack (Myocardial Infarction)

Clinical Trials  
Trial (Clinical Trials)

Beta-blocker (Adrenergic beta-Antagonists)

# The MetaMap/MMTx Algorithm

---

- Parsing
  - Using SPECIALIST minimal commitment parser, SPECIALIST lexicon, a part of speech tagger
- Variant generation
  - Using SPECIALIST lexicon, Lexical Variant Generation (LVG)
- Candidate retrieval
  - From the Metathesaurus
- Candidate evaluation
- Mapping construction

# Technical Details and Availability

---

- <http://skr.nlm.nih.gov/>
  - Click 'Research Information' for technical details
- <http://mmtx.nlm.nih.gov/>
  - Click 'Documentation' and 'Prerequisites'
- Use restrictions
  - Sign UMLS license agreement; then access MetaMap and download MMTx using UMLS ID/password
  - Respect UMLS constituent vocabulary copyrights

# MetaMap Demo

---

<http://skr.nlm.nih.gov>

# Example: normal processing

---

Phrase: "lung cancer."

Meta Candidates (8):

1000 Lung Cancer (Malignant neoplasm of lung) [Neoplastic Process]

1000 Lung Cancer (Carcinoma of lung) [Neoplastic Process]

861 Cancer (Malignant Neoplasms) [Neoplastic Process]

861 Lung [Body Part, Organ, or Organ Component]

861 Cancer (Cancer Genus) [Invertebrate]

861 Lung (Entire lung) [Body Part, Organ, or Organ Component]

861 Cancer (Specialty Type - cancer) [Biomedical Occupation or Discipline]

768 Pneumonia [Disease or Syndrome]

Meta Mapping (1000):

1000 Lung Cancer (Carcinoma of lung) [Neoplastic Process]

Meta Mapping (1000):

1000 Lung Cancer (Malignant neoplasm of lung) [Neoplastic Process]

# Example: Variants (-v)

---

Phrase: "lung cancer."

lung cancer [noun] variants (n=1):

lung cancer{[noun], 0=[]}

lung [noun] variants (n=9):

lung{[noun], 0=[]} lungs{[noun], 1="i"} pneumonia{[noun], 5="ds"}  
pneumoniae{[noun], 5="ds"} pneumonias{[noun], 5="ds"}  
pneumonic{[adj], 2="s"} pulmonal{[adj], 4="ss"} pulmonary{[adj],  
2="s"} pulmonic{[adj], 2="s"}

cancer [noun] variants (n=4):

cancer{[noun], 0=[]} cancerous{[adj], 3="d"} cancers{[noun], 1="i"}  
carcinomatous{[adj], 2="s"}

...

# Example: Compound mappings

---

Phrase: "obstructive sleep apnea."

Meta Candidates (8):

...

Meta Mapping (1000):

1000 Obstructive sleep apnoea (Sleep Apnea, Obstructive) [Disease or Syndrome]

Meta Mapping (901):

827 Obstructive (Obstructed) [Functional Concept]

901 Apnea, Sleep (Sleep Apnea Syndromes) [Disease or Syndrome]

Meta Mapping (851):

827 Obstructive (Obstructed) [Functional Concept]

827 Sleep [Organism Function]

827 APNOEA (Apnea) [Pathologic Function]

...

**without**  
**--best\_mappings\_only**

# Example: show sources (-G)

---

Phrase: "scorpion sting."

Meta Candidates (4):

1000 Scorpion sting {MDR,DXP} [Injury or Poisoning]

861 Sting (Sting Injury  
{MTH,MSH,MDR,RCD,SNM,SNOMEDCT,SNMI,WHO}) [Injury or  
Poisoning]

694 Scorpion (Scorpions  
{LCH,MSH,MTH,SNM,SNOMEDCT,SNMI,CSP,RCD,NCBI}) [Invertebrate]

694 SCORPION (Scorpion antigen {MTH,LNC}) [Immunologic Factor]

Meta Mapping (1000):

1000 Scorpion sting {MDR,DXP} [Injury or Poisoning]



# Example: restrict to sources (-GR LCH)

---

Phrase: "scorpion sting."

Meta Candidates (1):

694 Scorpion (Scorpions {LCH}) [Invertebrate]

Meta Mapping (694):

694 Scorpion (Scorpions {LCH}) [Invertebrate]

# Example: restrict to STs (-J neop)

---

Phrase: "lung cancer."

Meta Candidates (3):

1000 Lung Cancer (Malignant neoplasm of lung) [Neoplastic Process]

1000 Lung Cancer (Carcinoma of lung) [Neoplastic Process]

861 Cancer (Malignant Neoplasms) [Neoplastic Process]

Meta Mapping (1000):

1000 Lung Cancer (Carcinoma of lung) [Neoplastic Process]

Meta Mapping (1000):

1000 Lung Cancer (Malignant neoplasm of lung) [Neoplastic Process]

Questions?

# UMLS Documentation and Support

---

- UMLS homepage

- <http://umlsinfo.nlm.nih.gov/>

- UMLSKS homepage

- <http://umlsks.nlm.nih.gov>

- UMLSUSERS-L

- subscribe to discussion list

- NLM Customer Service email:

- [custserv@nlm.nih.gov](mailto:custserv@nlm.nih.gov)

# Explore

---

- Register: sign the license agreement
- Create UMLSKS account
- Explore Knowledge Sources
- Download files or request DVD
- Create subsets using MetamorphoSys



Center  
Drive

Library  
of  
Medicine

**Thank you**