

# Einführung in das Unified Medical Language System (UMLS)

Marcus Nitzschke

29. November 2011

# Inhaltsverzeichnis

- 1 Problemstellung
- 2 UMLS
- 3 Metathesaurus
- 4 Semantic Network
- 5 SPECIALIST Lexicon
- 6 Zusammenfassung

# Inhaltsverzeichnis

- 1 Problemstellung
- 2 UMLS
- 3 Metathesaurus
- 4 Semantic Network
- 5 SPECIALIST Lexicon
- 6 Zusammenfassung

# Heterogenität

- Vielzahl an (medizinischen) Vokabularen und Thesauri
- starke Heterogenität untereinander
- Unterschiede in
  - Größe
  - Begriffsordnung
  - Terminologie
  - Syntax
  - Sprachen

# Beispiel

|              | SNOMED CT | Gene Ont. | MeSH    |
|--------------|-----------|-----------|---------|
| # Terms      | 800.000   | 33.000    | 181.141 |
| # Concepts   | 300.000   | -         | -       |
| # Relationen | 1.000.000 | -         | -       |

atrial fibrillation – auricular fibrillation

Addison's disease – Addison Disease

# Beispiel

|              | SNOMED CT | Gene Ont. | MeSH    |
|--------------|-----------|-----------|---------|
| # Terms      | 800.000   | 33.000    | 181.141 |
| # Concepts   | 300.000   | -         | -       |
| # Relationen | 1.000.000 | -         | -       |

atrial fibrillation – auricular fibrillation

Addison's disease – Addison Disease

# Folgen

- verteiltes, dezentrales Wissen
- Redundanzen
- keine formalen Strukturen zur semantischen Verarbeitung zwischen Quellen

# Inhaltsverzeichnis

- 1 Problemstellung
- 2 UMLS**
- 3 Metathesaurus
- 4 Semantic Network
- 5 SPECIALIST Lexicon
- 6 Zusammenfassung



# Einführung

- Unified Medical Language System
- entwickelt von der National Library of Medicine, USA
- 1986 erstmals vorgestellt
- Integration über 100 vorhandener maschinenlesbarer Vokabulare und Thesauri
- Bereitstellung von Entwicklungstools
- selber keine Endanwendung!

# Inhaltsverzeichnis

- 1 Problemstellung
- 2 UMLS
- 3 Metathesaurus**
- 4 Semantic Network
- 5 SPECIALIST Lexicon
- 6 Zusammenfassung

# Allgemein

- Integriert Terme einzelner Vokabulare in Konzepte
- Integration durch Mappings zwischen einzelnen Quellen
- Quell-Informationen der Terme bleiben erhalten
- über 5Mio. Terme
- mehrsprachig
  - 62% englisch

# Mappings

- vorwiegend auf syntaktischen Eigenschaften beruhend
  - exakte Matches
  - normalisierte Terme
- zusätzliches Expertenwissen

# Konzepte

- Konzepte umfassen
  - Synonyme
  - Übersetzungen
  - semantischen Typ
- jedes Konzept besitzt min. einen semantischen Typ
- jedes Konzept besitzt “bevorzugten Term” der aus gerankten Quellen hervorgeht

# Unique Identifier

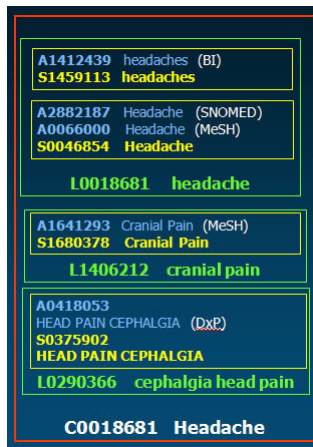
Concept Unique Identifier (CUI) Konzept-IDs

Lexical (term) Unique Identifier (LUI) IDs für unters. Terme  
desselben Konzeptes

String Unique Identifier (SUI) IDs für lexikalische Varianten der  
Terme

Atom Unique Identifier (AUI) IDs für Ursprungs-Term inkl. Quelle

# Beispiel



# Verwendung

- Metathesaurus stellt nur integrierte Rohdaten zur Verfügung
- Einbindung in andere Anwendungssysteme
- Dateiformate:
  - RRF (Rich Release Format)
  - ORF (Original Release Format)

```
C0001175|ENG|P|L0001175|V0|S0010340|Y|A0019182||  
M0000245|D000163|MSH|PM|D000163|Acquired  
Immunodeficiency Syndromes|0|N|1792|
```



# Inhaltsverzeichnis

- 1 Problemstellung
- 2 UMLS
- 3 Metathesaurus
- 4 Semantic Network**
- 5 SPECIALIST Lexicon
- 6 Zusammenfassung

# Allgemein

- Katalog von semantischen Typen und Relationen
- Konzepte werden durch Relationen miteinander in Verbindung gesetzt
- bildet high-level Netzwerk/Ontologie der biomedizinischen Domäne

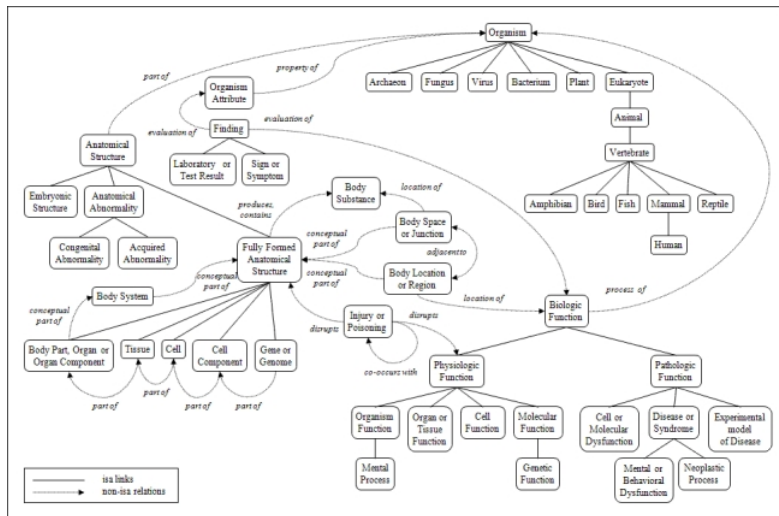
# Semantische Typen

- 135 Semantische Typen
- Entity
  - “Gen”
  - “Kohlenhydrat”
  - “Krankheit”
- Event
  - “Geistiger Prozess”
- Konzepten wird der präziseste Typ zugewiesen

# Semantische Relationen

- 54 Semantische Relationen
- hierarchisch:
  - isa
  - part of
- nicht-hierarchisch:
  - Physically related to
  - Spatially related to
  - Temporally related to
  - ...
- Bsp.: “Arzneimittel behandeln Krankheiten”
- nicht alle Verbindungen sinnvoll, siehe Übersicht

# Übersicht



# Inhaltsverzeichnis

- 1 Problemstellung
- 2 UMLS
- 3 Metathesaurus
- 4 Semantic Network
- 5 SPECIALIST Lexicon**
- 6 Zusammenfassung

# Allgemein

- allgemeines Lexikon biomedizinischer Begriffe
- über 250.000 Terme
- Quellen:
  - Metathesaurus
  - MEDLINE abstracts
  - ...
- ermöglicht NLP-Anwendungen

# Einträge

- Eigenschaften der Einträge

- Morphologische

- POS (11 Tags)

- Orthographische

- Syntaktische

- Wortstamm

```
{ base=anaesthetic  
spelling_variant=anesthetic  
entry=E0008769  
cat=noun  
variants=reg }
```



# Lexical Tools

- Normalizer - Norm
  - Normalisiert Terme auf einheitliches Format
  - Terme werden in Lexicon und Konzepten genutzt
- word index generator - Wordind
  - Tokenizer - teilt Zeichenketten in Wörter
  - produziert Metathesaurus Wortindex
- lexical variant generator - lvg
  - Beugen, Konjungieren
  - Interpunktion

## weitere Tools

- MetamorphoSys
  - erlaubt das Erstellen von Teilmengen des Metathesaurus
  - Bsp.: alle deutschen Terme; alle Terme vom Typ "Krankheit"
- UMLS Terminology Services (UTS)
  - Zugriff auf Wissensquellen
  - Durchsuchen der Quellen

# Inhaltsverzeichnis

- 1 Problemstellung
- 2 UMLS
- 3 Metathesaurus
- 4 Semantic Network
- 5 SPECIALIST Lexicon
- 6 Zusammenfassung**

## UMLS

- ... integriert verschiedenste Vokabulare und Thesauri
- ... stellt integrierte Daten maschinenlesbar zur Verfügung

## UMLS

- ... integriert verschiedenste Vokabulare und Thesauri
- ... stellt integrierte Daten maschinenlesbar zur Verfügung
- ... ist keine Endanwendung

## UMLS

- ... integriert verschiedenste Vokabulare und Thesauri
- ... stellt integrierte Daten maschinenlesbar zur Verfügung
- ... ist keine Endanwendung
- ... bringt jedoch zahlreiche Tools mit

## UMLS

- ... integriert verschiedenste Vokabulare und Thesauri
- ... stellt integrierte Daten maschinenlesbar zur Verfügung
- ... ist keine Endanwendung
- ... bringt jedoch zahlreiche Tools mit

# Quellen



Barbara Heller.

Unified Medical Language System (UMLS).

Medizinische Begriffs- und Dokumentationssysteme,  
WS2000/2001.



Olivier Bodenreider.

The Unified Medical Language System (UMLS): integrating  
biomedical terminology.

*Nucleic Acids Research, 2004, Vol. 32, Database issue, 2003.*



Halit Erdogan, Olivier Bodenreider, and Esra Erdem.

Finding Semantic Inconsistencies in UMLS Using Answer Set  
Programming.

*In Proceedings of the Twenty-Fourth AAAI Conference on  
Artificial Intelligence (AAAI-10), 2010.*



E. Jiménez-Ruiz, B.Cuenca Grau, R. Berlanga, and

L. Horrocks