

Medizinische Begriffs- und Dokumentationssysteme WS 2000/2001

Unified Medical Language System (UMLS) Teil I

16.01.2001

Barbara Heller

Email: heller@imise.uni-leipzig.de

Terminologische Hausgemeinschaft

**Integrated Vocabularies &
Internet-/Intranet-based
Applications**

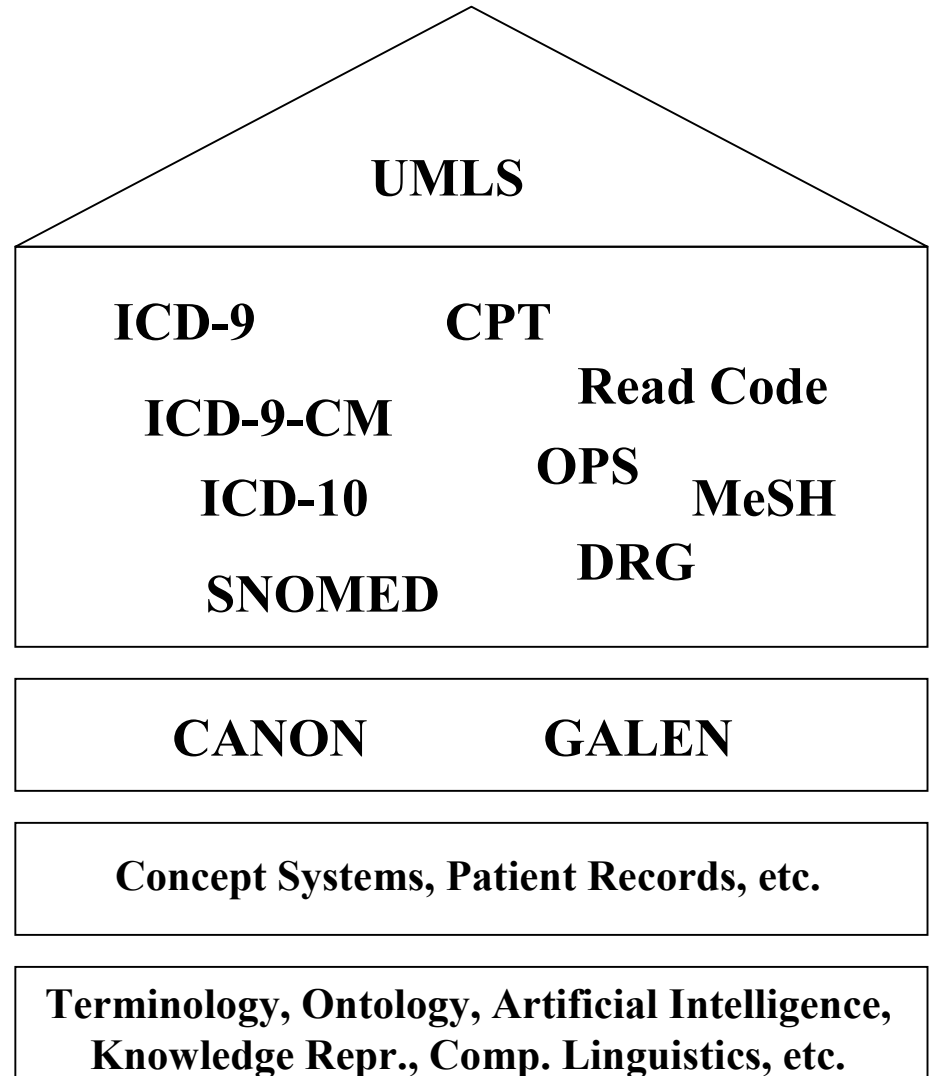
**Diseases and Procedure
Classifications I & II**

Other Medical Vocabularies

**Formal Reconstruction of
Medical Vocabularies**

Standards

Basic Sciences



Zwei unterschiedliche Ansätze

- Unified Medical Language System (**UMLS**)
National Library of Medicine (NLM)
Ziel:
 - Integration von maschinenlesbaren Source Vocabularies (MeSH, ICD, SNOMED, usw.) über einen Metathesaurus sowie damit dotierten Wissensquellen wie MEDLINE
 - Konzeptuelle Verknüpfung von Benutzeranfragen und relevanter Informationen
- Generalized Architecture for Language Encyclopedias, and Nomenclatures in Medicine (**GALEN**)
EU-Projekt mit Koordinator A. Rector in Manchester
Ziel:
 - Medical concept representation as foundation for the next generation of multilingual coding systems
 - Terminology server for data entry, multilingual language generation and code conversion

UMLS - Einführung

- Forschungsprogramm der UNITED STATES National Library of Medicine (NLM) in Bethesda, USA
- Entwicklung in Kooperation mit zahlreichen Partnern seit 1986
- Seit 1990 jährliche Herausgabe des UMLS-Systems auf CD
- Seit 1995 kurzfristige Aktualisierung über WWW-Server-Implementierung
- Informationen unter: www.nlm.nih.gov/research/umls/
 - Ausführliche Dokumentation
 - Umfassende Bibliographie-Liste
 - Links zu UMLS-Applications
 - Lizenzvertrag

Ausgangsproblem

Vielzahl medizinischer Thesauren

- mit unterschiedlichem Umfang
- mit uneinheitlicher Begriffsordnung
- mit uneinheitlicher Terminologie
- mit uneinheitlichen Schreibweisen
- in unterschiedlichen Sprachen

Schwächen existierender Begriffssysteme

Benutzung von Codes zur Identifikation von (komplexen) Begriffen sowie zur Repräsentation von Begriffsrelationen und von Begriffswissen

- Präkoordinierte Konzepte in fixierten Hierarchien
 - ➔ kombinatorische Explosion, Redundanz
- Hierarchisierung von nicht-hierarchischen Begriffsbeziehungen
- Vermengung von generischen und partitiven Beziehungen
- Keine Kontrolle der Kombination von Begriffen
- Implizite Kontextabhängige Semantik
- Fehlen formaler Strukturen
 - ➔ erschwert eine rechnergestützte semantische Interpretation vzw. Verarbeitung

Defizite traditioneller Kodiersysteme (SNOMED)

1	DB-62110	Diabetic nephropathy	
2	DB-61000	Diabetes mellitus	
	G-C025	Causing	
	D7-11000	Nephropathy	
3	DB-61000	Diabetes mellitus	} Fehlende formal Sprache
	G-C025	Causing	
	DF-00000	Disease	
	G-C006	Located in	
	T-71000	Kidney	

Defizite traditioneller Kodiersysteme (SNOMED)

4 DB-62110 Diabetic nephropathy
G-C006 Located in
T-71000 Kidney

Redundanz



Normalform

5 DB-62110 Diabetic nephropathy
G-C006 Located in
T-11000 Bone

Inkohärenz



**Sensible
Statements**

6 DB-62110 Diabetic nephropathy
Kimmelstiel-Wilson syndrome
Diabetic glomerulopathy

**Differen-
zierte
Bedeutung**

Ziel des UMLS

- Zusammenfassung medizinischer Begriffsordnungen in einer gemeinsamen Datenstruktur durch Integration maschinenlesbarer Source Vocabularies (ICD, MeSH, ...)
- Integrierte Sicht und erleichterter Zugriff auf:
 - Biomedizinische Literatur
 - Krankenblattdaten
 - Faktendatenbanken
 - wissensbasierte Systeme

Komponenten von UMLS

UMLS Metathesaurus

Integration von biomedizinischen
Thesauren, kontrollierten
Vokabularen und Klassifikationen

➔ **Concepts, Terms, Strings**

UMLS Information

Sources Map

Verweise auf elektronische Daten-
banken mit relevanter Information

UMLS Semantic Network

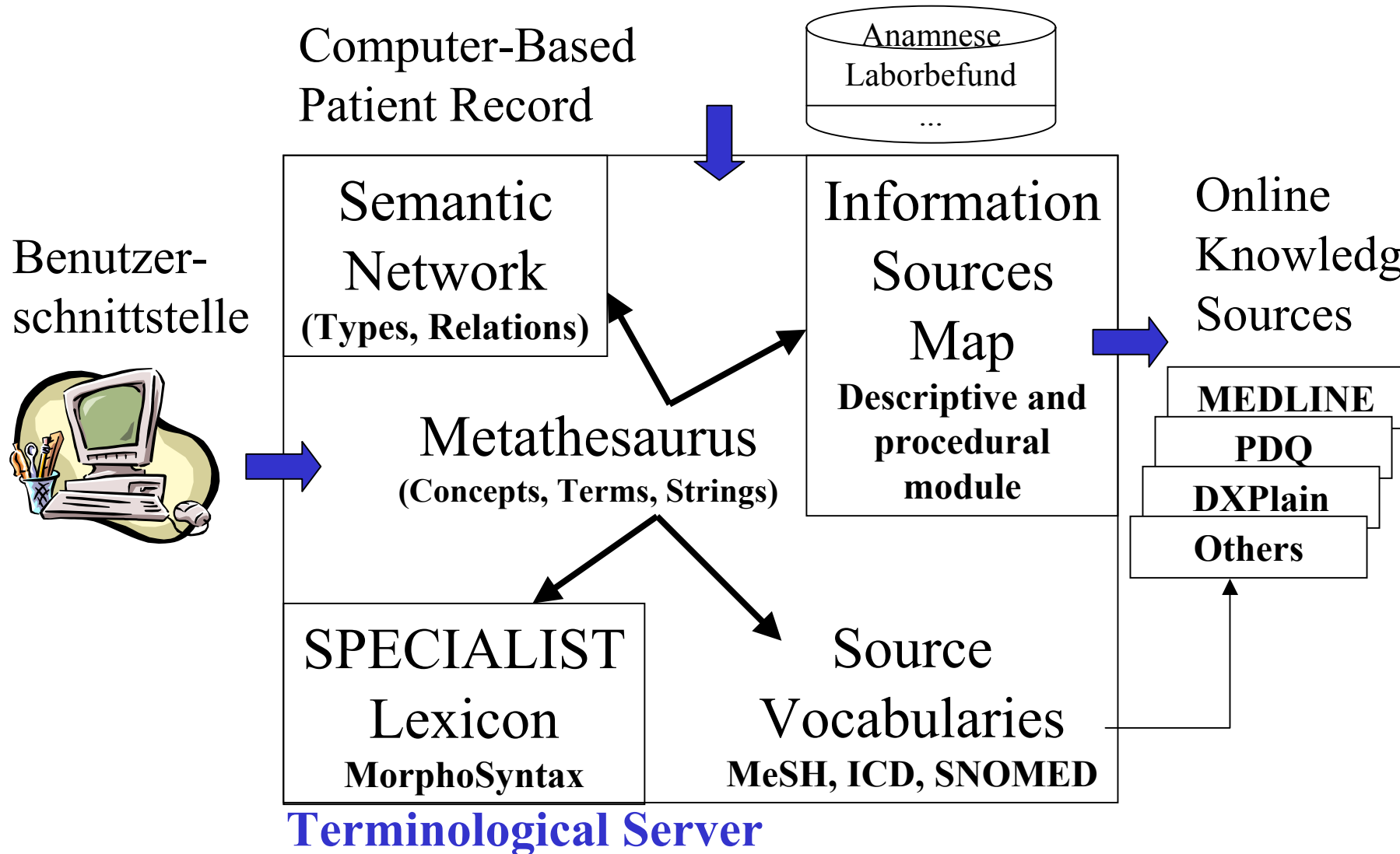
konsistente Kategorisierung aller
Begriffe des Metathesaurus und
Darstellung von semantischen
Beziehungen zwischen diesen
Kategorien

➔ **Types, Relations**

SPECIALIST Lexicon

lexikalische Information; Programme
zur maschinellen Sprachverarbeitung

Komponenten von UMILS



UMLS Metathesaurus

- Enthält Informationen über
 - Biomedizinische Konzepte
 - Begriffe aus kontrollierten Vokabularen und Klassifikationen
 - Administrative health data
 - Bibliographische und Voll-Text-Datenbanken
 - Expertensysteme

UMLS Metathesaurus

- Namen, Bedeutungen, hierarchischer Kontext, Attribute und Beziehungen bleiben erhalten so wie sie im Quell-Vokabular enthalten sind
 - + Basisinformationen
 - + Neue Beziehungen zwischen Begriffen (terms) aus verschiedenen Quell-Vokabularen

UMLS Metathesaurus

- Anwendung:
 - Bereitstellung von Informationen, die Computerprogramme für die Interpretation von Nutzeranfragen nutzen können
- Erstellung:
 - Maschinelle Bearbeitung von maschinenlesbaren Versionen der Quell-Vokabularien
 - Überprüfung von Hand
 - Ergänzung durch Experten
- Zielgruppe:
 - Systementwickler / Datenbankentwickler
 - Bibliothekare / andere Fachleute für Informationsverarbeitung

UMLS Metathesaurus

- Basiert auf der Definition von Konzepten (concepts)
- Zu Concepts wird angegeben:
 - Synonyme, lexikalische Varianten, Übersetzungen
 - Attribute, z.B. Semantischer Typ, Kategorie zu der er gehört, Position im hierarchischen Kontext von verschiedenen Quell-Vokabularien
 - Definition
 - Nutzerinformationen
 - Name der gewählten Datenbank
 - Für MeSH terms Informationen über die Qualifier, die in MEDLINE verwendet werden

UMLS Metathesaurus

- Inhalt - 2000 edition
 - 730000 concepts
 - 1.5 Mio concept names von über 50 verschiedenen biomedizinischen Vokabularien, die meisten in verschiedenen Sprachen

(Bsp.: The 2000 Metathesaurus includes the French, Finnish, German, Italian, Portuguese, Russian (transliterated), and Spanish translations of NLM's Medical Subject Headings (MeSH)).

UMLS Metathesaurus

Beispiele für Quell-Vokabularien (SOURCE VOCABULARIES)

HL7

Health Level Seven Vocabulary (HL7). Ann Arbor (MI): Health Level Seven, 1998. Contact: Health Level Seven; 3300 Washtenaw Avenue, Suite 227, Ann Arbor, MI 48104-4250; Phone: 734-677-7777; Fax: 734-677-6622; E-mail: HQ@HL7.ORG; <http://www.HL7.ORG>.

ICD10

International Statistical Classification of Diseases and Related Health Problems (ICD-10). 10th rev. Geneva (Switzerland): World Health Organization, 1998.

UMLS Metathesaurus

CONCEPTS

- Konzept** Jedes Konzept hat einen eindeutigen **concept identifier (CUI)**
- Term** Jeder englische Konzept-Name zeigt auf alle lexikalischen Varianten mit gleicher Bedeutung und hat einen **common term identifier (LUI)**
- String** Jeder eindeutige Konzept-Name in jeder Sprache hat einen eindeutigen **string identifier (SUI)**

UMLS Metathesaurus

CONCEPTS

Concepts (CUIs)	Terms (LUIs)	Strings (SUIs)
C0004238	L0004238	S0016668
(preferred) Atrial Fibrillation Atrial Fibrillations Auricular Fibrillation Auricular Fibrillations	(preferred) Atrial Fibrillation Atrial Fibrillations	(preferred) Atrial Fibrillation
		S0016669 Atrial Fibrillations
	L0004327	S0016899
	(synonym) Auricular Fibrillation Auricular Fibrillations	(preferred) Auricular Fibrillation
		S0016900
		(plural variant) Auricular Fibrillations

UMLS Metathesaurus

STRINGS

Strings with multiple meanings

<u>Concepts (CUIs)</u>	<u>Terms (LUIs)</u>	<u>Strings (SUIs)</u>
C0009264	L0215040	S0288775
cold temperature	cold temperature	cold temperature
	L0009264	S0007170
	Cold <1> Cold	Cold <1>
		S0026353 Cold
C0009443	L0009443	S0026747
Common Cold	Common Cold	Common Cold
	L0009264	S0007171
	Cold <2> Cold	Cold <2>
		S0026353 Cold

C0024117	L0498186	S0837575
Chronic Obstructive Airway Disease	Chronic Obstructive Airway Disease	Chronic Obstructive Airway Disease
	L0008703	S0837576
	Chronic Obstructive Lung Disease	Chronic Obstructive Lung Disease
	L0009264	S0829315
	COLD <3>	COLD <3>
		S0474508 COLD

UMLS Metathesaurus

types of relationships

Relationships in the Metathesaurus come from the sources themselves or are created by the Metathesaurus editors to link concepts that would not otherwise be connected.

There are nine types of relationships that exist in the Metathesaurus.

Broader (RB)	has a broader relationship.
Narrower (RN)	has a narrower relationship.
Other related (RO)	has relationship other than synonymous, narrower, or broader.

UMLS Metathesaurus

types of relationships

LIKE (RL)	the two concepts are similar or "alike". In the current edition of the Metathesaurus, most relationships with this attribute link MeSH supplementary concepts which are largely chemicals. A number of the concepts linked by this relationship may be synonymous and will be in a single concept identifier in future editions of the Metathesaurus. Source-specific SNMI98 mappings to ICD2000 codes and ALT98 mappings to CPT codes also have this REL, along with the "mapped_to" RELA.
Parent (PAR)	has parent relationship in a Metathesaurus source vocabulary.
Child (CHD)	has child relationship in a Metathesaurus source vocabulary.
Sibling (SIB)	has sibling relationship in a Metathesaurus source vocabulary.
AQ	is an allowed qualifier for a concept in a Metathesaurus source vocabulary.
QB	can be qualified by a concept in a Metathesaurus source vocabulary.

UMLS Metathesaurus

Kontexte für Quellen (contexts)

Full Contexts:

Full Contexts are made for sources that have rich hierarchies. For these sources, all levels of terms **may have Ancestors, Parents, Children and Siblings.**



Beispiel: **MSH2000 (Medical Subject Headings (MeSH))**

Title Contexts:

Title Contexts are built for sources which do not have true hierarchies, but supply "section titles."

These titles are used as Ancestors for the terms under them, but making full contexts for these sources may not be possible because the number of Siblings among all the terms that fall under a title would be excessive. The contexts for these sources are attached only to the lowest level, not to the hierarchical terms. Title Contexts have **Ancestors, but no Children or Siblings.**



Beispiel: **ICD-10**

Mini Contexts:

Mini Contexts are similar to Title contexts in that hierarchical terms do not receive contexts, and no Siblings are computed. However, unlike the pure title contexts, **some preferred terms may have other preferred terms as Parents or Children.** The set of preferred terms under each "section title" has full contexts (without Siblings).



Beispiel: **ICD91**

UMLS Metathesaurus

Attribute (Data Elements)

Es gibt 3 logische Gruppen von Attributen:

1. Concept names

- Ausdruck synonymer Beziehungen zwischen Strings und Terms mit gleicher Bedeutung
- Es gibt folgende Attribute:

CUI	Unique Identifier for Concept	LRL	Least Restriction Level
LUI	Unique Identifier for Term	SUI	Unique Identifier for String
TS	Term Status	STT	String Type
LAT	Language of Term	STR	String

UMLS Metathesaurus

Attribute (Data Elements)

2. Relationships

- Beschreibung von Relationen zwischen verschiedenen Konzepten
- Drücken nicht aus, dass die Konzepte Synonym sind
- Können aus den Quell-Vokabularien abgeleitet oder durch den Metathesaurus hinzugefügt sein
- Es gibt folgende Attribute:

REL **Related Concepts**

COC **Co-occurring Concepts**

ATX **Associated Expressions**

UMLS Metathesaurus

Attribute (Data Elements)

3. Attributes

- Zusätzliche Information zur Bedeutung des Konzeptes
- Identifikation spezieller Arten von Terms
- Information über Quellen
- Mehr als 100 Attribute
- Beispiele:

DEF **Definition**

STY **Semantic Type**

CXT **Hierarchical context**

RN **Registry Number**

UMLS Metathesaurus

Beispiele für Attribute (Data Elements)

TS **Term Status * (M)**

One alpha character

Valid values:

P Preferred Name

S Synonym

s Suppressable synonym (abbreviated form of another term)

An indication of whether the term is the preferred name of the concept to which it is linked, or a synonym of the preferred name. There is a separate preferred name for each language.

A suppressable synonym is a shortened form (not an LV) of another term.

UMLS Metathesaurus

concept name indexes

- 3 Indexes to the concept names
 - Word Index
 - Normalized Word Index (for english strings only)
 - Normalized String Index (for english strings only)

UMLS Metathesaurus

concept name indexes

Word Index

1. Abbreviation of language of word
2. Lowercase Word
3. CUI
4. LUI
5. SUI

Beispiel:

Lung Disease, Obstructive ENG|disease|C0024117|L0024117|S0058458

Obstructive Lung Disease ENG|disease|C0024117|L0024117|S0068168

UMLS Metathesaurus

concept name indexes

Normalized Word Index

1. Abbreviation of language of normalized word (always ENG in this edition of the Metathesaurus)
2. Lowercase, normalized word
3. CUI
4. LUI
5. SUI

Beispiel:

Lung Disease, Obstructive	ENG disease C0024117 L0024117 S0058458
Lung Diseases, Obstructive	ENG disease C0024117 L0024117 S0058463
Obstructive Lung Disease	ENG disease C0024117 L0024117 S0068168
Obstructive Lung Diseases	ENG disease C0024117 L0024117 S0068169

UMLS Metathesaurus

concept name indexes

Normalized String Index

1. Abbreviation of language of normalized string (always ENG in this edition of the Metathesaurus)
2. Lowercase, normalized string
3. CUI
4. LUI
5. SUI

Beispiel:

Lung Disease, Obstructive	ENG disease lung obstructive C0024117 L0024117 S0058458
Lung Diseases, Obstructive	ENG disease lung obstructive C0024117 L0024117 S0058463
Obstructive Lung Disease	ENG disease lung obstructive C0024117 L0024117 S0068168
Obstructive Lung Diseases	ENG disease lung obstructive C0024117 L0024117 S0068169