

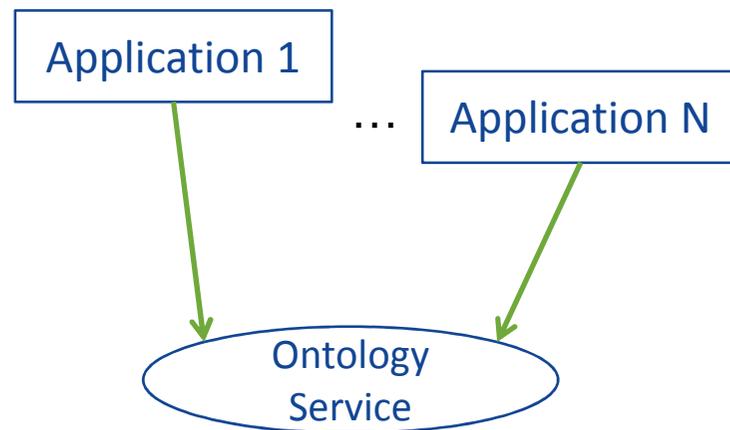


„Klassifikation und Retrieval von eHealth-Dokumenten auf der Basis von Ontologie-Diensten“

Inhalt

- Ontologie-Dienste
- Klassifikation
- Systemarchitektur
- Retrieval

TMF und cloud4health: Text-Mining für die medizinische Forschung, 2015



Warum ?

- Verwendung von **gemeinsamen, kontrollierten Vokabularien**
- **“endpoints”** für die Verwaltung und Bereitstellung von Ontologien
 - Analogie zu Verzeichnisdiensten (->LDAP)

Möglicher Standard

- Common Terminology Services 2 (**CTS2**)
- Standardisierung von **OMG** and **HL7**

Ontologie-Dienste (Ausdrucksstärke)



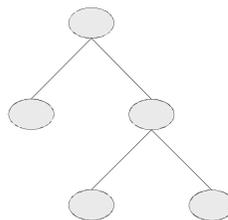
Warum hohe Ausdruckstärke ?

- Viele der zu verwaltenden Artefakte sind **formal-semantic** definiert (*SNOMED, NCI, Gene Ontology* \subseteq *EL++ -Ontologien* !)
- **Intrinsische** Regeln
(*Grippaler Infekt* \subset *Krankheit* UND *causative-agent* (*Virus*) (*SNOMED!*))
- **explizite** Domain-abhängige Regeln
(*Geschäftsfähigkeit* $\dashv\vdash$ *setztVoraus* $\dashv\vdash$ *Volljährigkeit*)

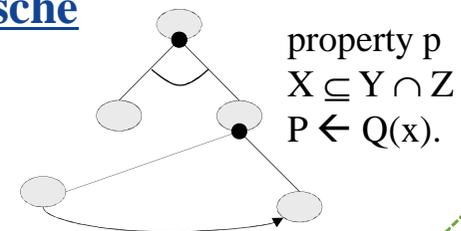
Glossare & Codelisten



Taxonomien



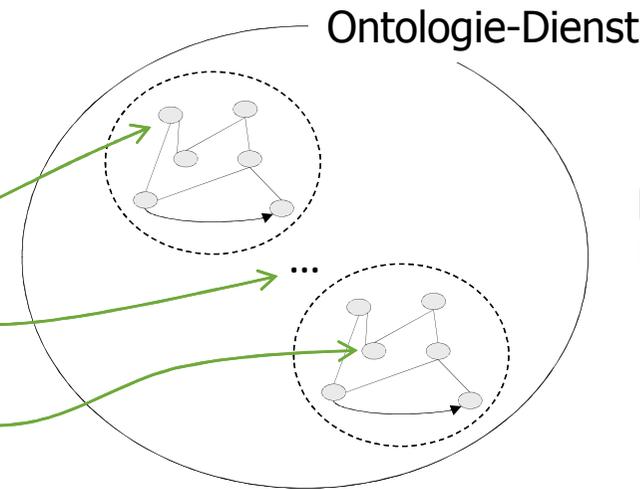
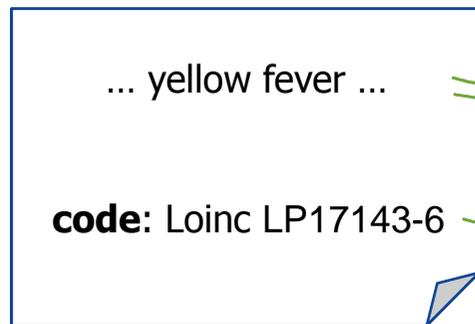
Formal-logische Ontologien



Klassifikation I



CDA, FHIR, openEHR - Dokumente



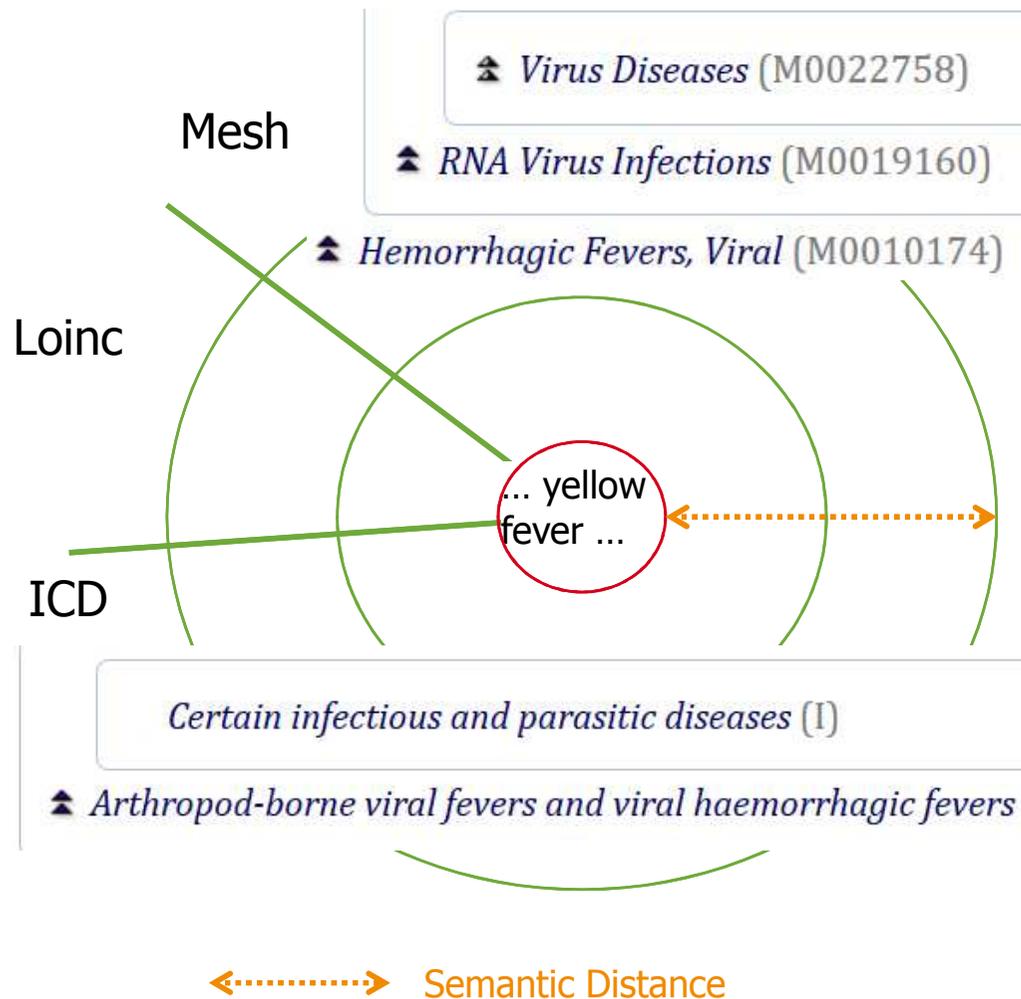
ICD, OPS,
Mesh, HL7-v3,
FHIR, ATC,
UCUM, LOINC

Zuordnung von Dokumenten zu Konzepten
mit Hilfe von

310 Terminologien
326451 Konzepte
558421 Bezeichnungen (Terme)
663490 Assoziationen

- **narrativen** Dokumentteilen
 - Information Retrieval (**IR**) / **Mining** - Verfahren
 - .. : N - Relationen
- **Semantic Types** (-> kodierte Attribute)
 - Explizit
 - .. : 1 - Relationen

Klassifikation II



- Dienst-Funktionalität:

\mathcal{K}_1 : Doc \mapsto SetOf (Concept, Dist)

\mathcal{K}_2 : Unit \mapsto SetOf (Concept, Dist)

- Verfahren:

\mathcal{K}_1 :

for all narrative parts p **in** doc

for all units u **in** p

relate (doc, $\mathcal{K}_2(u)$)

\mathcal{K}_2 :

C = *search-concept-candidates* (u)

for all designations d **of** c **in** C

C-final += fine-match (d, c, u)

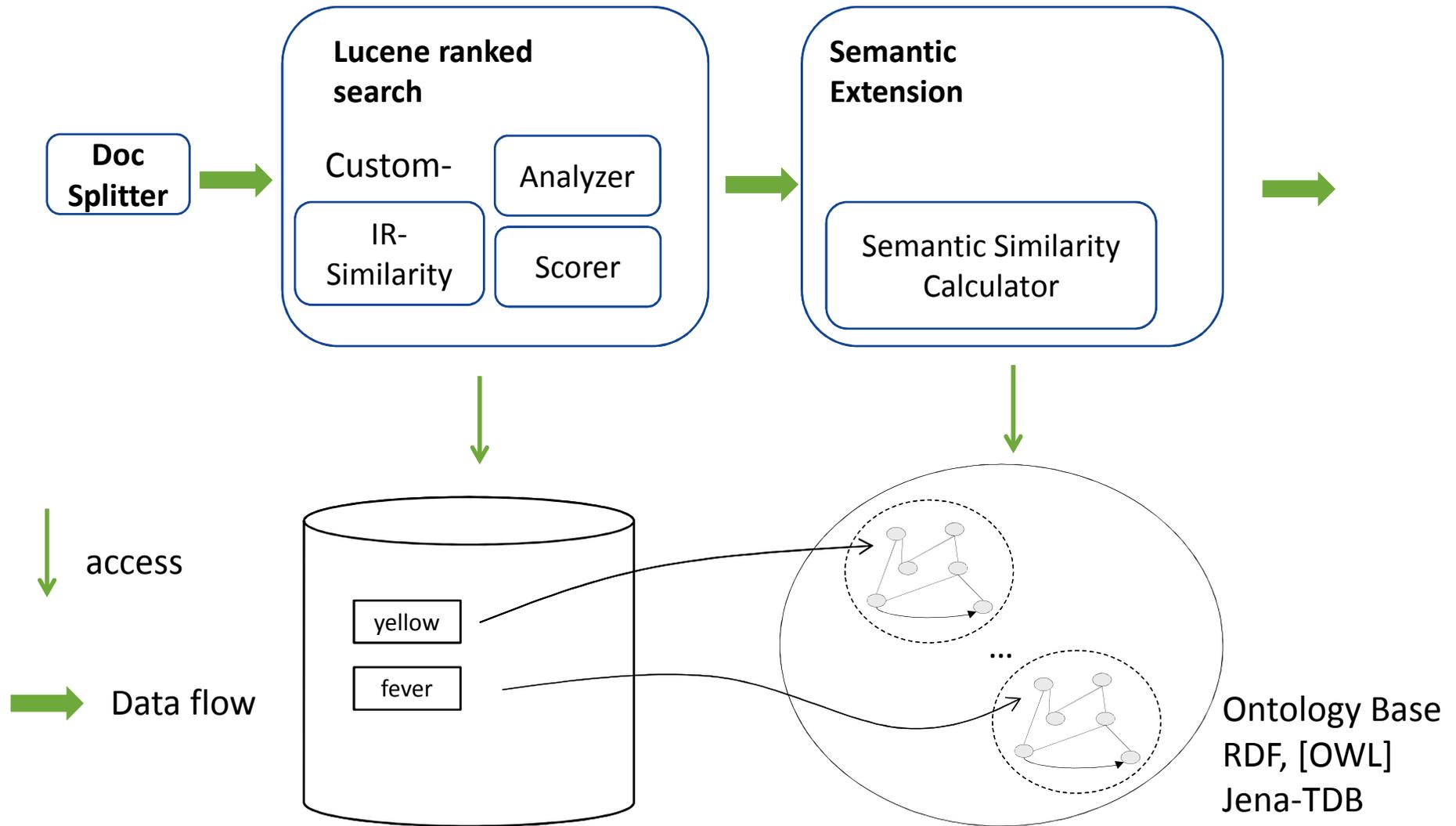
C-final = *semantic-extent* (C-final)



- Dienste-Anforderungen
 - Erweiterung um **Klassifikation** basierend auf **IR-search** (*search-concept-candidates()*)
 - und **semantischen Verfahren** (*semantic-extent()*)
 - Möglichkeit zur Bildung von (Teil-)**Wissensnetzen** (*relate()*)

- Erfüllung (Technologie): Integration von
 - de facto - Industrie – IR-Standard **Lucene** (Kern von SOLR und ES)
 - W3C
 - Standards für semantische Technologien
 - **RDF, SPARQL, OWL**
 - RDF-based **Linked Data**

Systemarchitektur



Retrieval (Dokument <-> Ontologie - Exploration)



Terminologie Navigator ⓘ

Suche

Auswahl: -

Top-100-Treffer (relevantere zuerst) von mehr als 300 (in 0,372 Sekunden)

Loinc2.44, LP7084-9

- Radiology (LP34218-5)
 - Body systems (LP32505-7)
 - Nervous system (LP7441-1)

Brain ▲ ▼

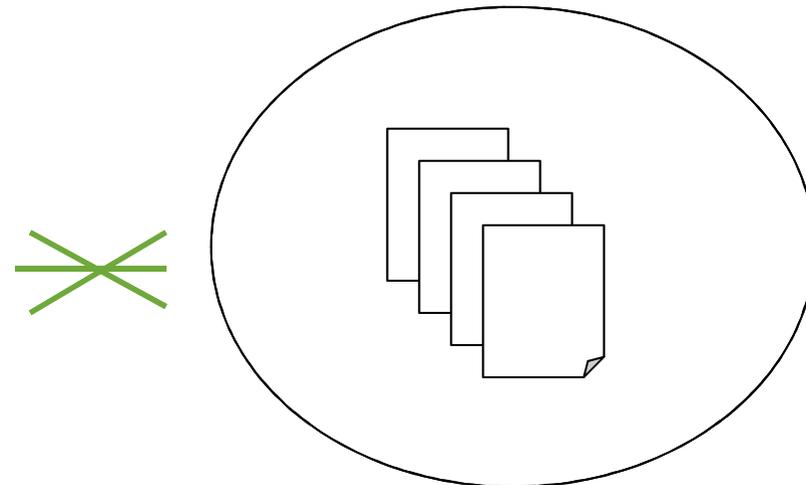
Mesh2014en, M0002865

- Anatomy (A)
 - Nervous System (M0014665)
 - Central Nervous System (M0003801)

Brain ▲ ▼

- Blood-Brain Barrier (M0002730)
- Brain Stem (M0002886)
- Cerebral Ventricles (M0003903)

Document-Pool



Retrieval (Anfragen)



- „Klassisches“ **semantisches Retrieval**
 - *Suche nach „Gehirn“ findet „Alzheimer“ und „Schlaganfall“ - Dokumente*

Ausblick

- **Integrative** , strukturierte Suche mit Hilfe aller vernetzten Artefakte
 - Dokumenttypen und -inhalte
 - Intrinsische Regeln
 - Explizite Regeln
 - Ontologie-Entitäten
- Z.B.
 - „Finde alle **FHIR-Dokumente**, dessen **Diagnose-Abschnitte durch Viren hervorgerufene Krankheiten** enthalten, die **unterhalb Krankheit X** eingeordnet sind“



- Vielen Dank für Ihre Aufmerksamkeit