



Medizinische Universität Graz

Nutzung von unstrukturierten Daten der elektronischen Patientenakte

Stefan Schulz

Markus Kreuzthaler

Institut für Medizinische Informatik,
Statistik und Dokumentation

Nutzungspotentiale

- Kohortenbildung für prospektive Studien
- Bildung von Patientengruppen nach inhaltlichen Kriterien (retrospektive Forschung, Biobank)
- Verbesserung der Nutzbarkeit elektronischer Krankenakten
 - Entscheidungsunterstützung
 - semantisches Retrieval
 - benutzeroptimierte Zusammenfassung ("Quick View")
 - patientenübergreifende Recherche
- Qualitätssicherung
- Prädiktive Analysen für Klinik und Verwaltung
- Optimierung der Kodierung von Diagnosen und Leistungen)

Sprachökonomie

Pat. mit rez.
HWI und VUR

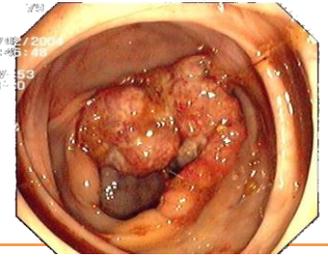
rezent?
rezidivierend?

Harnwegsinfekt ?
Hinterwandinfarkt?

Vesicoureteral
reflux?



Synonyme, Schreibvarianten



Colon-Ca	Kolon-Ca	Kolon- karzinom	Colon- carcinom
Colon- Karzinom	Kolonkrebs	Dickdarm- krebs	Dickdarm-Ca
Malignom des Kolon	Dickdarm- karzinom	Dickdarm-Ca	Bösartige Neubildung am Dickdarm
Bösartiger Dickdarm- tumor	maligne Neoplasie des Dickdarms	Karzinom des Dickdarms	maligne NPL des Colon

Schreib-, Tipp-, Diktierfehler



Simvastatin

Sinvastatin

Simvastastin

Simvastain

Simvastad

Simbastatin

Simavstatin

Simavastatin

Simastatin

Symvastatin

Simvastation

Simvaststin

Simvatatin

Simvatin

Simvatstain

Simvstatin

Dokumentenrecherche ohne Semantik

"Makroskopie: "Resektat nach Whipple": Ein noch nicht eröffnetes Resektat, bestehend aus einem distalen Magen mit einer kleinen Kurvaturlänge von 9,5 cm und einer großen Kurvaturlänge von 13,5 cm, sowie einem duodenalen Anteil von 14 cm Länge. 2 cm aboral des Pylorus zeigt die Dünndarmwandung eine sanduhrartige Stenose. Im Magen- und Duodenallumen reichlich zähflüssiger Schleim, sanguinolent; die Schleimhaut ist insgesamt livide. Auf lamellierenden Schnitten zähfestes weißliches, teilweise nodulär konfiguriertes Gewebe, ohne dass Gallengänge manifest werden. Der distale Anteil des Ductus pancreaticus ist leicht erweitert und von der Papilla Vateri aus 4,5 cm weit sondierbar, wobei er hier in einer peripankreatischen Narbenzone abbricht. Eine Gallengangsmündung lässt sich makroskopisch nicht abgrenzen. Die berichtete Duodenumstenose liegt 2,5 cm oral der Papilla Vateri und steht mit der beschriebenen Narbenzone in direktem Zusammenhang.

Dokument wird gefunden

mit den Suchwörtern:

"Whipple", "Magen",
"Pylorus"

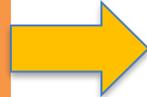
Keine Treffer für:

"Pankreatikoduodenektomie", "Resektion",
"Duodenum", "Zwölffingerdarm", "Pankreas",
"Bauchspeicheldrüse", "Gallengang", "Pankreasgang",
"Ductus pancreaticus", "Papille", "Magenresektion"

Semantische Interpretation

St. p. TE eines exulc.
sek.knot.SSM (C43.5) li
US dors. 5/11 Level IV
2,4 mm Tumordurchm.
Sentinell LK ing. li.
tumorfrei

NLP



Code (SNOMED CT)	Wert	Kontext
254730000 Superficial spreading malignant melanoma of skin		392521001 History of
301889008 Excision of malignant skin tumor		392521001 History of
47224004 Skin of posterior surface of lower leg 7771000 Left		
81827009 Diameter 258673006 millimeter	2.41	
258403002 Lymph node level IV		
94339008 Secondary malignant neoplasm of inguinal lymph nodes		15240007 Current 2667000 Absent



Tools und Ressourcen für maschinelle Analyse medizinischer Sprache

- Ontologien / Referenzterminologien
 - Axiomatische Beschreibungen, z.B.
MitralValve subclassOf partOf some Heart
- Interface-Terminologien / Lexika
 - Fachtermini und deren Beziehungen, z.B. Synonyme: "SSMM" = "Superficially spreading malignant melanoma"
- Textkorpora (z.B. Kliniktexte, Veröffentlichungen, Web)
 - Annotierte Korpora: Wörter <-> Wortklassen, Textpassagen <-> Kontext (z.B. Diagnose, Familienanamnese, Labor, Medikation)
 - Nicht annotierte Korpora: z.B. Worthäufigkeiten, Wortkollokationen
- NLP (Natural Language Processing) Pipeline

SNOMED CT, ICD-10, ICD-O, LOINC (UMLS)

MUG-GIT: German Interface Terminology for SNOMED CT

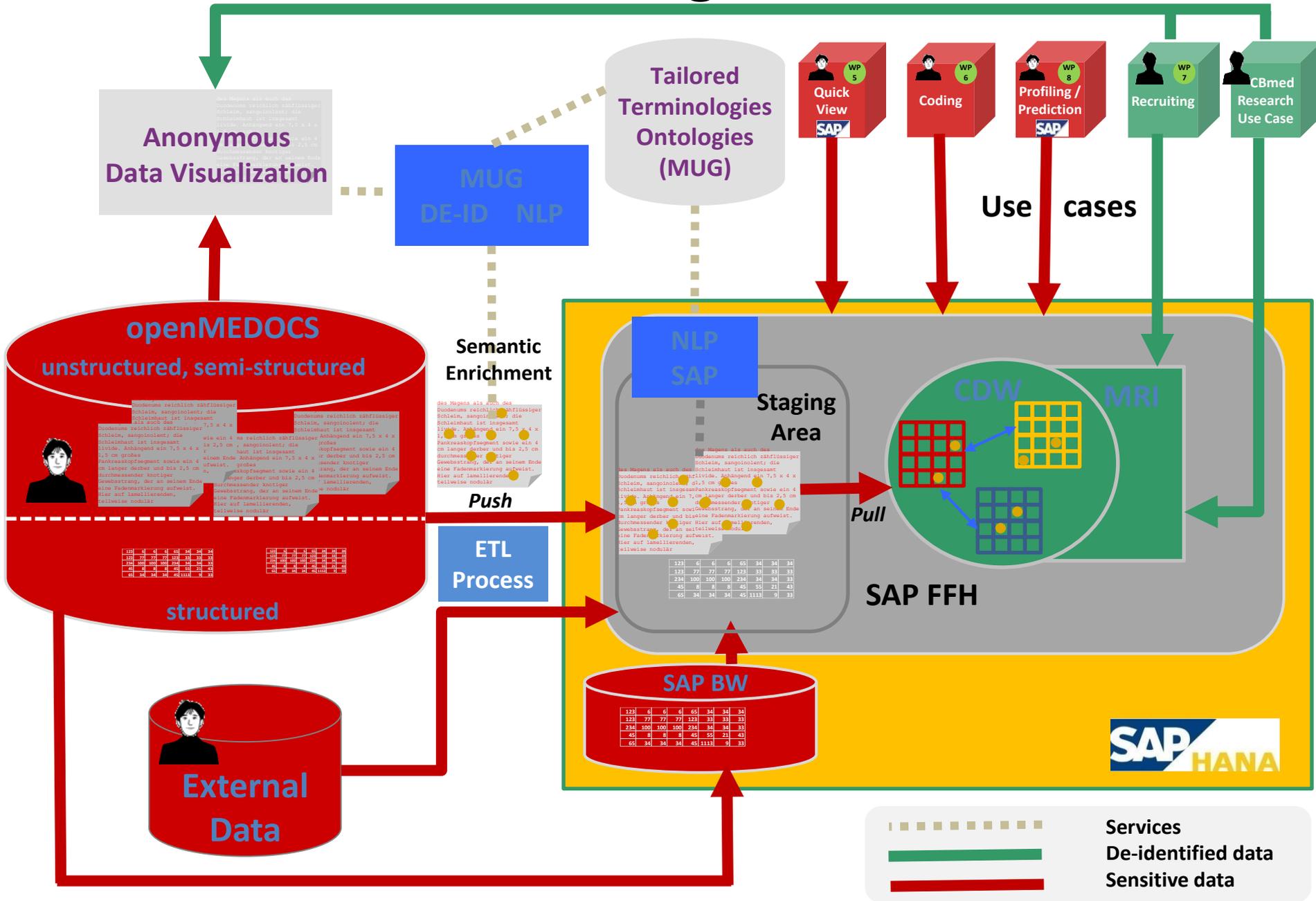
Anonymisierte Extrakte aus Klinikinformationssystemen

Averbis-Tools +
Eigenentwicklungen,
verwendet in:
epaNAVI
SEMCARE
CBmed --IICCAB

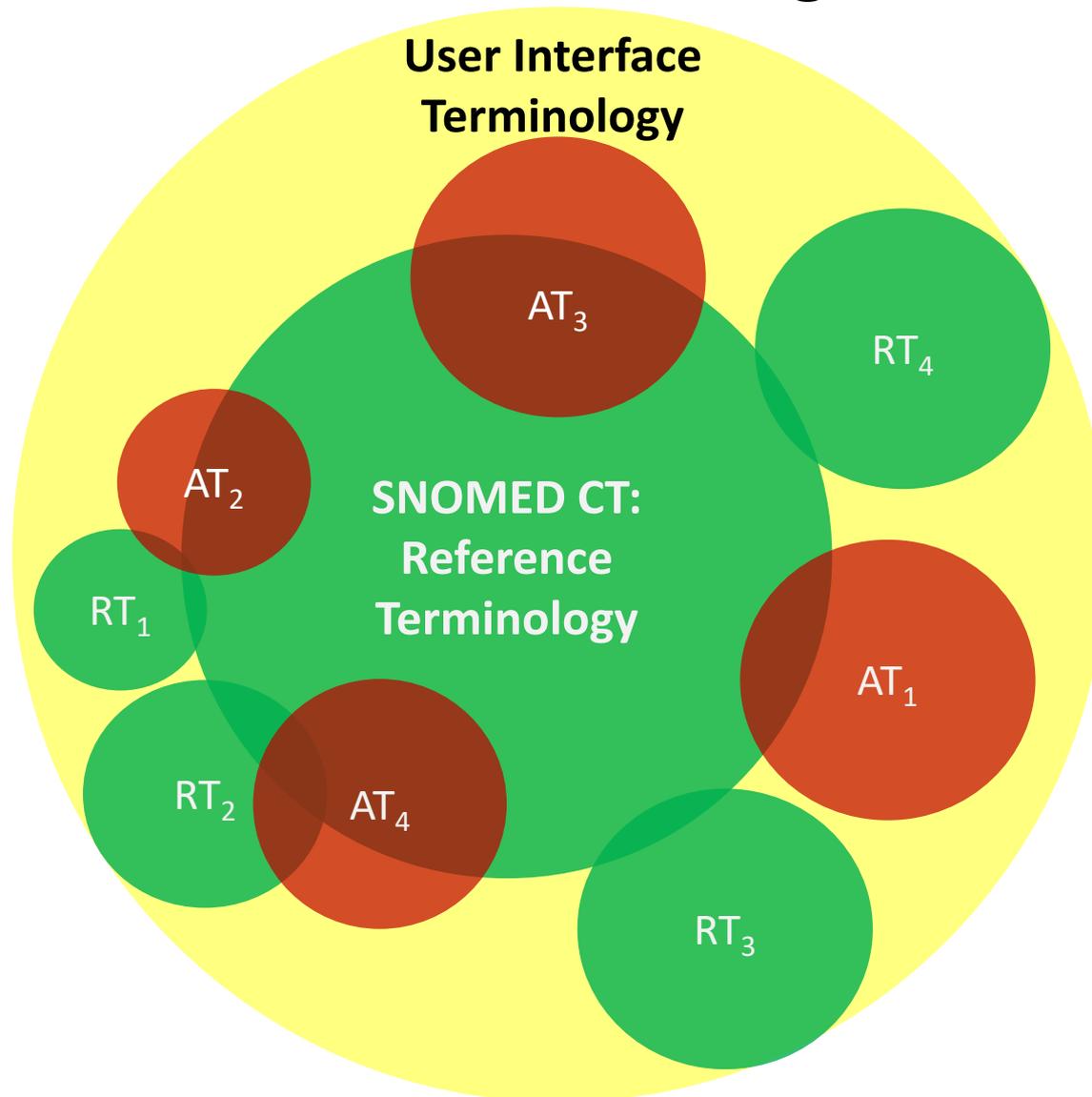
Fazit

- Gering strukturierte Texte in der elektronischen Krankenakte:
 - oft einzige Informationsquelle
 - oft verlässlicher und nachhaltiger als strukturierte Inhalte
 - Liefert wertvolle Daten für primäre und sekundäre Nutzung
 - bisher kaum erschlossene "Goldmine"
- Breites Spektrum an semantischen / linguistischen Ressourcen, die jedoch auf die deutsche Medizinsprache zugeschnitten werden müssen
- Anlehnung an internationale Standards sinnvoll. Wichtigstes Beispiel: SNOMED CT als umfassende klinische Referenzterminologie
- MUG-IMI: umfassende Erfahrung mit medizinischer Semantik, Softwarewerkzeugen und Ressourcen

Anwendung CBmed



Interface-, Aggregations- und Referenzterminologien



Kontakt

Stefan Schulz

(Univ. Prof. Dr. med.)

Institut für Medizinische
Informatik, Statistik und
Dokumentation

stefan.schulz@medunigraz.at

