

SNOMED CT IN AVERBIS HEALTH DISCOVERY

...aus der Nutzerperspektive

Stefan Schulz, Markus Kreuzthaler, Amila Kugic, Andrea Berghold

Institut für Medizinische Informatik, Statistik und Dokumentation

SMITH - PheP - 18. Oktober 2021

SNOMED CT - wozu?

Highlights

- ▶ Umfassende, feingranuläre Repräsentation klinischer Inhalte
- ▶ Internationaler Standard → Interoperabilität
- ▶ Ontologiebasiert, Polyhierarchisch → semantische Suche
- ▶ Harmonisiert gut mit HL7-FHIR als Informationsmodell

Drawbacks

- ▶ D-A-CH: Mitglieder von SNOMED Intl., Übersetzungsaktivitäten am Anfang
- ▶ In den meisten Ländern noch kein Routineeinsatz
- ▶ Kaum Erfahrungen als Terminologie für Text Mining

SCT-GIT - SNOMED CT - German Interface Terminology

- Klinisches NLP erfordert Termsammlungen, die den *Sprachgebrauch*, nicht die *Sprachnorm* abbilden

Häufigkeit von SNOMED Preferred Terms und ihrer Übersetzungen

Treffer Google

- | | |
|---|-----|
| – Englisch: "Secondary malignant neoplasm of liver" | 100 |
| – Schwedisch: "sekundär malign levertumör" | 1 |
| – Deutsch: "Sekundäre maligne Neoplasie der Leber" | 1 |

Häufigkeit klinisch gebräuchlicher Synonyme

- | | |
|---------------------------------|-----------|
| – Englisch: "liver metastases" | 1.230.000 |
| – Schwedisch: "levermetastaser" | 217.000 |
| – Deutsch: "Lebermetastasen" | 204.000 |

Ähnliche Beobachtungen in klinischen Korpora / PubMed:

Z.B. kein einziger Treffer für "Elektrokardiogramm" in 30.000 Kardiologie-Arztbriefen

SCT-GIT - SNOMED CT - German Interface Terminology

- ▶ Seit 2014, mit Hilfe von 1-3 Studierenden der Medizin: Übersetzung eines Kernvokabulars aus englischem SNOMED CT
- ▶ Algorithmische Erzeugung von Varianten, Kombinationen, Komposita aus englischen SNOMED-Descriptions
- ▶ Scoring nach Vorkommen in Referenzkorpora, Abgleich mit Negativliste
- ▶ Gefilterte Version für NLP (nicht länger als 6 Token, maximal 100 Synonyms pro Konzept): derzeit für ca. 294.000 SNOMED-Konzepte 4,3 Mio Terme
- ▶ Angepasst an Averbis Health Discovery

► Kernvokabular

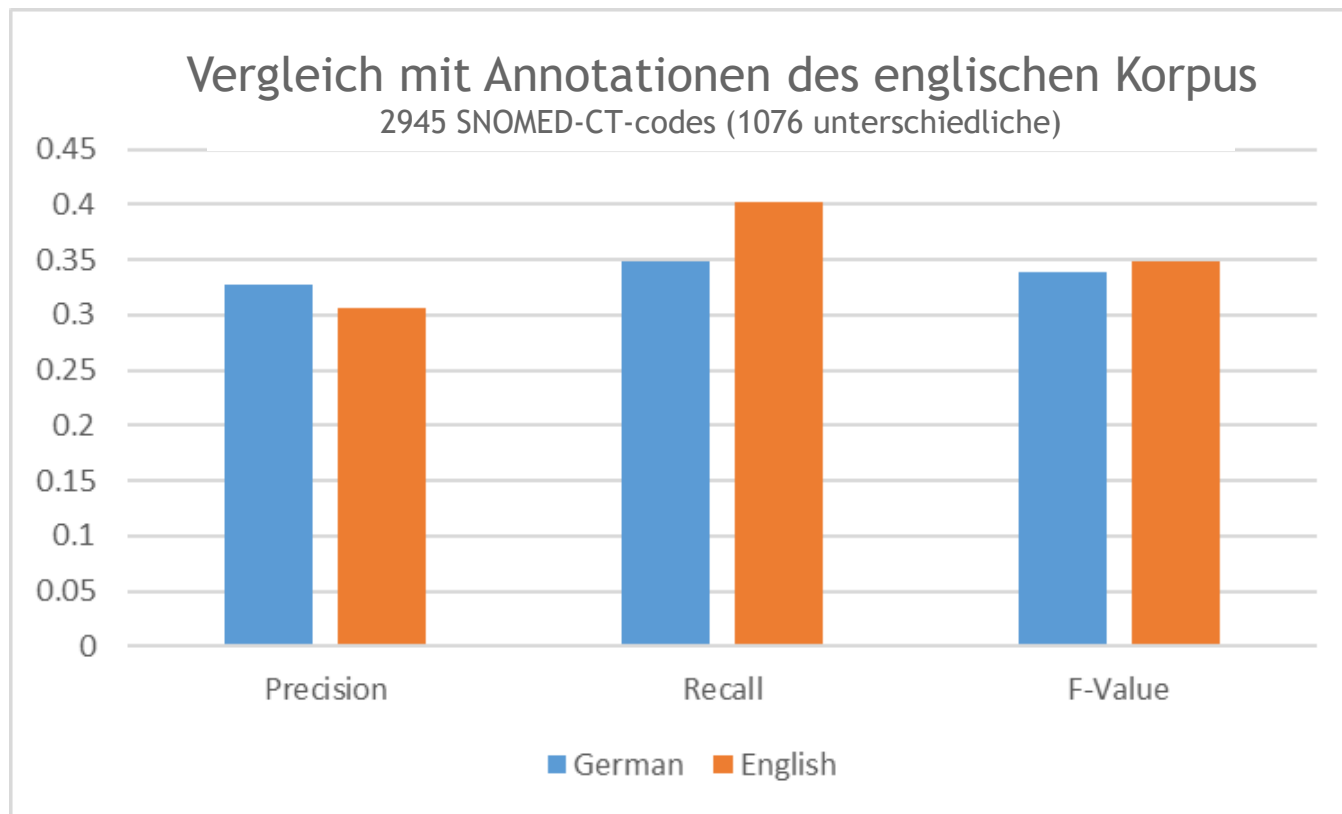
English	L Count	German 1	German 2	German 3	German 4
burn	1	1264 Brandverletzung NN F	Brandwunde NN F	Verbrennung NN F	
normal	1	1264 normales JJ	normenhaftes JJ		
ankle	1	1254 Knöchel NN M			
wrist	1	1251 Handgelenk NN N			
drug	1	1244 Wirkstoff NN M	Arznei NN F	Arzneimittel NN N	Droge NN F
second	1	1244 zweites JJ	Sekunde NN F	Sekunden-	%VOID% 2. %VOID%
uncertain	1	1227 unsicheres JJ			
abdominal	1	1222 abdominales JJ	Bauch-	abdominelles JJ	
membrane	1	1210 Membran NN F			
liver	1	1207 Hepar NL N	Leber NN F		
microgram	1	1202 %VOID% µg %VOID%	Mikrogramm NN N	Mikrogramm NL N	
middle	1	1193 mittleres JJ	Mitte NN F	Mittel--	
ulcer	1	1180 Ulzeration NN F	Ulkus NN N	Geschwür NN N	
upper limb	2	1180 oberes JJ Extremität NN F	Arm NN M	oberes JJ Gliedermaße NN F	OE NL F
fluoroscopic	1	1171 Durchleuchtungs-	durchleuchtungsgestütztes JJ	fluoroskopisches JJ	
effect	1	1170 Effekt NN M	Auswirkung NN F	Wirkung NN F	Folge NN F
service	1	1158 Service NN M	Dienst NN M	Service NN N	
vehicle	1	1154 Fahrzeug NN N			
external	1	1149 äußeres JJ	externes JJ	auswärtiges JJ	
internal	1	1149 inneres JJ	internes JJ	internistisches JJ	
of foot	2	1149 des Fußes	_Fuß_		

► Automatisch generierte Interface-Terme

99451000119105	0.833	Cerebral infarction due to stenosis of carotid artery (disorder)	Hirnfarkt verursacht durch Stenose der A. carotis
99451000119105	0.833	Cerebral infarction due to stenosis of carotid artery (disorder)	Hirnfarkt verursacht durch Stenose der A. karotis
99451000119105	0.833	Cerebral infarction due to stenosis of carotid artery (disorder)	Schlaganfall wegen Stenose der Halsschlagader
99451000119105	0.833	Cerebral infarction due to stenosis of carotid artery (disorder)	Insult wegen Stenose der Halsschlagader
99451000119105	0.833	Cerebral infarction due to stenosis of carotid artery (disorder)	Schlaganfall wegen Karotisstenose
99451000119105	0.833	Cerebral infarction due to stenosis of carotid artery (disorder)	Insult wegen Karotisstenose
99451000119105	0.800	Cerebral infarction due to stenosis of carotid artery (disorder)	Gehirnfarkt verursacht durch Verengung der Halsschlagader

SCT-GIT - SNOMED CT - German Interface Terminology

► Validierung gegen klin. Parallelkorpus mit Expertenannotationen



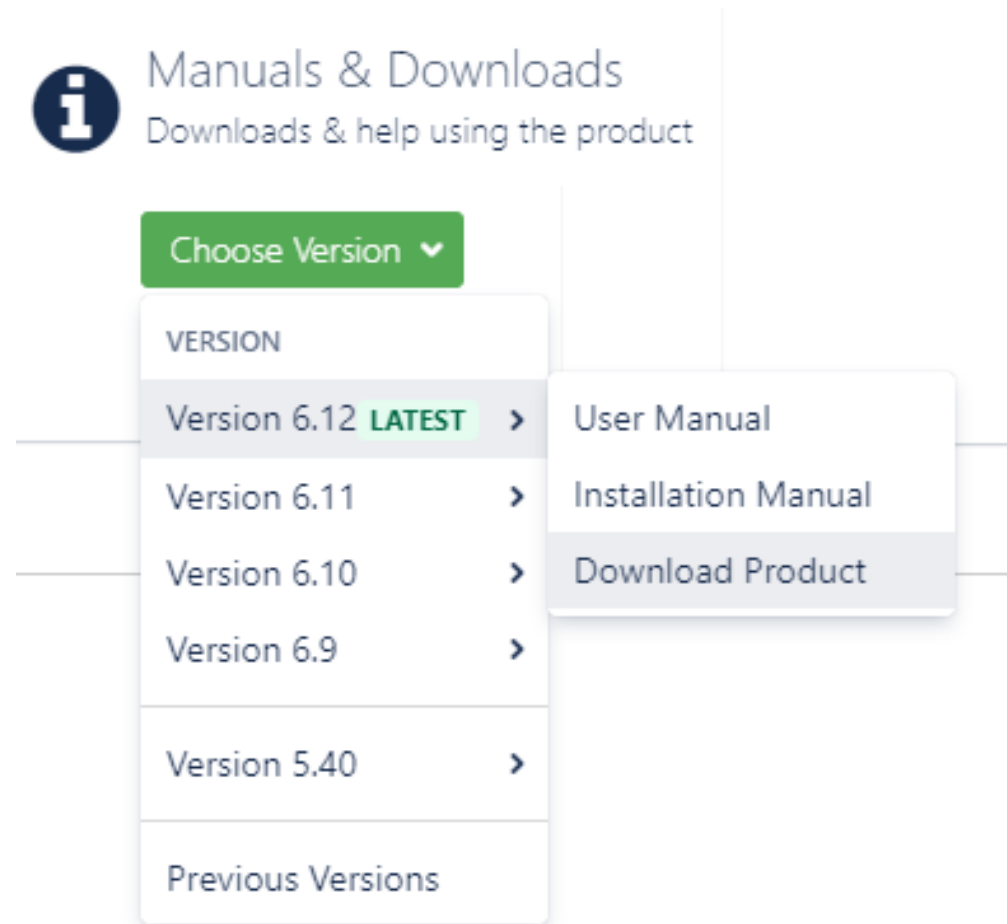
- Unterschiede nicht signifikant zwischen
 - NLP-Tagging auf Basis des englischen SNOMED auf englischem Korpus
 - NLP-Tagging auf Basis unserer Interface-terminologie auf bedeutungsgleichem deutschem Korpus
- Inter-Annotator-Agreement der manuellen Annotationen 0,4 (Krippendorffs Alpha): bekanntes Problem großer Terminologien mit ähnlichen Konzepten und Möglichkeiten von Prä- und Postkoordination

Nutzung und Quellen

- ▶ Derzeitiger Status
 - ▶ Prototyp
 - ▶ IP: Medizinische Universität Graz, Ansprechpartnerin Andrea Berghold (andrea.berghold@medunigraz.at)
 - ▶ Innerhalb der MI-I zusammen mit Averbis Health Discovery für Zwecke der Forschung entgeltfrei nutzbar
 - ▶ Erfordert Registrierung (Feedback an Entwickler erwünscht)
- ▶ <https://user.medunigraz.at/stefan.schulz/mugit/>
 - ▶ Terminologie als OBO File
 - ▶ 2-3 Releases pro Jahr
 - ▶ Transitive closure file (alle Ancestors für jedes SNOMED-Konzept)
 - ▶ Negativliste für Code-term-Kombinationen
 - ▶ Einfacher Python-Beispielcode

Step-by-Step

- ▶ AHD 6 lokal installieren
- ▶ OBO-Datei herunterladen
- ▶ Terminologie anlegen
- ▶ Terminologie importieren
- ▶ Terminologie übergeben
- ▶ Pipeline anlegen
- ▶ API-Token generieren
- ▶ Über REST-API auf HD zugreifen
- ▶ JSON-Output beliebig weiterverarbeiten
- ▶ Ggf. OBO-Datei und Negativliste an eigene Bedürfnisse anpassen



<https://help.averbis.com/display/AKB/Health+Discovery>

Step-by-Step

- ▶ AHD 6 lokal installieren
- ▶ OBO-Datei herunterladen
- ▶ Terminologie anlegen
- ▶ Terminologie importieren
- ▶ Terminologie übergeben
- ▶ Pipeline anlegen
- ▶ API-Token generieren
- ▶ Über REST-API auf HD zugreifen
- ▶ JSON-Output beliebig weiterverarbeiten
- ▶ Ggf. OBO-Datei und Negativliste an eigene Bedürfnisse anpassen

<https://user.medunigraz.at/stefan.schulz/mugit/>

	Parent Directory	-
	MUGIT_de_drugs_nlp_alt.obo	02-Aug-2021 14:41 120M
	SCT-GIT_de_drugs_nlp.obo	11-May-2021 13:11 151M
	SCT-GIT_de_large.dat	11-May-2021 13:12 647M
	SCT-GIT_de_nlp.dat	11-May-2021 13:11 294M
	SCT-GIT_de_precise.dat	11-May-2021 13:10 53M
	SCT-GIT_fuer_AHD.pptx	17-Feb-2021 12:00 231K
	active_fsn.dat	06-Apr-2022 12:55 20M
	ahd_extrahiere_sct.py	18-Aug-2022 09:50 7.6K
	closure.dat	05-Apr-2022 16:55 63M
	neg_trans_code.dat	18-Aug-2022 09:49 1.5M
	old_versions/	11-May-2021 13:09 -
	readme.txt	18-Aug-2022 09:57 3.9K

Step-by-Step

- ▶ AHD 6 lokal installieren
- ▶ OBO-Datei herunterladen
- ▶ Terminologie anlegen
- ▶ Terminologie importieren
- ▶ Terminologie übergeben
- ▶ Pipeline anlegen
- ▶ API-Token generieren
- ▶ Über REST-API auf HD zugreifen
- ▶ JSON-Output beliebig weiterverarbeiten
- ▶ Ggf. OBO-Datei und Negativliste an eigene Bedürfnisse anpassen

Terminologien



Editor



Verwaltung

Terminologie anlegen

Name: ?

Label: ?

Version:

Konzept-Typ:

Hierarchisch: ☐

Verschlüsselter Export: ☐

Zugewiesene Sprachen

 Deutsch ↑ ↓ ×

Verfügbare Sprachen

←  Diverse

←  Englisch

←  Spanisch

←  Französisch

←  Italienisch

←  Portugiesisch

OK Abbrechen

Step-by-Step

- ▶ AHD 6 lokal installieren
- ▶ OBO-Datei herunterladen
- ▶ Terminologie anlegen
- ▶ Terminologie importieren
- ▶ Terminologie übergeben
- ▶ Pipeline anlegen
- ▶ API-Token generieren
- ▶ Über REST-API auf HD zugreifen
- ▶ JSON-Output beliebig weiterverarbeiten
- ▶ Ggf. OBO-Datei und Negativliste an eigene Bedürfnisse anpassen



The screenshot shows a web interface titled "Terminologie importieren". It features two main sections: "Importformat" with a dropdown menu currently set to "OBO Importer", and "Terminologiedatei" with a file selection area containing a "Datei auswählen" button and the text "Keine ausgewählt". At the bottom right, there are two buttons: "Import" (light blue) and "Abbrechen" (orange).

"Refresh" zum Statusupdate

Step-by-Step

- ▶ AHD 6 lokal installieren
- ▶ OBO-Datei herunterladen
- ▶ Terminologie anlegen
- ▶ Terminologie importieren
- ▶ Terminologie übergeben  Übergabe an die Textanalyse (2771 / 293579)
- ▶ Pipeline anlegen
- ▶ API-Token generieren
- ▶ Über REST-API auf HD zugreifen
- ▶ JSON-Output beliebig weiterverarbeiten
- ▶ Ggf. OBO-Datei und Negativliste an eigene Bedürfnisse anpassen



Step-by-Step

- ▶ AHD 6 lokal installieren
- ▶ OBO-Datei herunterladen
- ▶ Terminologie anlegen
- ▶ Terminologie importieren
- ▶ Terminologie übergeben
- ▶ **Pipeline anlegen**
- ▶ API-Token generieren
- ▶ Über REST-API auf HD zugreifen
- ▶ JSON-Output beliebig weiterverarbeiten
- ▶ Ggf. OBO-Datei und Negativliste an eigene Bedürfnisse anpassen



Pipeline-Konfiguration

🔗 Verwendete Annotatoren der Pipeline *sct-git* (6013 | 105.44)

- 1 LanguageDetection
- 2 HealthPreprocessing
- 3 GenericTerminologyAnnotator

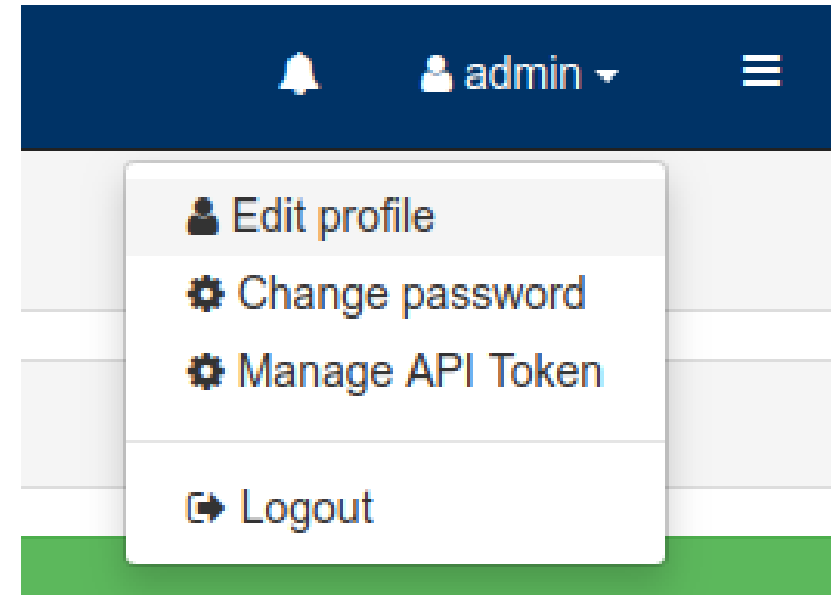
Parameter

terminologyNames	A	1 ausgewählt ▼
useSegmentLookup	☐	Inaktiv
useStemLookup	☒	Aktiv

- 4 Negations
- 5 HealthPostprocessing

Step-by-Step

- ▶ AHD 6 lokal installieren
- ▶ OBO-Datei herunterladen
- ▶ Terminologie anlegen
- ▶ Terminologie importieren
- ▶ Terminologie übergeben
- ▶ Pipeline anlegen
- ▶ **API-Token generieren**
- ▶ Über REST-API auf HD zugreifen
- ▶ JSON-Output beliebig weiterverarbeiten
- ▶ Ggf. OBO-Datei und Negativliste an eigene Bedürfnisse anpassen



Step-by-Step

- ▶ AHD 6 lokal installieren
- ▶ OBO-Datei herunterladen
- ▶ Terminologie anlegen
- ▶ Terminologie importieren
- ▶ Terminologie übergeben
- ▶ Pipeline anlegen
- ▶ API-Token generieren
- ▶ Über REST-API auf HD zugreifen
- ▶ JSON-Output beliebig weiterverarbeiten
- ▶ Ggf. OBO-Datei und Negativliste an eigene Bedürfnisse anpassen

```
url = "https://myserver.com/health-discovery/rest/v1/textanalysis/projects"
url = url + "/SNOMED/pipelines/sct-git/analyseText?annotationTypes="
url = url + "de.averbis.extraction.types.Concept&language=de"

hds = {'Content-type': 'text/plain; charset=utf-8',
      'api-token': 'm9580086b54ee7fb7161cc82b8314f8886637085f4b1ed4bcb0e0b509c30977c'}
```

Step-by-Step

- ▶ AHD 6 lokal installieren
- ▶ OBO-Datei herunterladen
- ▶ Terminologie anlegen
- ▶ Terminologie importieren
- ▶ Terminologie übergeben
- ▶ Pipeline anlegen
- ▶ API-Token generieren
- ▶ Über REST-API auf HD zugreifen
- ▶ JSON-Output beliebig weiterverarbeiten
- ▶ Ggf. OBO-Datei und Negativliste an eigene Bedürfnisse anpassen

```
myResponse = requests.post(url,
                           data=row.encode('utf-8'),
                           headers=hds)

if (myResponse.ok):
    jData = json.loads(myResponse.content)
else:
    print("Response not OK")
for annotation in jData['payload']:
    if (annotation['type'] == 'de.averbis.extraction.types.Concept'):
        snomed_id = annotation['conceptID']
```

Step-by-Step

- ▶ AHD 6 lokal installieren
 - ▶ OBO-Datei herunterladen
 - ▶ Terminologie anlegen
 - ▶ Terminologie importieren
 - ▶ Terminologie übergeben
 - ▶ Pipeline anlegen
 - ▶ API-Token generieren
 - ▶ Über REST-API auf HD zugreifen
 - ▶ JSON-Output beliebig weiterverarbeiten
- ▶ Ggf. OBO-Datei und Negativliste an eigene Bedürfnisse anpassen

```
3527087
3527088 [Term]
3527089 id: 291849004
3527090 name: Accidental methandriol poisoning
3527091 synonym: "Accidental methandriol poisoning" EXACT []
3527092 synonym: "versehentliche Methandriolvergiftung" EXACT PREF []
3527093 synonym: "akzidentelle Methandriolvergiftung" EXACT []
3527094 synonym: "zufällige Methandriolvergiftung" EXACT []
3527095 synonym: "versehentliche Methandriolintoxikation" EXACT []
3527096 synonym: "akzidentelle Methandriolintoxikation" EXACT []
3527097 synonym: "zufällige Methandriolintoxikation" EXACT []
3527098
```

```
2755 231008003 Rückenmarkdurchtrennung
2756 23131001 Salmonella IIIa 13,22:z4,z23:-
2757 231346001 lokales Anästhetikum
2758 23141003 keuchende Atmung Gasping for breath (finding)
2759 231472009 Barbituratabusus Barbiturate dependence (disorder)
2760 231472009 Barbituratmissbrauch Barbiturate dependence (disorder)
2761 231877006 schlimm
2762 232663008 Atemwegeeingriff
2763 232663008 Atemwegeprocedere
2764 232663008 Atemwegeprozedere
2765 232663008 Atemwegeprozedur
2766 232663008 Atemwegeverfahren
2767 232717009 Bypass Coronary artery bypass grafting (procedure)
2768 232853003 LVOT
```

Demo

:.			
Patient	116154003		Patient
Aufnahme	111984006		Ingestion
Aufnahme	129273005		Admission - action
Telefonischer	359993007		Telephone
Patient	116154003		Patient
in Behandlung bei Fr. OÄ Schönfeld in PIA.			
Patient	116154003		Patient
Aufnahme	111984006		Ingestion
Aufnahme	129273005		Admission - action
Zeit	410669006		Time (property)
schlafen	258158006	negated	Sleep, function
Quetiapin	386850001		Quetiapine
Dosis	408102007		Unit dose
mg	258684004		milligram
Seroquel	108383008		Quetiapine fumarate
Dosis	408102007		Unit dose
mg	258684004		milligram
Seroquel	108383008		Quetiapine fumarate
Prolong.			
Ende	261782000		End
Drogen	6525002	negated	Dependent drug abuse
rauche	77176002		Smoker
Mal	400096001		Melanocytic nevus

Weiterführende Aktivitäten

- ▶ SNOMED on FHIR Arbeitsgruppe
- ▶ Annotationsguideline für klinische Texte mit SNOMED CT und FHIR
- ▶ Bei Bedarf Trainingssession, Kontakt: stefan.schulz@averbis.com
Voraussetzung:
 - ▶ Averbis Health Discovery Version 6 installiert
 - ▶ Beherrschung der Grundlagen der Averbis Health Discovery