

SCT-GIT: Deutschsprachige Interface-Terminologie für die Referenzterminologie SNOMED CT

Stefan Schulz

Vollversammlung MUG-IMI, 19.12.2018



Begriffsklärung – SCT-GIT (I)

- **SCT**: SNOMED CT: globaler Terminologiestandard
 - Ehemals „systematized *nomenclature* of medicine-clinical terms“
 - *Ontologiebasierte Referenzterminologie*
 - Fachterme und formale Beschreibungen für alle Bereiche der klinischen Medizin (Anatomie, Erkrankungen, Prozeduren, Organismen, Substanzen, ...)
 - Ca. 341.000 Konzepte, 895.000 englische Terme
 - Herausgegeben von SNOMED International (ehemals IHTSDO)
- SNOMED in D-A-CH
 - Österreich: ab 2019 landesweite Lizenz (Bund + Länder + SVs)
 - Schweiz Lizenz seit 2016, Deutschland in Verhandlungen
- SNOMED in Graz
 - Annotationsvokabular für klinische Texte in CBmed - Projekten
 - S. Schulz in SNOMED-AGs und SIGs (Modelling WG)
 - Diverse internationale Projekte
(*SemanticHealthNet*, *AssessCT*)



Taxonomy
Search
Favorites
Refset

Search

Options

Search Mode: Partial matching search mode

Status: Active concepts only

Group by concept

Filter results by Language

english
172

Filter results by Semantic Tag

disorder
147

procedure
10

situation
10

navigational concept
3

finding
2

Filter results by Module

SNOMED CT core
172
module (core metadata concept)

Type at least 3 characters Example: shou fra

neop lung

172 matches found in 0.223 seconds.

Neoplasm of lung	Neoplasm of lung (disorder)
Lung neoplasm screening	Screening for malignant neoplasm of lung (procedure)
Benign neoplasm of lung	Benign neoplasm of lung (disorder)
Neoplasm of hilus of lung	Neoplasm of hilus of lung (disorder)
Neoplasm of lung (disorder)	Neoplasm of lung (disorder)
H/O: malignant neoplasm of lung	Personal history of primary malignant neoplasm of lung (situation)
Neuroendocrine neoplasm of lung	Neuroendocrine neoplasm of lung (disorder)
Benign neoplasm of hilus of lung	Benign neoplasm of hilus of lung (disorder)
Family history of neoplasm of lung	Family history of neoplasm of lung (situation)
Primary malignant neoplasm of lung	Primary malignant neoplasm of lung (disorder)
Benign neoplasm of lung (disorder)	Benign neoplasm of lung (disorder)
Neoplasm of left lower	Neoplasm of left lower lobe

Concept Details

Summary
Details
Diagram
Expression
Refsets
Members

References

Stated

Inferred

Parents

Malignant tumor of lung (disorder)
Primary malignant neoplasm of intrathoracic organs (disorder)
Primary malignant neoplasm of respiratory tract (disorder)

Primary malignant neoplasm of lung (disorder)
☆

Finding site → Lung structure
Associated morphology → Malignant neoplasm, primary

93880001 | Primary malignant neoplasm of lung (disorder) |
en Primary malignant neoplasm of lung
en Primary malignant neoplasm of lung (disorder)
en Lung cancer

Children (31)

Carcinoma of lung parenchyma (disorder)
Epithelioid hemangioendothelioma of lung (disorder)
Non-Hodgkin's lymphoma of lung (disorder)
Non-small cell lung cancer (disorder)
Overlapping malignant neoplasm of bronchus and lung (disorder)
Pancoast tumor (disorder)
Pleuropulmonary blastoma (disorder)
Primary acinar cell carcinoma of lung (disorder)

Begriffsklärung – SCT-GIT (I)

- GIT: German Interface Terminology / Grazer Interface-Terminologie
- Was ist eine Interfaceterminologie?

The diagram shows a blue box representing a SCT-GIT entry. Inside the box, the text is as follows:

- Primary malignant neoplasm of lung (disorder) (with a star icon and a document icon)
- SCTID: 93880001
- 93880001 | Primary malignant neoplasm of lung (disorder) |
- en Primary malignant neoplasm of lung
- en Primary malignant neoplasm of lung (disorder)
- en Lung cancer

Three yellow callout boxes point to specific parts of the entry:

- „Fully specified name“ points to the main term: Primary malignant neoplasm of lung (disorder)
- „Preferred term“ points to the SCTID and the full name: 93880001 | Primary malignant neoplasm of lung (disorder) |
- „Synonym“ points to the English synonyms: en Lung cancer

Begriffsklärung – SCT-GIT (I)

- GIT: German Interface Terminology / Grazer Interface-Terminologie
- Was ist eine Interfaceterminologie?
 - *“a systematic collection of health care–related phrases (terms) that supports clinicians' entry of patient-related information into computer programs”¹*
 - *„collections of language expressions that actually occur in medical documentation. Such interface terms are typically the building blocks of clinical narratives but also are used as text values for structured data entry”²*
 - Interface-Terme
 - Knapp, mehrdeutig, oft abgekürzt („Ca“)
 - Spezifisch für Fachgebiete, Berufsgruppen, Sprachvarianten („Filiae“)
 - Bedeutungswandel („Neurose“, „larviert depressiv“)

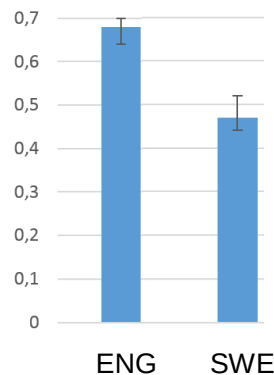
1. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. J Am Med Inform Assoc. 2006 May-Jun;13(3):277-88. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems.
2. Schulz S, Rodrigues JM, Rector A, Chute CG. Interface Terminologies, Reference Terminologies and Aggregation Terminologies: A Strategy for Better Integration. Stud Health Technol Inform. 2017;245:940-944.

„Vocabulary Mismatch“

Beispiel Pubmed

Vorzugsterm (SNOMED CT)	Anzahl	Synonym	Anzahl
Primary malignant neoplasm of lung	0	Lung cancer Bronchial carcinoma	120682 3452
Cerebrovascular accident	3819	Stroke	191559
Block dissection of cervical lymph nodes	1	Neck dissection	7512
Electrocardiographic procedure	1	Electrocardiogram ECG	33670 55120
Backache	3489	Back pain	38132
Capillary blood specimens	32	Capillary blood samples	574

Beispiel ASSESS CT-Parallelkorpus: SNOMED CT – Term coverage



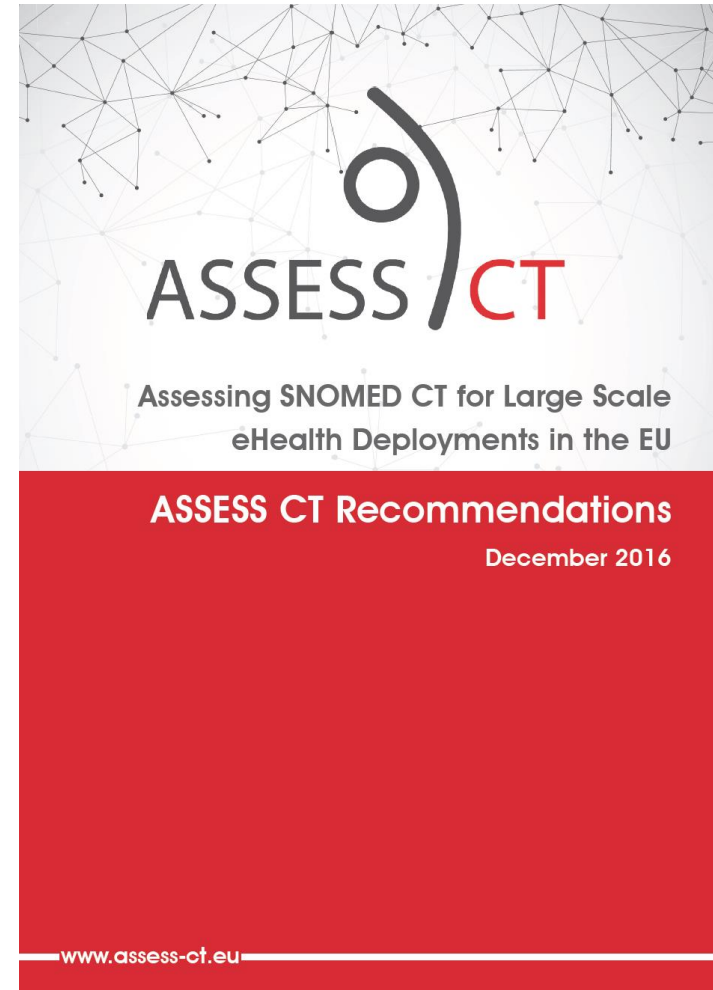
Beispiel Kardiologie Graz

Vorzugsterm (ICD, OPS)	Anzahl	Synonym	Anzahl
Aortenklappenstenose	3749	Aortenstenose	3126
Hirninfarkt	7	Schlaganfall	65
Elektrokardiogramm	0	EKG	12208
Koronare Herzerkrankung	331	KHK	18455
Nicht-ST-Hebungsinfarkt	498	NSTEMI	3839
Magnetresonanztomographie	2	NMR	17



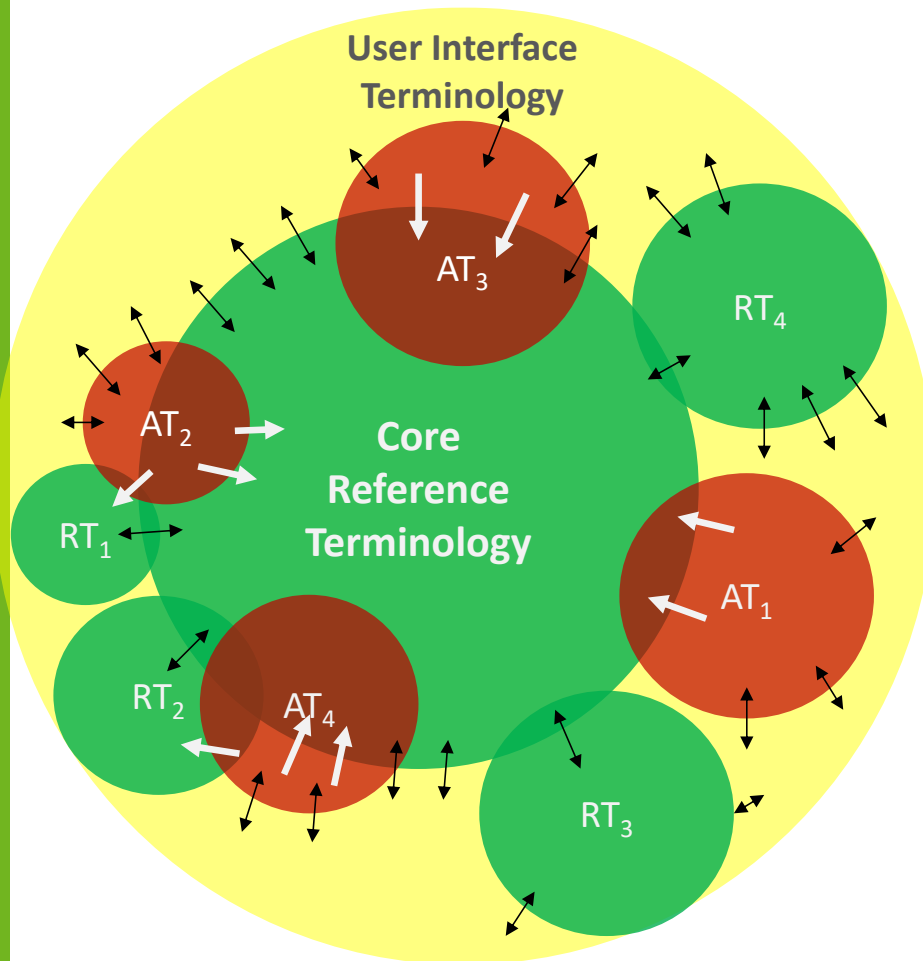
Empfehlung ASSESS-CT zu Referenzterminologien

“Their translation into all languages where they are used is not mandatory, provided this role is fulfilled by user interface terminologies. (...). User interface terminologies acquire their necessary semantic import by linkage to reference terminologies“¹



1. http://assess-ct.eu/fileadmin/assess_ct/final_brochure/assessct_final_brochure.pdf

Empfehlung ASSESS-CT zu Referenzterminologien



AT: aggregation terminology,
RT: reference terminology

„SNOMED CT should be part of an ecosystem of terminologies, including international aggregation terminologies (e.g., the WHO Family of Classifications), and user interface terminologies, which address multilingualism in Europe and clinical communication with multidisciplinary professional language and lay language“¹

1. http://assess-ct.eu/fileadmin/assess_ct/final_brochure/assessct_final_brochure.pdf

Gegenüberstellung Referenzterminologien- Interfaceterminologien

- Referenzterminologien:
 - Primär sprachenunabhängige Repräsentationseinheiten (Konzepte): Eigenschaften der Objekte, die von diesen denotiert werden
 - Maximal eindeutige Labels, unterstützt durch textuelle und / oder formale (ontologische) Definitionen
- Interfaceterminologien:
 - Sammlungen von sprachlichen Ausdrücken, die in schriftlicher und mündlicher Kommunikation verwendet werden.
 - Fundierung durch Verknüpfung zu Referenzterminologien
 - Interfaceterme sind häufig ambig. Ihre Bedeutung wird durch Kontexte der Termbenützung bestimmt



Verwendung von Interfaceterminologien:

- Strukturierte Dateneingabe
 - „Value sets“
 - Dropdown Menu
 - Autocomplete
- Term / Concept Retrieval:
 - Werkzeuge zur manuellen Kodierung
 - Beispiel ID Diacos
 - ICD – Interfaceterminologie: DIMDI-Diagnosenthesaurus
- Natural Language Processing
 - Automatische Extraktion und Kodierung relevanter Information aus klinischen Texten



Positionen zu SNOMED-Lokalisation für den deutschen Sprachraum

- Übersetzung der SNOMED-Vorzugsterme: keine Priorität
 - Kostspielig, zeitaufwändig
 - Beispiel Schweden (10 SEK $\cong$ 1 €)

 primär malign tumör i lunga  

SCTID: 93880001

93880001 | primär malign tumör i lunga |

Artikel Diskussion Visa Redigera Redigera wikitext

Lungcancer [redigera | redigera wikitext]

Lungcancer är en cancersjukdom som orsakas av

<https://www.socialstyrelsen.se/publikationer2011/national-terminology-classification-resource-snomed-ct-project-plan/Documents/Project-plan-Snomed-CT-2007.pdf>

Annual cost of project for the period 2007-2010 and for 2011 (thousands of SEK)

Year	2007 *)	2008	2009	2010	2007-2010	2011
Project management – project manager, administration, including communication and coordination of information	2 700	4 600	4 600	4 600	16 500	1 300
International participation in IHTSDO – board, committees.	500	730	730	730	2 690	750
Web-based training in Snomed CT including development. 1)	500	1 000	500		2 000	
SNOMED CT licence - including startup costs. 2)	2 600	1 050	1 050	1 050	5 750	1 050
Translation of Snomed CT, including IT solution and administration. 3)	6 700	21 200	21 200	21 200	70 300	10 500
Mapping and harmonisation. 4)	1 200	5 350	6 350	4 350	17 250	500
Multi-professional terminology work	1 900	5 650	5 650	1 000	14 200	500
National cooperation and exchange of experience	200	700	700	700	2 300	500
Administration. 5)		1 400	1 400	1 400	4 200	2 400
Evaluation of the project on two occasions. 6)			1 250		1 250	1 250
Support in order to encourage use, pilot project. 7)		5 000	5 000	5 000	15 000	
TOTAL	16 300	46 680	48 430	40 030	151 440	18 750

Positionen zu SNOMED-Lokalisation für den deutschen Sprachraum

- Übersetzung der SNOMED-Vorzugsterme: keine Priorität
 - Kostspielig, zeitaufwändig, schlechte Benutzerakzeptanz¹
 - Ausreichende Kenntnis der englischen sollte bei Zielgruppen (professionellen Kodierern, Implementierern) vorhanden (SNOMED - ärztliche Kodierung kein realistisches Szenario)
- Priorität
 - Inkrementelle Erstellung von Interfaceterminologien
 - Verlinkung von Interfaceterminologien mit Referenzterminologien statt Übersetzung von Referenzterminologien
 - Entwicklung und Validierung von Methoden und Werkzeugen
- Terminologie-Management
 - Referenzterminologien: Top-down, zentralisiert, Standardisierung
 - Interfaceterminologien: Bottom-up, dezentral, „Crowdsourcing“, gemeinsame Erstellung und Nutzung (annotierter) Korpora

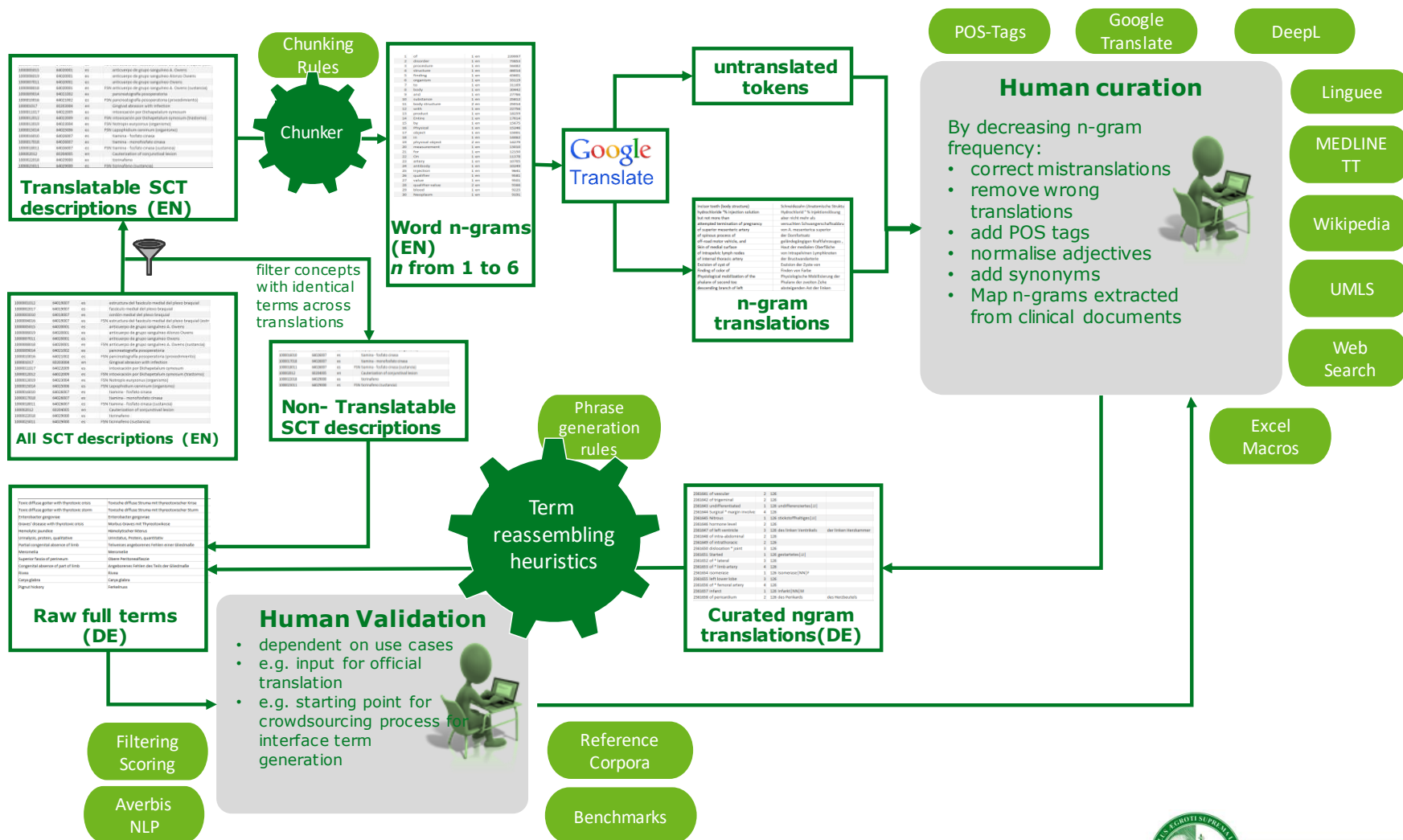
1. Højten AR, Elberg PB, Andersen SK. SNOMED CT adoption in Denmark--why is it so hard? Stud Health Technol Inform. 2014;205:226-30.

MUG-GIT: Erstellung einer deutschen Interface-terminologie für SNOMED CT (I)

- Absicht: Vokabular zur semantischen Annotation deutschsprachiger Kliniktexte
- Limitierte Ressourcen, inkrementelles Vorgehen
- Modularisierung: Zerlegung englischer Terme in Wort-N-Gramme
- Hohe Wiederholungsrate von Teilphrasen wie
 - "*Magnetic resonance imaging*" in 627 SNOMED -Termen
 - "*Second degree burn*" in 166 SNOMED-Termen
- Maschinelle Übersetzung der N-Gramme
- Editieren der N-Gramm-Liste nach Häufigkeit:
 - Manuelle Revision, Ergänzung von Synonymen, POS-Tags
 - Anreicherung durch Terme aus anderen deutschen Terminologien und klinischen Corpora



MUG-GIT: Erstellung einer deutschen Interface-terminologie für SNOMED CT (II)



MUG-GIT: Erstellung einer deutschen Interface-terminologie für SNOMED CT (III)

- Inkaufnahme "schlechter" Übersetzungen und ausufernde Permutationen bei langen SNOMED-CT-Termen
- Kernvokabular:
 - gepflegt durch medizinische Hilfskräfte unter Aufsicht
 - Richtlinien, z.B. keine Schreibvarianten (c/k/z - Problem), Akronyme nur im Kontext (kein Eintrag für "CT", aber für "Schädel-CT"), keine Übersetzung unvollständiger N-Gramme
- Derzeitiger Stand:
 - 4.425.948 Interface-Terme, algorithmisch generiert aus Kernvokabular mit 111.605 deutschen Übersetzungen von 95.298 englischen N-Grammen: Arithmetisches Mittel: 13 Interfaceterme pro SNOMED-CT-Konzept, Median: 4



N-gram - Kernvokabular

N-Gramm (EN)	Tokens	Anzahl	N-Gramm (DE)	N-Gramm (DE)	N-Gramm (DE)
vaginal	1	1478	vaginales JJ	Scheiden-	
fluoroscopic guidance	2	1477	Durchleuchtungskontrolle NN F		
disc	1	1476	Scheibe NN F		
lower limb	2	1473	unteres JJ Extremität NN F	Bein NN N	
brain	1	1468	Gehirn NN N	Hirn NN N	Encephalon NN N
preparation	1	1464	Zubereitung NN F	Aufbereitung NN F	Präparation NN F
method	1	1463	Verfahren NN N	Methode NN F	
of bone	2	1462	des Knochens	_Knochen_	
Red	1	1455	rotes JJ		
Monitoring	1	1453	Überwachung NN F	Monitoring NN N	
Computed	1	1453	berechnetes JJ	Computer-	
phalanx	1	1449	Phalanx NN F		
subsp.	1	1449			
anastomosis	1	1447	Anastomose NN F	Anastomosierung NN F	
vessel	1	1446	Blutgefäß NN N	Gefäß NN N	
Computed tomography	2	1443	Computertomographie NN F		
uterus	1	1436	Uterus NN M	Gebärmutter NN F	
difficulty	1	1432	Schwierigkeit NN F		
elbow	1	1429	Ellbogen NN M	Cubitus NN M	Ellbogengelenk NN N
high	1	1429	hohes JJ		
food	1	1423	Lebensmittel NN N	Speise NN F	Nahrungsmittel NN N
Observation	1	1423	Beobachtung NN F		
using fluoroscopic	2	1422			
unable	1	1421	unfähiges JJ		
Peripheral	1	1419	peripheres JJ		
unable to	2	1418	unfähig zu		
Vascular	1	1417	vaskuläres JJ	Gefäß-	
using fluoroscopic guidance	3	1416	mit Durchleuchtungskontrolle		
Benign neoplasm	2	1415	gutartiges JJ Neubildung NN F	gutartiges JJ Neoplasie NN F	benignes JJ Neoplasie NN F



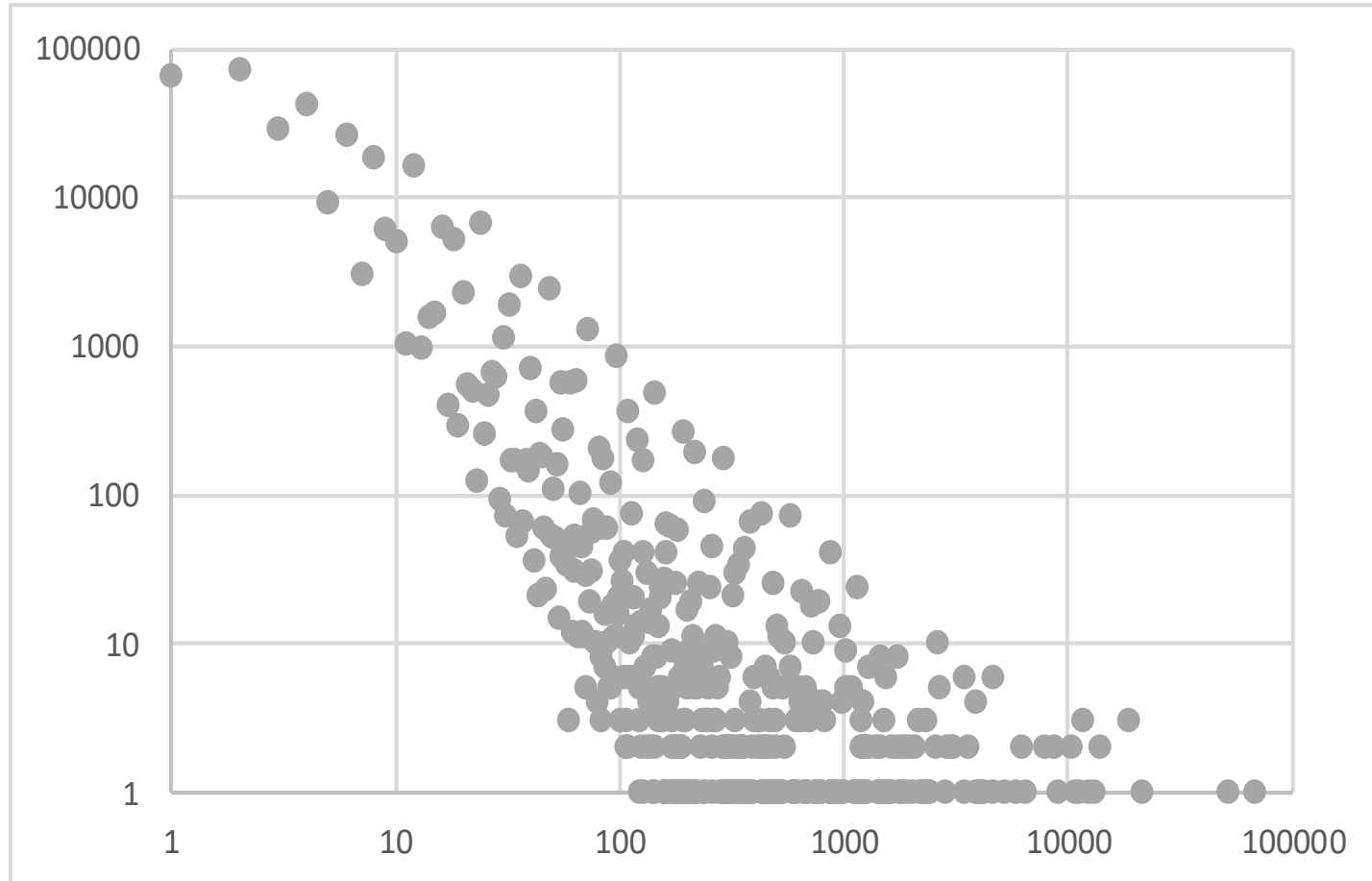
Automatische generierte Interfaceterminologie

	<i>Konzept-ID</i>	<i>Englischer Term</i>	<i>Deutscher Term</i>
20170315_240011_002	126952004	Neoplasm of brain	Gehirneubildung
20170315_240011_003	126952004	Neoplasm of brain	Neubildung des Hirns
20170315_240011_004	126952004	Neoplasm of brain	Hirneubildung
20170315_240011_005	126952004	Neoplasm of brain	Neoplasie des Gehirns
20170315_240011_006	126952004	Neoplasm of brain	Gehirneoplasie
20170315_240011_007	126952004	Neoplasm of brain	Neoplasie des Hirns
20170315_240011_008	126952004	Neoplasm of brain	Hirneoplasie
20170315_240011_009	126952004	Neoplasm of brain	Neoplasma des Gehirns
20170315_240011_010	126952004	Neoplasm of brain	Gehirneoplasma
20170315_240011_011	126952004	Neoplasm of brain	Neoplasma des Hirns
20170315_240011_012	126952004	Neoplasm of brain	Hirneoplasma
20170315_241010_001	126953009	Neoplasm of cerebrum	Neubildung des Großhirns
20170315_241010_002	126953009	Neoplasm of cerebrum	Neoplasie des Großhirns
20170315_241010_003	126953009	Neoplasm of cerebrum	Neoplasma des Großhirns
20170315_242015_001	126954003	Neoplasm of frontal lobe	Neubildung des Frontallappens
20170315_242015_002	126954003	Neoplasm of frontal lobe	Neubildung des Lobus frontalis
20170315_242015_003	126954003	Neoplasm of frontal lobe	Neoplasie des Frontallappens
20170315_242015_004	126954003	Neoplasm of frontal lobe	Neoplasie des Lobus frontalis
20170315_242015_005	126954003	Neoplasm of frontal lobe	Neoplasma des Frontallappens
20170315_242015_006	126954003	Neoplasm of frontal lobe	Neoplasma des Lobus frontalis
20170315_243013_001	126955002	Neoplasm of temporal lobe	Neubildung des Temporallappens
20170315_243013_002	126955002	Neoplasm of temporal lobe	Neubildung des Lobus temporalis
20170315_243013_003	126955002	Neoplasm of temporal lobe	Neoplasie des Temporallappens
20170315_243013_004	126955002	Neoplasm of temporal lobe	Neoplasie des Lobus temporalis
20170315_243013_005	126955002	Neoplasm of temporal lobe	Neoplasma des Temporallappens



Verteilung

Häufigkeit



Generierte Synonyme
pro Konzept



SCT-GIT: offene Fragen / nächste Schritte

- Studentische Arbeitskräfte, Training
- Verbessertes Tooling
- Softwarepflege / Dokumentation
- Manuelle und automatische Qualitätskontrolle
- Benchmarking
- Einbindung von maschinellen Übersetzungstools
- Termakquisition aus klinischen Korpora
- Umgang mit Abkürzungen und Akronymen (PhD Oleynik)
- Anwendung der Methode auf andere Sprachen (Spanisch → Portugiesisch mit PUC-Paraná in Curitiba, Brasilien)



SCT-GIT: Ko-operative Entwicklung einer deutschen Interface-Terminologie

- Günstige Rahmenbedingungen
 - Datenintegration / Sekundärnutzung / semantische Suche: Thema in geförderten Großprojekten (D: BMBF-MI, A: CBmed, ELGA)
 - Wachsendes Interesse an internationalen Terminologien (SNOMED CT, LOINC, RadLex...) und Ontologien (GO, HPO, ...)
 - SNOMED-CT-Einführung im deutschsprachigen Raum
 - A: National Release Center @ ELGA GmbH
 - CH: eHealth Suisse
 - DE: De-Facto-SNOMED-CT-Nutzung in BMBF-MI-Projekten
- Synergieeffekte vs. Ressourcenverschwendung
- Idee: Crowdsourcing-Plattform für Entwicklung deutschsprachigen Interface-Terminologien: GIT-CP



Mögliche Spezifikation für GIT-CP

- Web-basierte Crowdsourcing-Plattform
- Registrierung als User → Commitment zu Kooperation
- Zentrales Datenelement:
Mapping Interface Term – Externer Code
"DM" - 81827009 |*Diameter (qualifier value)*
- Wichtige Attribute:
 - Ersteller, Erstellungsart, Datum, klinisches Fachgebiet, Nutzergruppe
Max Muster, manuell, 20170803, Dermatologie Graz, Ärzte
 - Beispielannotation, z.B.
"ein 3 cm im DM haltender Tumor"
 - Validierung / Kommentierung durch andere User
Sven Piefke, 20180912, ★★★★★
„Hä?? Beispiel unverständlich!“



GIT-CP – Offene Fragen

- Technisch
 - Versionierung (GIT – Zielsysteme)
 - Schnittstellen zu lokalen Annotationsplattformen
 - Intelligente Tools (z.B. recommender services)
- Rechtlich / Organisatorisch
 - Koordination
 - Nachhaltige Finanzierung
 - Qualitätssicherung
 - Eigentumsrechte
 - Verwertung
 - Datenschutz



Fazit

- Text Mining deutschsprachiger Medizintexte benötigt Interfaceterminologien, die den alltäglichen Sprachgebrauch abbilden
- Die Verlinkung von Interfaceterminologien mit Referenzterminologien hat Priorität gegenüber der Übersetzung von Referenzterminologien
- Die Referenzterminologie SNOMED-CT gewinnt an Bedeutung und ist in Österreich ab 2019 verfügbar
- IMI konzentriert SNOMED-CT-Kompetenz im deutschsprachigen Raum
- Terminologiemanagement
 - Referenzterminologien: top-down, zentralisiert
 - Interfaceterminologien: bottom-up, dezentral
- Zeit ist reif für die kooperative, verteilte Erstellung von medizinischen Interfaceterminologien

