



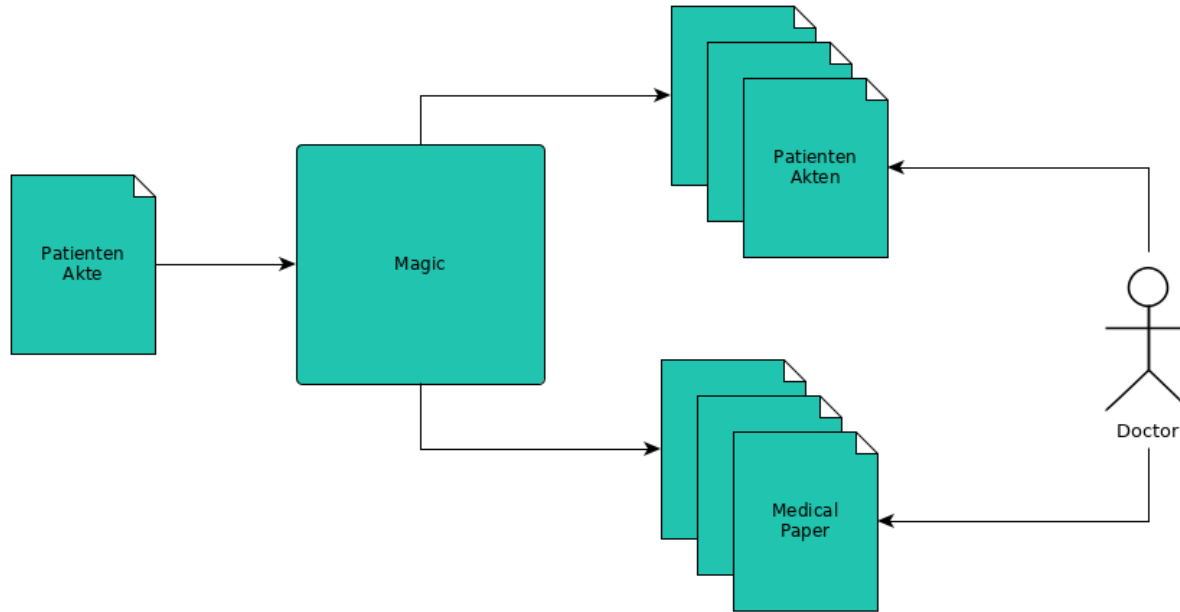
CLINICAL NAMED ENTITY LINKING

Julius Brückner
Kevin Mattutat
Arnold Schwarz

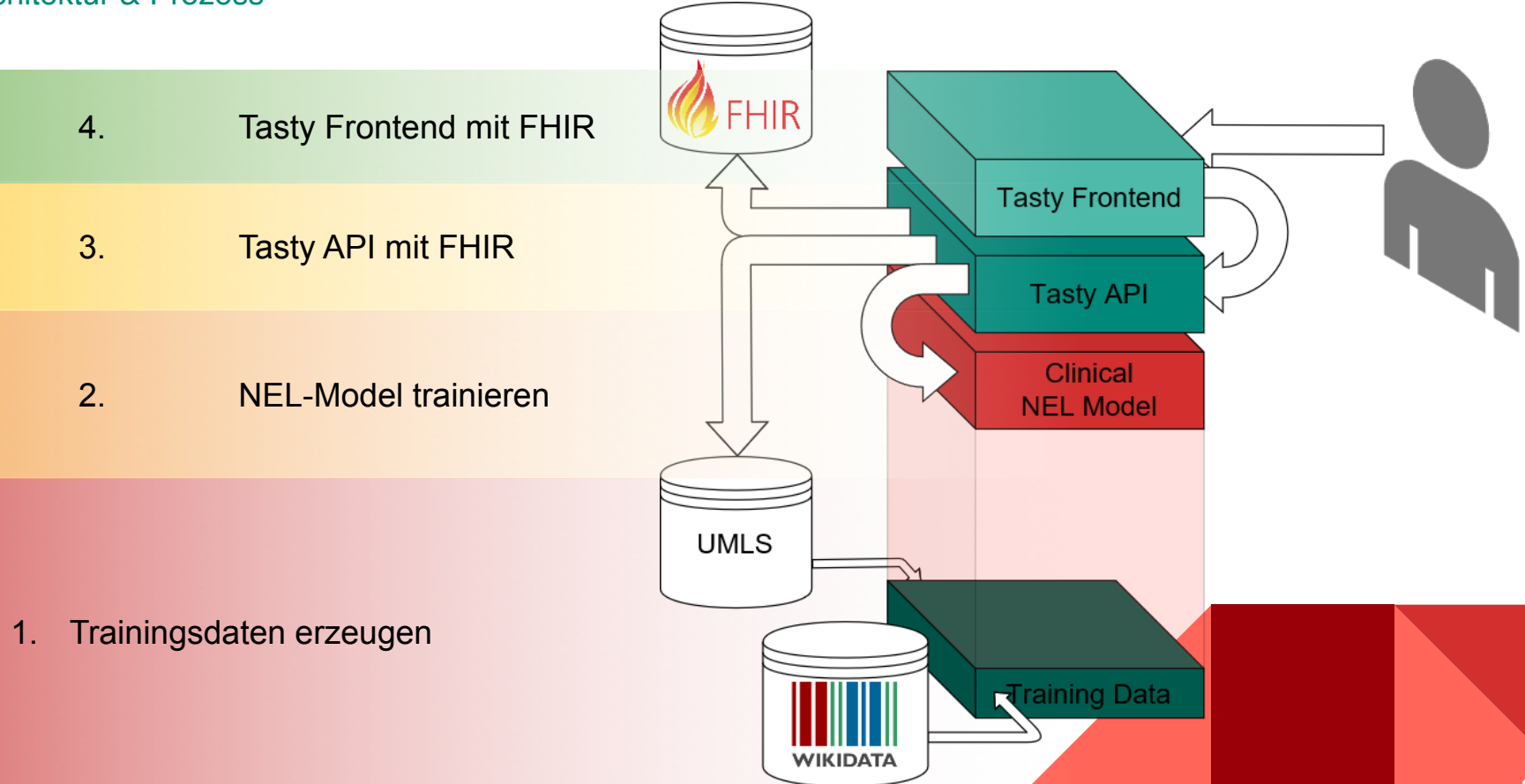
Kleines Recap

- Was war eigentlich unsere Aufgabe?
- Wie hatten wir geplant zu Lösen?
- Was hat uns davon Abgehalten?

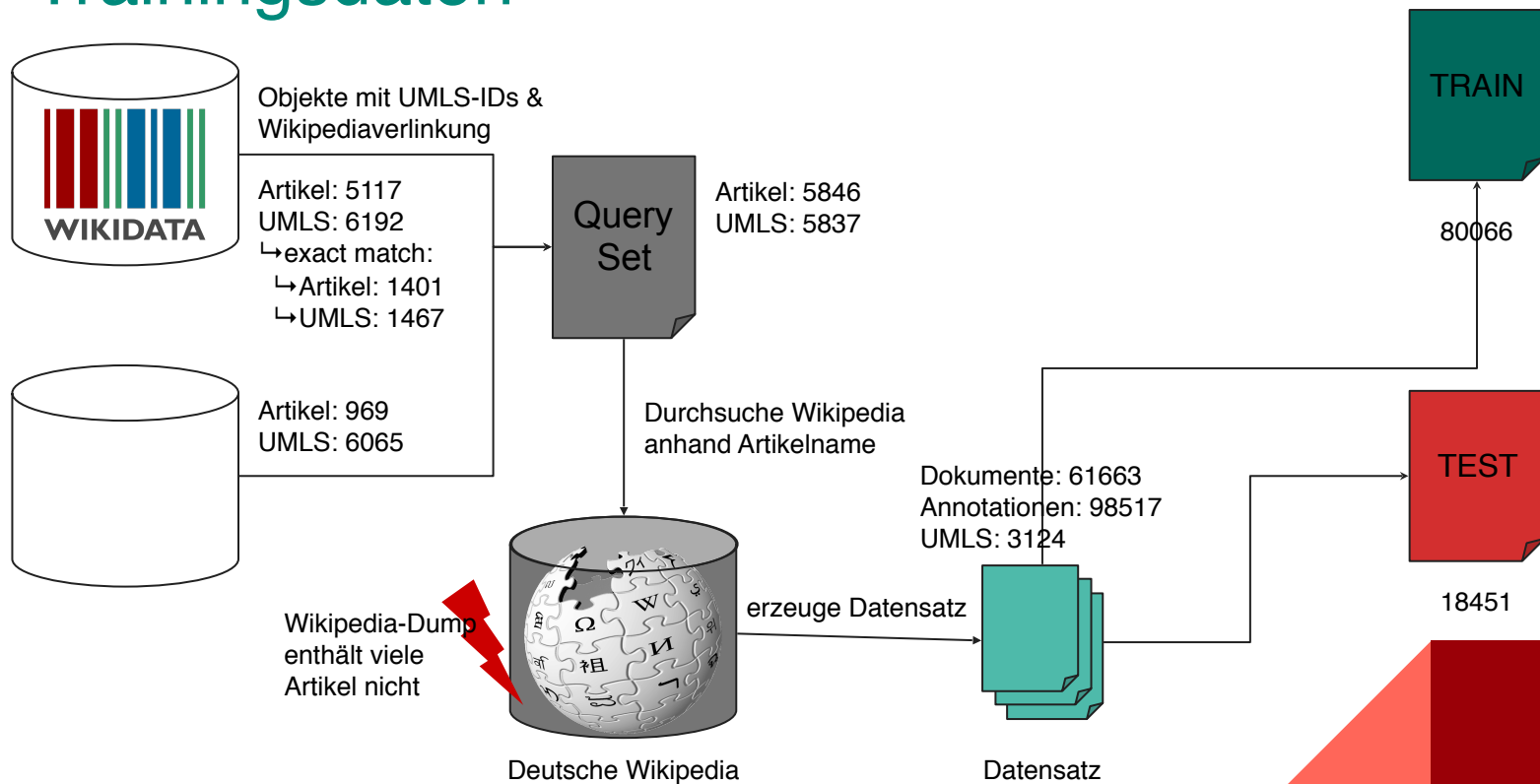
Das große Ganze



Architektur & Prozess

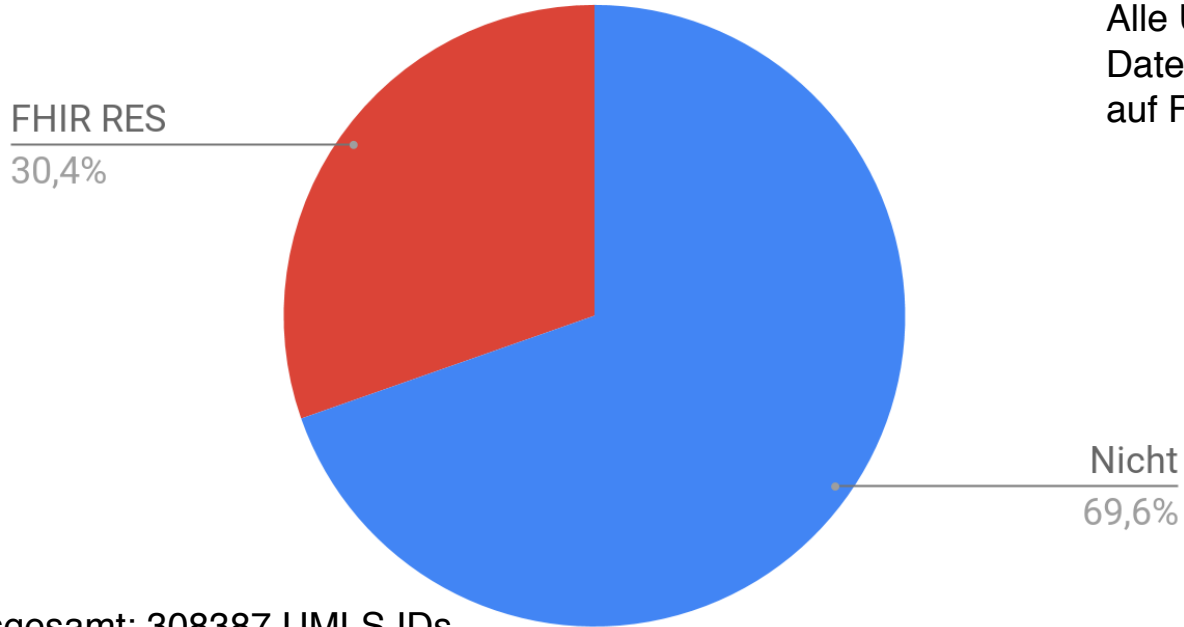


Trainingsdaten



FHIR Daten Analyse

Identifizierbare FHIR Ressourcen

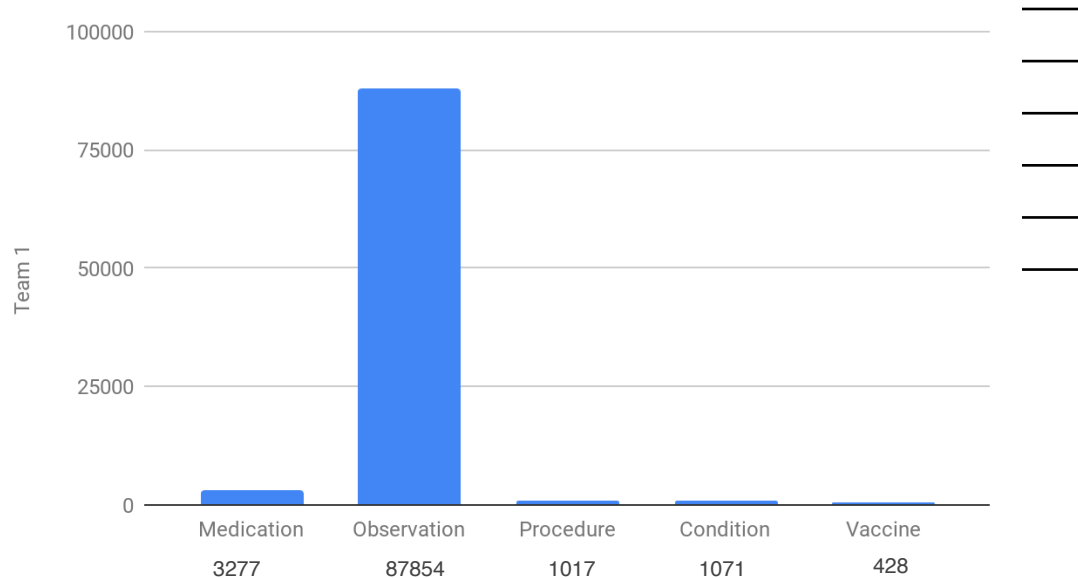


Alle UMLS IDs aus der UMLS Datenbank und ihre Mapbarkeit auf FHIR Ressourcen.

Insgesamt: 308387 UMLS IDs

FIHR Daten Analyse

FIHR Ressourcen Verteilung



Verteilung der FHIR
Ressourcen-Typen auf die
mapbaren UMLS IDs

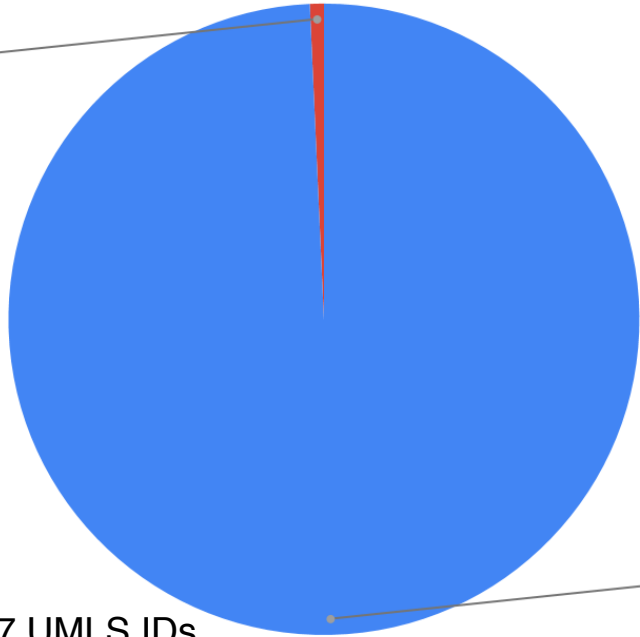
Insgesamt: 93647 UMLS IDs

Trainingsdaten Analyse (WIKIDATA)

Trainierbare UMLS IDs

Trainingsdaten

0,7%
2192



UMLS IDs die mit genügend Text beschrieben sind und daher als Trainingsdaten genutzt werden können. Auf die Anzahl aller UMLS IDs in der Datenbank.

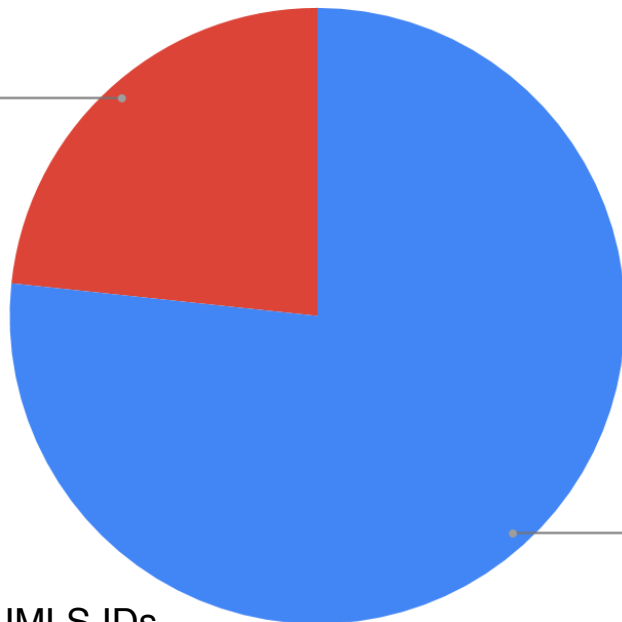
Nicht
99,3%

Insgesamt: 308387 UMLS IDs

Trainingsdaten zu FHIR Daten

Trainingsdaten auf FHIR Ressourcen

Auf FHIR
23,3%



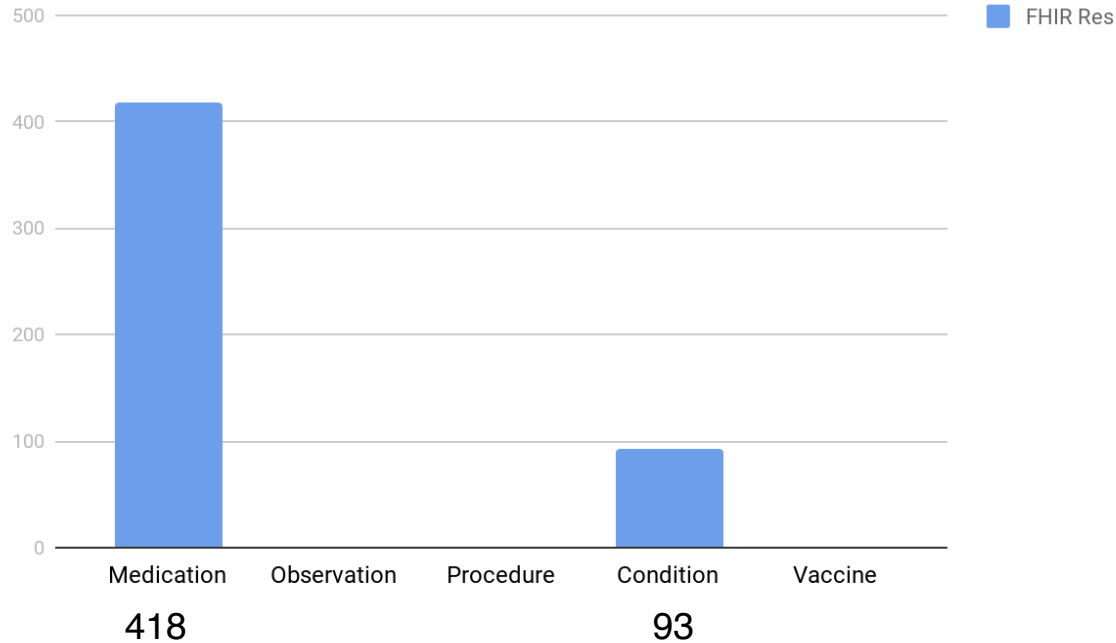
Nicht auf FHIR
76,7%

UMLS IDs in den Trainingsdaten die sich auf FHIR Ressourcen mappen lassen

Insgesamt: 2192 UMLS IDs

Verteilung der FHIR Res. in WIKIDATA

FHIR Res Verteilung

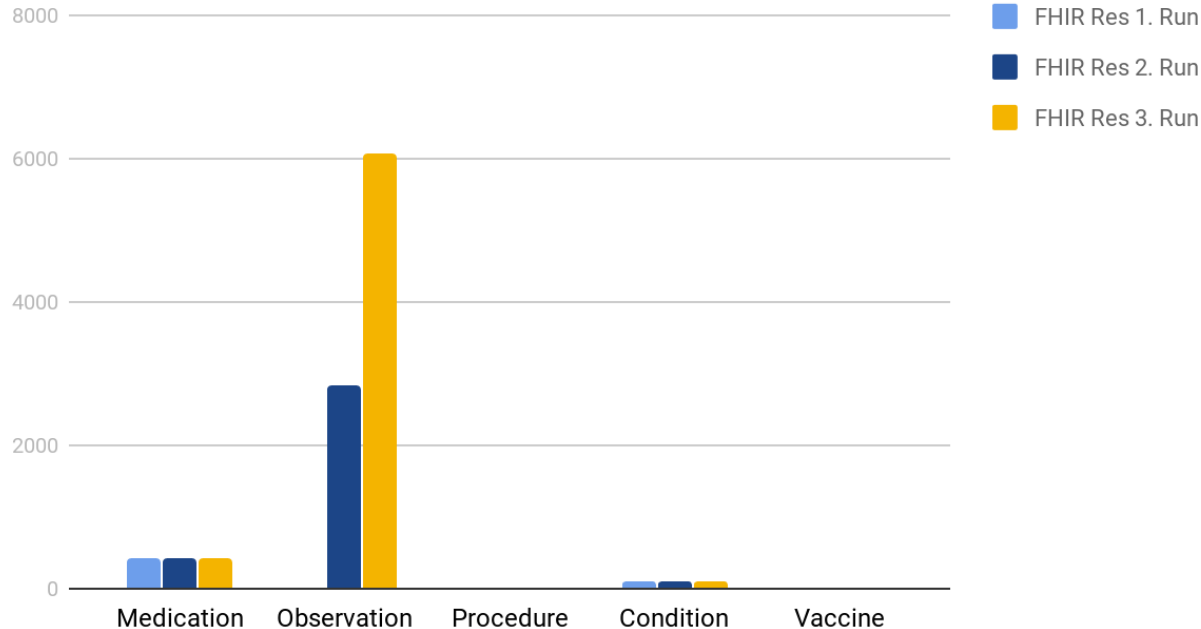


Verteilung der FHIR Ressourcen-Typen auf die mapbaren und gelernten UMLS IDs

Insgesamt: 511 UMLS IDs

Verteilung der FHIR Res. in WIKIDATA

FHIR Res Verteilung



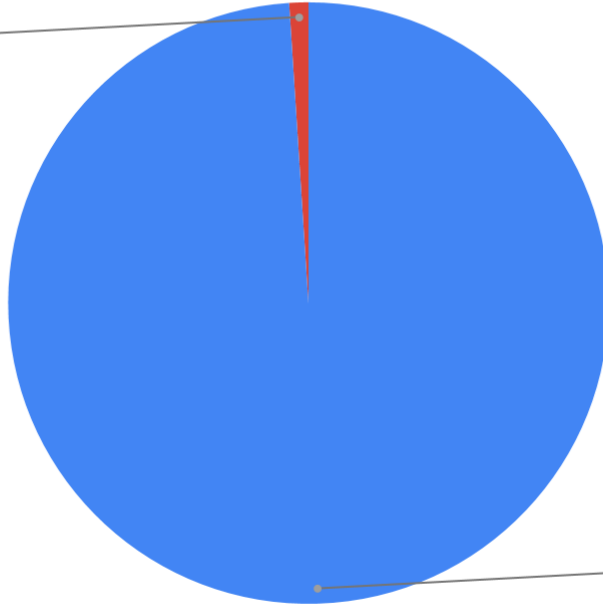
Verteilung der FHIR Ressourcen-Typen auf die mapbaren und gelernten UMLS IDs

Mithilfe der LOINC Datenbank konnten LOINC IDs auf Wikipedia und UMLS IDs gemappt werden

Trainingsdaten Analyse

Tranirbare UMLS IDs

Trainingsdaten
1,0%



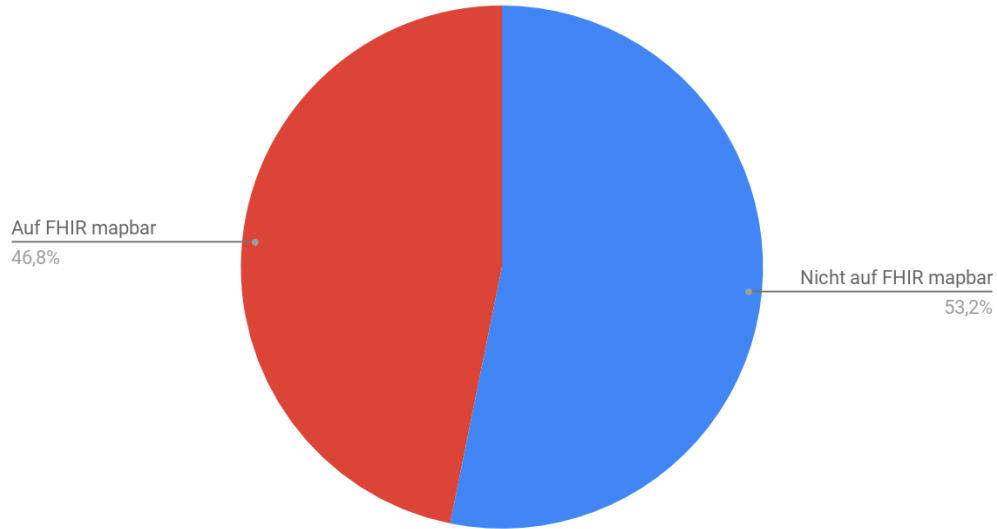
Nicht Trainierbar
99,0%

Nach dem aus LOINC weitere Dokumente extrahiert werden konnten. Wegen dopplungen konnten aber nur 969 Verwertet werden.

Insgesamt: 308387 UMLS IDs

Trainingsdaten zu FHIR Daten

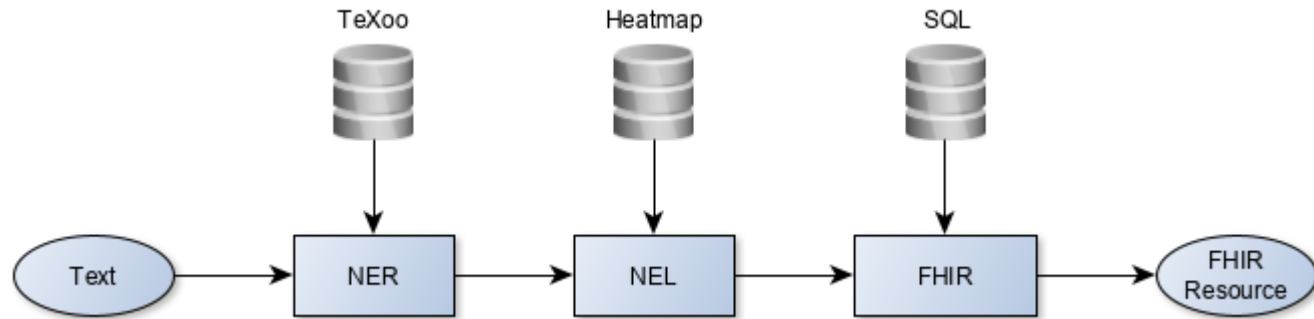
Points scored



UMLS IDs in den Trainingsdaten die sich auf FHIR Ressourcen mappen lassen steigt erheblich.

Insgesamt: 3161 UMLS IDs

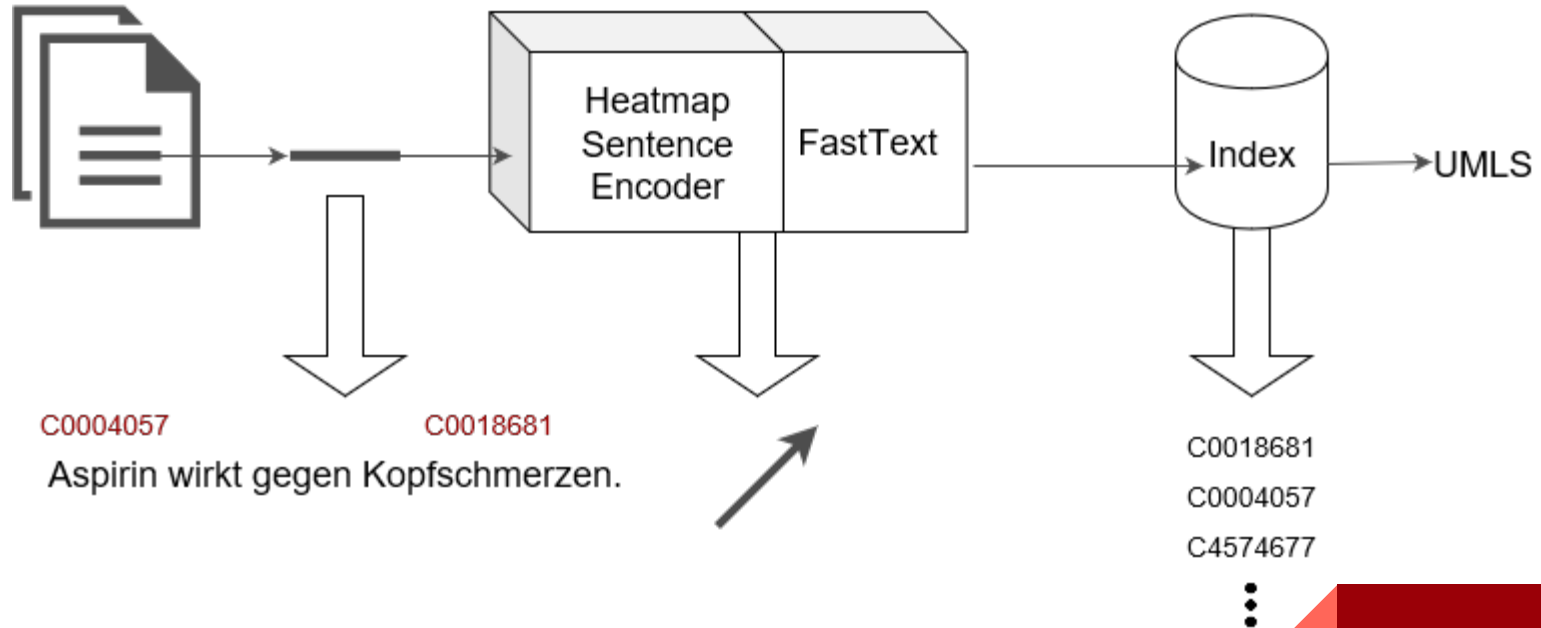
Datenfluss im Backend



Implementierte FHIR Ressourcen

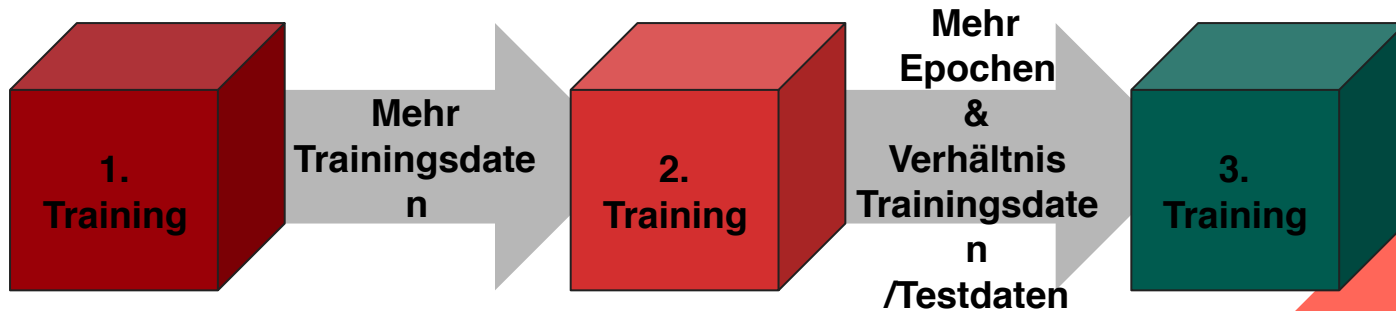
ImmunizationResource -identifier: Identifier[*] -status: String -statusReason: CodeableConcept -vaccineCode: CodeableConcept -patient: Reference -encounter: Reference -occurrenceDateTime: Date -occurrenceString: String -recorded: Date -primarySource: Boolean -reportOrigin: CodeableConcept -location: Reference -manufacturer: Reference -lotNumber: String -expirationDate: Date -site: CodeableConcept -route: CodeableConcept -doseQuantity: Quantity -performer: ImmunizationPerformer[*] -note: TextAnnotation[*] -reasonCode: CodeableConcept[*] -reasonReference: Reference[*] -isSubpotent: Boolean -subpotentReason: CodeableConcept[*] -education: ImmunizationEducation[*] -programEligibility: CodeableConcept[*] -fundingSource: CodeableConcept -reaction: ImmunizationReaction[*] -protocolApplied: ImmunizationProtocol[*]	ProcedureResource -identifier: Identifier[*] -instantiatesCanonical: String[*] -instantiatesUri: URI[*] -basedOn: Reference[*] -partOf: Reference[*] -status: String -statusReason: CodeableConcept -category: CodeableConcept -code: CodeableConcept -subject: Reference -encounter: Reference -performedString: String -recorder: Reference -asserter: Reference -performer: ProcedurePerformer[*] -location: Reference -reasonCode: CodeableConcept[*] -reasonReference: Reference[*] -bodySite: CodeableConcept[*] -outcome: CodeableConcept -report: Reference[*] -complication: CodeableConcept[*] -complicationDetail: Reference[*] -followUp: CodeableConcept[*] -note: TextAnnotation[*] -focalDevice: ProcedureFocalDevice[*] -usedReference: Reference[*] -usedCode: CodeableConcept[*]
ConditionResource -identifier: Identifier[*] -clinicalStatus: CodeableConcept -verificationStatus: CodeableConcept -category: CodeableConcept[*] -severity: CodeableConcept -code: CodeableConcept -bodySite: CodeableConcept[*] -subject: Reference -encounter: Reference -onsetString: String -abatementString: String -recordedDate: Date -recorder: Reference -asserter: Reference -stage: ConditionStage[*] -evidence: ConditionEvidence[*] -note: TextAnnotation[*]	ObservationResource -status: String -code: CodeableConcept -identifier: Identifier[*] -onPlan: Reference[*] -partOf: Reference[*] -category: CodeableConcept[*] -subject: Reference -focus: Reference[*] -encounter: Reference -effectiveDateTime: Date -issued: Instant -performer: Reference[*] -valueString: String -dataAbsentReason: CodeableConcept -interpretation: CodeableConcept[*] -note: TextAnnotation[*] -bodySite: CodeableConcept -method: CodeableConcept -specimen: Reference -device: Reference -referenceRange: Interpretation[*] -hasMember: Reference[*] -derivedFrom: Reference[*] -component: ResultComponent[*]
MedicationResource -identifier: Identifier[*] -code: CodeableConcept -status: String -manufacturer: Reference -form: CodeableConcept -amount: Ratio -ingredient: Ingredient[*] -batch: MedicationBatch	

Training und Evaluation



Training und Evaluation

Training	Train	Test	Epochs	A@1	A@3	P@1	P@3	R@1	R@3	MAP
1	79724	7591	2	4.58	7.94	4.58	2.65	4.58	7.94	8.35
2	89892	8625	2	5.54	10.02	5.54	3.34	5.54	10.02	10.17
3	80066	18451	4	9.49	16.08	9.49	5.36	9.49	16.08	15.54



Training und Evaluation - Probleme

Problem:

Erworbene Taubheit (Innenohrschaden) kann als Folge von z.B. (Meningokokken-)Meningitis, Enzephalitis, Scharlach, Masern, Tuberkulose, Osteomyelitis, Mittelohr-Erkrankungen, Otosklerose, (Baro-)Trauma u. a. (bei absoluter Taubheit stets mit Innenohr- oder Hörnervbeteiligung) auftreten.

(Insgesamt 10 UMLS-ID)

Lösung:

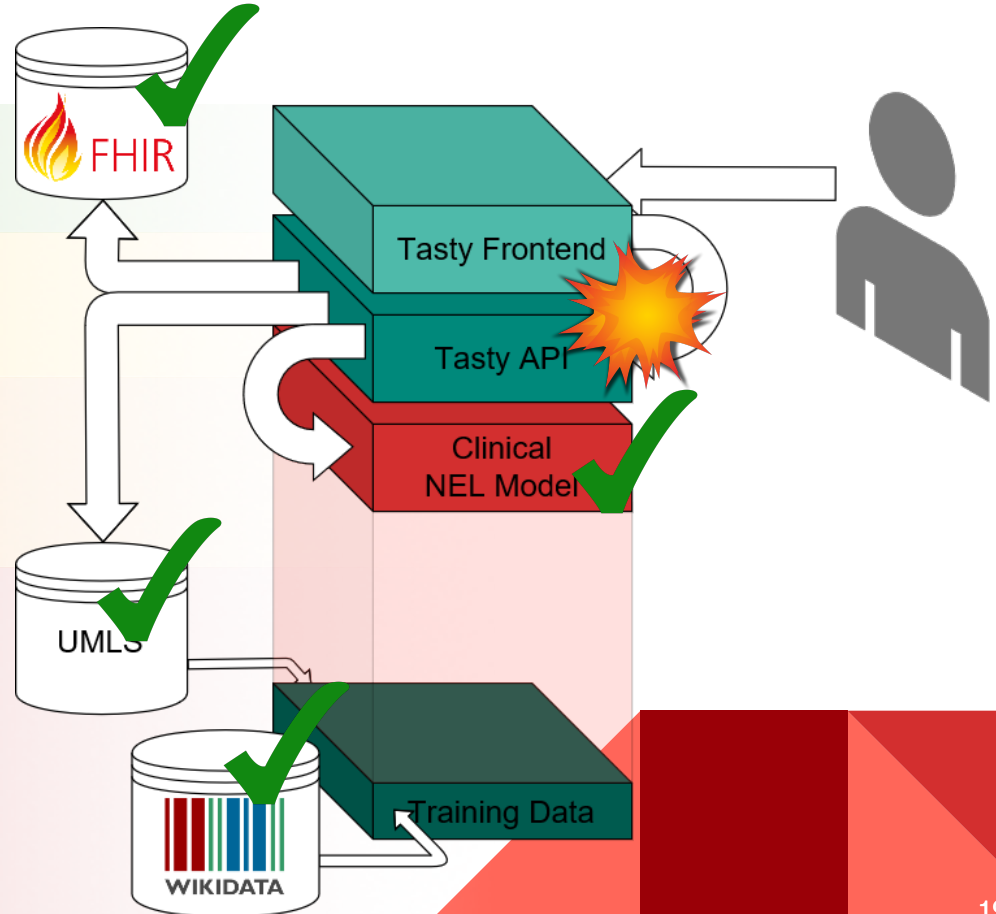
Solche Ausreißer herausfiltern

Weitere Verbesserungen:

- Besser gelabelte Daten
- dem Ziel entsprechender Daten

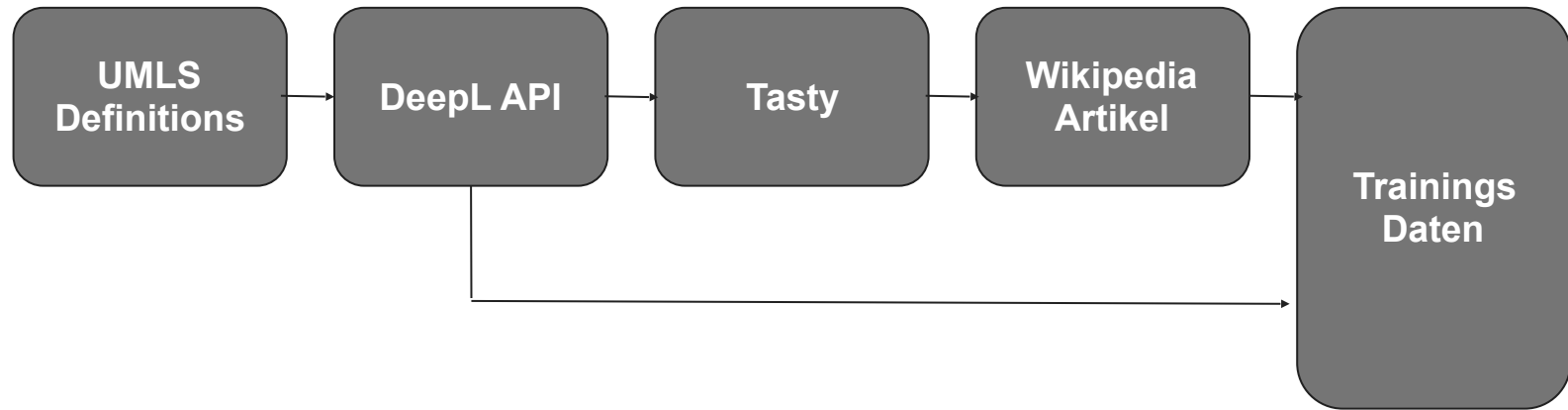
Was fehlt?

1. Trainingsdaten erzeugen
2. NEL-Model trainieren
3. Tasty API mit FHIR
4. Tasty Frontend mit FHIR



Ausblick

- Verbesserung der Datenlage durch Übersetzung



- Fixen der Tasty API
 - Verursacht durch fehlerhafte Interpretation der Konfiguration
 - Und was sonst noch schief hängt