



Nutzenpotential von FHIR für unstrukturierte Daten

Vortrag

Datum	Zeit	Ort
22. April 2020	15:35 - 15:55 Uhr	_Stage A, Halle 6.2

Thema dieses Vortrags ist die FHIR-konforme Repräsentation von Informationen, die durch Text-Mining aus klinischen Dokumenten extrahiert werden. Wichtige Anhaltspunkte zu Diagnosen, Symptomen, Scores oder Medikamenten liegen oft nur in unstrukturierter Form vor. Mittels Text-Mining-Verfahren extrahiert, werden sie FHIR-konform zugänglich gemacht und können im weiteren Behandlungskontext ebenso wie in der klinischen Forschung Anwendung finden. Auch administrative Prozesse wie die DRG-Kodierung können davon profitieren, so dass Krankenhäuser die Behandlungsqualität zu erhöhen, Ineffizienzen reduzieren und zusätzliche Einnahmequellen erschließen können.

Inhalte der elektronischen Patientenakte liegen oft nur in textueller Form vor, beispielsweise in Arztbriefen, der Verlaufsdokumentation, in Radiologie- oder Pathologieberichten. Die zunehmende Verbreitung automatischer Spracherkennung fördert diesen Trend. Es ist für Krankenhäuser von enormem Wert, auch textuelle Inhalte maschinell zu erschließen, um sie weiter zu verwerten: sei es zur Bereitstellung lernender, diagnoseunterstützender Systeme, sei es für die klinische Forschung, oder auch, um administrative Prozesse wie die DRG-Kodierung zu automatisieren. Hierzu müssen die Freitextdaten durch Verfahren der natürlichen Sprachverarbeitung (NLP, Text-Mining) in eine semantisch interoperable Form überführt werden. Als Repräsentationsformat für die so extrahierten Informationen bietet sich FHIR als führender Standard an. Wir haben einen Vorschlag erarbeitet, wie sich Extrakte aus Arztbriefen in FHIR repräsentieren lassen. Zusätzlich wurden zwei Erweiterungen des FHIR-Standards entwickelt, die den FHIR-konformen Einsatz von Text Mining ermöglichen.

Dieser Arbeit basiert auf dem kommerziellen Text-Mining Produkt Health Discovery der Averbis GmbH. Aus klinischen Texten extrahiert das Produkt über 50 verschiedene Informationstypen wie Diagnosen, Prozeduren, Medikamente, Laborwerte und Zeitangaben, um sie auf Terminologien wie ICD-10, OPS, ABDAMED oder LOINC abzubilden. Bereits jetzt verwaltet Health Discovery die klinischen Originaltexte ebenso wie die daraus gewonnenen Text-Mining-Extrakte in einem internen Datenmodell, welches dem von FHIR ähnelt. Im Rahmen dieser Arbeit wurden diese in das FHIR-Datenmodell überführt.

Die klinischen Texte selbst werden in FHIR in einer "Binary"-Ressource repräsentiert. Für die Text-Mining-Extrakte werden entsprechende FHIR-Ressourcen wie "Condition", "MedicationStatement" oder Observation genutzt. Für jede dieser "klinischen" Ressourcen wird zusätzlich eine FHIR-Ressource vom Typ "Provenance" erzeugt, welche Texte mit den dazugehörigen FHIR-Ressourceninstanzen verbindet. Alle auf diese Weise generierten Instanzen von FHIR-Ressourcen werden in einem "Bundle" vom Typ "Collection" zusammengefasst. Die "Provenance"-Ressource wird zudem um zwei FHIR-Extensions erweitert.

Die erste betrifft die Konfidenz einer Text-Mining-Annotation, welche dem Anwender offenlegt, wie wahrscheinlich ein Extrakt korrekt ist. Ihr Attribut "value" kann Werte zwischen 0 und 1 annehmen.

Die zweite Erweiterung betrifft die genaue Position der einem Extrakt zugrundeliegenden Textpassage im Originaltext, da bei Text-Mining die Rückverfolgbarkeit der gewonnenen Extrakte von großer Bedeutung ist. Hierfür wurde eine CharacterRange-Extension definiert, welche die Attribute "low" (Startposition im Text) und "high" (Endposition im Text) enthält.

Das Ergebnis dieser Arbeit ist eine FHIR-konforme Repräsentation klinischer Texte und ihrer durch Text-Mining gewonnenen semantischen Extrakte. Durch zwei Erweiterungen wurden bestehende Lücken in FHIR in Bezug auf das Text-Mining geschlossen, so dass einem FHIR-konformen Einsatz von Text-Mining für Zwecke wie Behandlungsunterstützung, klinische Forschung und Kodierung nichts im Wege steht.

Akteure

Speaker:



[Dr. Philipp Daumke](#), Geschäftsführer, Averbis GmbH