

Clinical Text Mining on FHIR

Philipp Daumke^a, Kai U. Heitmann^{b,c}, Simone Heckmann^{b,c}, Catalina Martínez-Costa^d, **Stefan Schulz**^{a,d}

^a *Averbis GmbH, Freiburg, Germany,*

^b *Gefyra GmbH, Hürth, Germany,*

^c *HL7 Germany, Berlin, Germany*

^d *Institute for Medical Informatics, Statistics and Documentation,
Medical University of Graz, Austria*

“

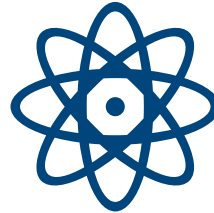
To build software that aims at removing the differences between structured and unstructured data



FIND



Turn text into actionable
information



UNDERSTAND



Understand and automate
cognitive processes



PREDICT



Enable better
business critical decisions

“

To build software that aims at removing the differences between structured and unstructured data



FIND



Turn text into actionable
information

Natural language processing (NLP)
Information extraction:



UNDERSTAND



Understand and automate
cognitive processes



PREDICT



Enable better
business critical decisions



To build software that aims at removing the differences between structured and unstructured data



FIND

Turn text into actionable
information

Natural language processing (NLP) Information extraction:

Klinischer Befund:

72-jährige Patientin in ausreichendem AZ, Körpergröße 1,57 m, Körpergewicht 72,8 kg. Haut und sichtbare Schleimhäute gut durchblutet. Kein Ikterus, keine Zyanose. Mundhöhle o.B.. Keine Struma, keine peripheren Lymphknotenschwellungen. Über Herz und Lungen war der klinische Befund unauffällig. RR bds. 160/80 mm Hg, Pulsfrequenz regelmäßig. Leber und Milz nicht vergrößert. Nierenlager klopfschmerzfrei. Keine Varizen, keine Oedeme. Fußpulse beiderseits tastbar. MER seitengleich.



Admission General Surgery	
Age	72
Gender	F
Weight	72.8
Height	157
BP	160/80
General status	GWB
Cyanosis	neg
Jaundice	neg
Edemas	neg

Text (unstructured / semistructured)

Information template

INFORMATION MODEL STANDARD: HL7 FHIR



- Resources:
 - data model components defining sets of properties that describe structures for data acquisition
 - models for frequently occurring documentation and information exchange tasks
- Examples:
 - *MedicationRequest*
 - *AdverseEvent*
 - *Procedure*
 - *Condition*

Condition (DomainResource)
identifier : Identifier [0..*] clinicalStatus : CodeableConcept [0..1] « ConditionClinicalStatusCodes! » verificationStatus : CodeableConcept [0..1] « ConditionVerificationStatus! » category : CodeableConcept [0..*] « ConditionCategoryCodes+ » severity : CodeableConcept [0..1] « Condition/DiagnosisSeverity? » code : CodeableConcept [0..1] « Condition/Problem/DiagnosisCo...?? » bodySite : CodeableConcept [0..*] « SNOMEDCTBodyStructures?? » subject : Reference [1..1] « Patient Group » encounter : Reference [0..1] « Encounter » onset[x] : Type [0..1] « dateTime Age Period Range string » abatement[x] : Type [0..1] « dateTime Age Period Range string » recordedDate : dateTime [0..1] recorder : Reference [0..1] « Practitioner PractitionerRole Patient RelatedPerson » asserter : Reference [0..1] « Practitioner PractitionerRole Patient RelatedPerson » note : Annotation [0..*]

<https://www.hl7.org/fhir>

FOCUS OF THE WORK



Klinischer Befund:

72-jährige Patientin in ausreichendem AZ, Körpergröße 1,57 m, Körpergewicht 72,8 kg. Haut und sichtbare Schleimhäute gut durchblutet. Kein Ikterus, keine Zyanose. Mundhöhle o.B.. Keine Struma, keine peripheren Lymphknotenschwellungen. Über Herz und Lungen war der klinische Befund unauffällig. RR bds. 160/80 mm Hg, Pulsfrequenz regelmäßig. Leber und Milz nicht vergrößert. Nierenlager klopfschmerzfrei. Keine Varizen, keine Oedeme. Fußpulse beiderseits tastbar. MER seitengleich.

NLP

Condition (DomainResource)
identifier : Identifier [0..*]
clinicalStatus : CodeableConcept [0..1] « ConditionClinicalStatusCodes! »
verificationStatus : CodeableConcept [0..1] « ConditionVerificationStatus! »
category : CodeableConcept [0..*] « ConditionCategoryCodes+ »
severity : CodeableConcept [0..1] « Condition/DiagnosisSeverity? »
code : CodeableConcept [0..1] « Condition/Problem/DiagnosisCo...?? »
bodySite : CodeableConcept [0..*] « SNOMEDCTBodyStructures?? »
subject : Reference [1..1] « Patient Group »
encounter : Reference [0..1] « Encounter »
onset[x] : Type [0..1] « dateTime Age Period Range string »
abatement[x] : Type [0..1] « dateTime Age Period Range string »
recordedDate : dateTime [0..1]
recorder : Reference [0..1] « Practitioner PractitionerRole Patient RelatedPerson »
asserter : Reference [0..1] « Practitioner PractitionerRole Patient RelatedPerson »
note : Annotation [0..*]

- Harmonization of clinical text mining output with FHIR.
- Identification of gaps in the current FHIR specification related to content essential for text mining.
- Introduction of FHIR extensions that allow text mining results to be represented without loss.

UIMA

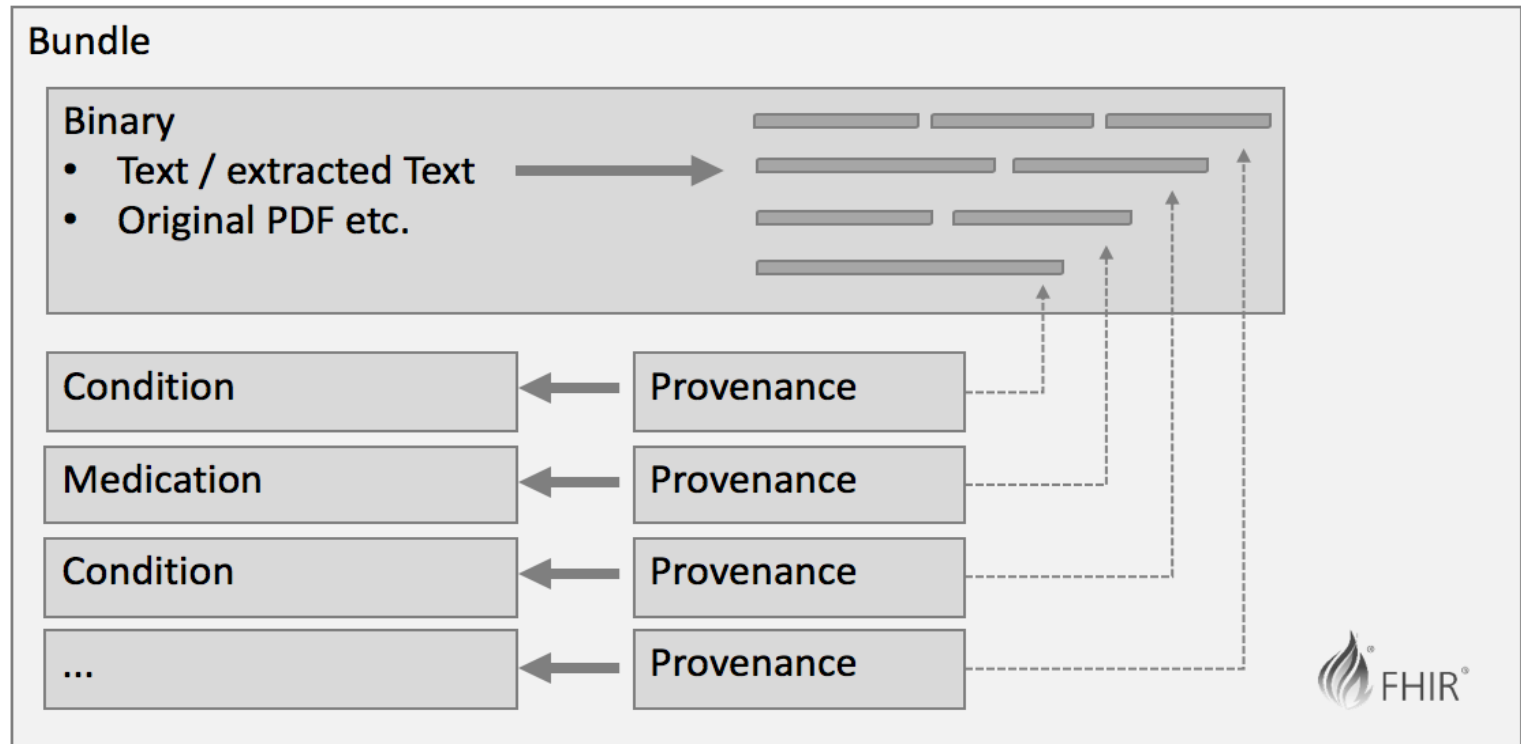
- *Unstructured Information Management Architecture* (UIMA): flexible and extensible text-mining framework
- All elements extracted from texts are predefined by Type systems:
 - Defines attributes and features
 - Common UIMA type: “Annotation: Attributes “begin” ... “end” specify a sequence of characters in a text to which the annotation relates
- Currently no UIMA type system specific to healthcare
- Bridging the gap between UIMA type system and FHIR:
 - Defining a FHIR-compatible type system (as output of the NLP system)
 - **Exposing a FHIR compatible interface**

FHIR-COMPATIBLE INTERFACE FOR TEXT MINING RESULTS

- Which are the right FHIR resources and containers to return a text mining result?
- How can document text annotations referring to clinical conditions and observations be linked together?
- How can two important text mining features, be specified in FHIR
 - (i) text span of an annotation
 - (ii) its confidence score
- How can the types of the **HEALTH discovery** text mining system be mapped to FHIR resources and which challenges occur?

RESULTS (I): FHIR RESOURCES AND CONTAINER FOR TEXT MINING RESULTS

- Text mining system processes documents (e.g. PDF) and returns extracted text with annotations
- Appropriate FHIR structure:
 - *Bundle* of type *collection*, grouping
 - Original document (Binary)
 - FHIR resource(s), filled by information extracted from the original document
- Provenance resource
 - Specifying data source, extraction agent (here NLP system) and target resource



RESULTS (II): FHIR EXTENSIONS CHARACTER RANGE AND CONFIDENCE

character-range specifies the place in a source text to which an annotation belongs

```
"extension": [{"url": "http://example.com/StructureDefinition/character-range/0.9",  
  "valueRange": [{"low": {"value": 201, "unit": 1},  
    "high": {"value": 219, "unit": 1}}]}
```

confidence specialises the type *Quantity*, stating the accuracy of annotations

```
"extension": [{"url": "http://example.com/StructureDefinition/confidence/0.9",  
  "valueQuantity": {"value": 0..1}}]}
```

RESULTS (III): MAPPING THE TEXT MINING OUTPUT TO FHIR RESOURCES

Averbis Health Discovery	FHIR Resource		
Diagnosis	Condition		
.concept*	.code*		
.clinicalStatus	.clinicalStatus		
.verificationStatus	.verificationStatus		
.belongsTo	.subject.display		
.side	.bodySite.text		
Medication	Medication		
.drug.ingredientConcept*	.ingredient*		
.doseForm*	.form*		
	MedicationStatement		
.status	.status		
.drug.strength.value	.dosage.doseQuantity.value		
.drug.strength.unit	.dosage.doseQuantity.unit		
.doseFrequency**	.dosage.timing**		
.administrations*	.dosage.method*		
LaboratoryValue	Observation		
.parameter*	.code*		
.fact.value	.valueQuantity.value		
.fact.unit	.valueQuantity.unit		
.interpretation	.interpretation.text		
.upperLimit.normalizedValue	.referenceRange.high.value		
.upperLimit.normalizedUnit	.referenceRange.high.unit		
.lowerLimit.normalizedValue	.referenceRange.low.value		
.lowerLimit.normalizedUnit	.referenceRange.low.unit	(*)	(codable) concept
		(**)	not yet further specified

EXAMPLE „THE PATIENT HAS A COLD“

```
{
  "resourceType": "Bundle",
  "id": "628320",
  "meta": {
    "versionId": "1",
    "lastUpdated": "2018-11-24T15:00:45.182+00:00"
  },
  "type": "collection",
  "entry": [
    {
      "fullUrl": "urn:uuid:f844ec9a-ef45-11e3-8bb6-00aa004d0001",
      "resource": {
        "resourceType": "Binary",
        "contentType": "text/plain",
        "content": "UGF0aWVudCBoYWQgYSBjb2xk"
      }
    },
    {
      "fullUrl": "urn:uuid:f844ec9a-ef45-11e8-8bb6-00aa004d0001",
      "resource": {
        "resourceType": "Device",
        "manufacturer": "Averbis GmbH",
        "model": "Health Discovery",
        "version": "5.6.0"
      }
    },
    {
      "fullUrl": "urn:uuid:f844ec9d-ef45-11e8-96f5-00aa004d0001",
      "resource": {
        "resourceType": "Condition",
        "clinicalStatus": "active",
        "verificationStatus": "unknown",
        "code": {
          "coding": [
            {
              "system": "http://hl7.org/fhir/sid/icd-10-cm",
              "version": "2018",
              "code": "J00",
              "display": "Acute nasopharyngitis [common cold]"
            }
          ]
        },
        "subject": {
          "display": "anonymous patient"
        }
      }
    },
    {
      "fullUrl": "urn:uuid:f844ec9d-ef45-11e8-96f3-00aa004d0001",
      "resource": {
        "resourceType": "Provenance",
        "target": [
          {
            "reference": "urn:uuid:f844ec9d-ef45-11e8-96f5-00aa004d0001"
          }
        ],
        "recorded": "2018-11-23T18:34:03.184859+01:00",
        "agent": [
          {
            "whoReference": {
              "reference": "urn:uuid:f844ec9a-ef45-11e8-8bb6-00aa004d0001"
            }
          }
        ],
        "entity": [
          {
            "extension": [
              {
                "url": "http://example.com/StructureDefinition/confidence/0.9",
                "valueQuantity": {
                  "value": 0.8
                }
              },
              {
                "url": "http://example.com/StructureDefinition/character-range/0.5",
                "valueRange": {
                  "low": {
                    "value": 15,
                    "unit": "1"
                  },
                  "high": {
                    "value": 19,
                    "unit": "1"
                  }
                }
              }
            ],
            "role": "source",
            "whatReference": {
              "reference": "urn:uuid:f844ec9a-ef45-11e3-8bb6-00aa004d0001"
            }
          }
        ]
      }
    }
  ]
}
```

CONCLUSION

- Proof of concept of alignment between UIMA based NLP type system and FHIR resources
- Modelling solution for additional NLP-relevant information as a FHIR bundle, addressing data provenance and data quality
- Necessary extensions to the FHIR standard currently being processed by the international HL7/FHIR committee

CONTACT

Philipp Daumke

Averbis GmbH

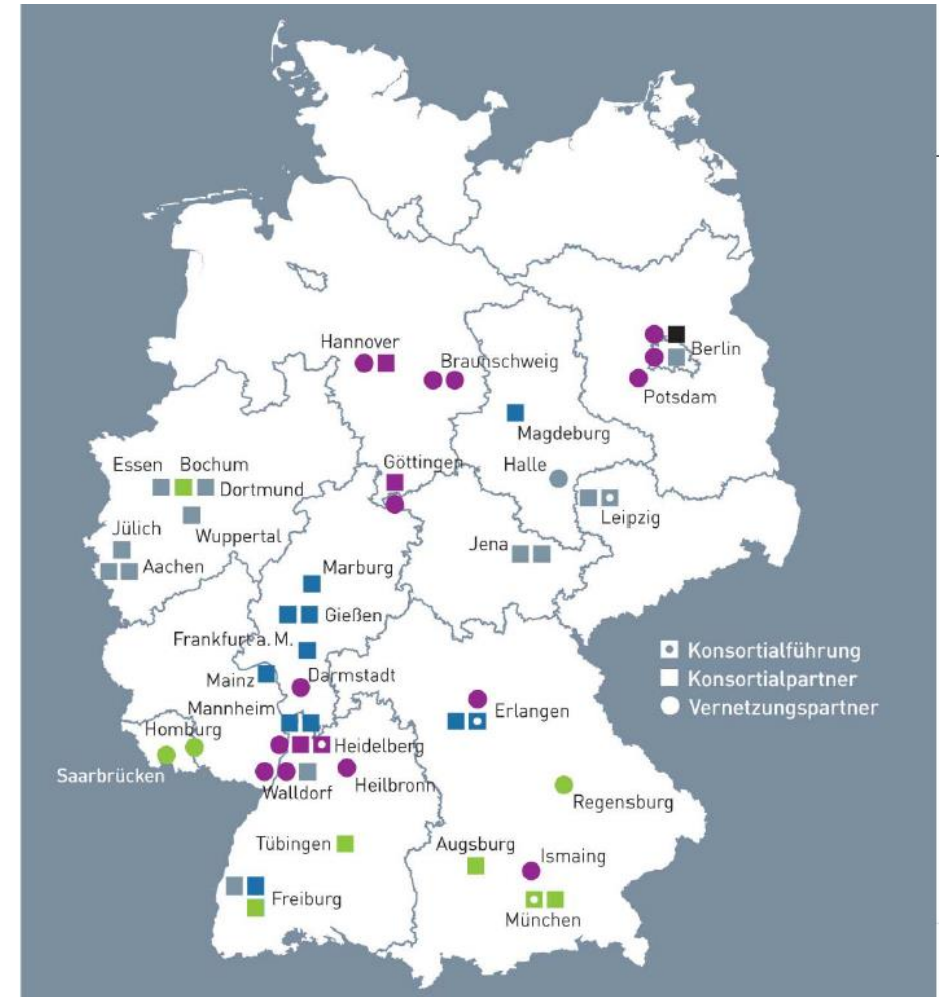
Freiburg, Germany

philipp.daumke@averbis.com

ACKNOWLEDGEMENTS

Averbis is the text mining partner in currently three consortia of the German „Funding Scheme Medical Informatics“

(Federal Ministry of Education and Research (BMBF) within the Medical Informatics Funding Scheme (FKZ 1ZZ1606A and 01ZZ1609A).



miracum



Smart Medical Information
Technology for Healthcare

DIFUTURE

Data Integration for Future Medicine