

# Automatische Extraktion von Medikamenteninformationen aus Arztbriefen

**Stefan Schulz, Thorsten Seddig**

**Universitätsklinikum Freiburg, Institut für Medizinische Biometrie und Medizinische Informatik  
Averbis GmbH, Freiburg**



# Strukturierte Dokumentation von Arzneiverordnungen

- Gegenwärtige strukturierte Daten in der stat. Routinedokumentation in Deutschland:
  - Diagnosen (ICD)
  - Prozeduren (OPS)
  - Häufig: Labordaten
  - Selten: Medikationsdaten !
- Gründe:
  - Keine Abrechnungsrelevanz
  - Andere Wahrnehmung von Medikamentenrisiken als z.B. In USA
  - Unterschiedliche Arbeitsabläufe
  - Kein verbindliches Kodierungssystem für Arzneimittel in D

# Strukturierte Dokumentation von Arzneiverordnungen

---

- Vorteile:
  - Entscheidungsunterstützung
  - Erkennung von Medikamenteninteraktionen und Kontraindikationen
  - Steuerung des Verschreibungsverhaltens  
(unter medizinischen und krankenhaushökonomischen Aspekten)
- Abrechnungsrelevanz?

# Hypothese

---

- ICD - Kodierung ist teilweise unvollständig (Nebendiagnosen !)
- In Folge: DRG – Gruppierung suboptimal: vermeidbare Einnahmeverluste
- Medikation gibt Hinweis auf unkodierte Nebendiagnosen
- Neue Diagnosehypothesen durch Interpretation der Medikationsdaten
- Komplettierung der Diagnosendokumentation nach Bewertung der Diagnosehypothesen
- Softwareunterstützung optimiert Leistung von Kodierern / Medizincontrollern

# Anwendungsbeispiel

## KIS

### Diagnosen:

K43.9 Narbenhernie

G81.0 Hemiparese re

### Prozedur:

5-536.44 Verschluss einer  
Narbenhernie

### Medikation:

Madopar 62,5mg 1-0-1

## Abrechnung 1

G08B Komplexe Rekonstruktion  
der Bauchwand

Basisbetrag: 3.465,50 €

# Anwendungsbeispiel

## KIS

### Diagnosen:

K43.9 Narbenhernie

G81.0 Hemiparese re

### Prozedur:

5-536.44 Verschluss einer  
Narbenhernie

### Medikation:

**Madopar 62,5mg 1-0-1**

## Abrechnung 1

G08B Komplexe Rekonstruktion  
der Bauchwand

Basisbetrag: 3.465,50 €

# Anwendungsbeispiel

## KIS

### Diagnosen:

- K43.9 Narbenhernie  
G81.0 Hemiparese re  
G20.21 M. Parkinson mit  
schwerster Beeinträchtigung

### Prozedur:

- 5-536.44 Verschluss einer  
Narbenhernie

### Medikation:

**Madopar 62,5mg 1-0-1**

## Abrechnung 1

G08B Komplexe Rekonstruktion  
der Bauchwand

Basisbetrag: 3.465,50 €

# Anwendungsbeispiel

## KIS

### Diagnosen:

- K43.9 Narbenhernie  
G81.0 Hemiparese re  
G20.21 M. Parkinson mit  
schwerster Beeinträchtigung

### Prozedur:

- 5-536.44 Verschluss einer  
Narbenhernie

### Medikation:

**Madopar 62,5mg 1-0-1**

## Abrechnung 1

G08B Komplexe Rekonstruktion  
der Bauchwand

Basisbetrag: 3.465,50 €

## Abrechnung 2

G08A Komplexe Rekonstruktion der  
Bauchwand mit äußerst  
schweren Komplikationen

Basisbetrag: 5.037,30 €

# Automatische Hypothesengenerierung

- Medikament: Abbildung auf Wirkstoff(e)
- Diagnosen / Wirkstoff - Matrix

	$W_1$	$W_2$	$W_3$	$W_4$	$W_5$		$W_n$
$D_1$		X					
$D_2$							X
$D_3$	X				X		
$D_4$				X			
$D_5$		X					
$D_6$			X				
$D_7$				X			
$D_n$	X						

# Pilotstudie

- Manuelle Erstellung eines einfachen Medikamententhesaurus
  - Parlodel (Medikamentname) → Bromocriptin (Wirkstoff)
  - Umprel (Medikamentname) → Bromocriptin (Wirkstoff)
  - Mexitil (Medikamentname) → Mexiletin (Wirkstoff)
  - Madopar (Medikamentname) → Levodopa (Wirkstoff)
  - Madopar (Medikamentname) → Benserazid (Wirkstoff)
- Manuelle Erstellung der Diagnosen – Wirkstoff – Matrix (MMI-Daten):
  - Diagnosengruppen: M. Parkinson, Arrhythmie, Epilepsie
  - Wirkstoffe: typische Medikamente:
- Beispiel:
  - Pramipexol → G20, G21, G22
  - Levetiracetam → G40, G41
  - Mexiletin → I44 - I49

# Extraktion von fallbezogenen Arzneimittelinformationen

- Quelle: Arztbrief
- Methode: Text Mining:
  - Identifikation von relevanten Dokumentenabschnitten
  - Erkennung von Medikamentennamen (Arzneimittel und / oder Wirkstoffe) durch Stringvergleich
- Vorteile:
  - Medikamenten- oder Wirkstoffnamen sind zumeist hoch spezifisch
  - Zumindest Namen der Medikamente bei Entlassung im Dokument gut zu identifizieren
- Nachteile:
  - Schreibfehler, Schreibvarianten
  - Formatierungsartefakte
  - Mehrdeutige Akronyme

# Identifikation der Textabschnitte

Diagnosen:

1. Colitis ulcerosa
2. Refluxoesophagitis I°
3. Chronisch obstruktive Atemwegserkrankung
4. Adipositas
5. Makro-CK

...

Die **Vorgeschichte** ist Ihnen ja gut bekannt, wir hatten seiner Zeit nach längerem Überlegen einen **Therapieversuch** mit **Budenofalk- Rektalschaum** begonnen. Am 20.04.03 berichtete mir Herr D., dass die Situation nicht zu konsolidieren begann. Wir sind so verblieben, daß er **Decortin** auf 15 mg reduzieren kann, weiterhin 3x 1 g **Pentasa**. **Metatrexat** war **abgesetzt**.

- **Durch Schlüsselwörter wie:**
  - Anamnese
  - Vorgeschichte
  - Therapieversuch

# Ergebnisse der Pilotstudie an Echtdaten des UKL Freiburg

- Details siehe Vortrag von Thorsten Seddig im GMDS-Hauptprogramm  
*„Generierung von Diagnosehypothesen für die DRG-Kodierung durch semantische Analyse der Medikation in Arztbriefen“*
- Stichprobentests:
  - Korrekte Erkennung der Medikation: [66 % – 100 %]
  - Unterkodierte Fälle / Selektierte Fälle: [17 – 25 %]
- Geringe DRG-Relevanz (Korrektur der unterkodierten Fälle bringt selten bessere Einstufung)

# Ausblick

---

- Erhöhung der Precision: automatische Erkennung von Formatierungsartefakten, fehlertolerante Suche
- Erhöhung des Recalls: Nichtberücksichtigung von konkurrierenden Diagnosen (z.B. Restless-Leg-Syndrom als häufige Indikation für Parkinsonmedikamente)
- Erweiterung der Wissensbasis (Füllen der Matrix)
  - Regelbasiert, z.B. durch Verwendung von MMI – Datenbasis
  - Probabilistisch, Füllen der durch (unsupervised) Machine Learning

# Wirkstoff / Diagnosenmatrix mit prädiktiven Werten

	$W_1$	$W_2$	$W_3$	$W_4$	$W_5$		$W_n$
$D_1$		0,9					
$D_2$							0,3
$D_3$	0,2				0,4		
$D_4$				0,5			
$D_5$		0,15					
$D_6$			0,05				
$D_7$				0,4			
$D_n$	0,1						

- Vergleich: automatisch „gelernte“ Wissensbasis gegenüber manuell erstellter Wissensbasis (z.B. MMI) bzgl. Prädiktion von Diagnosen

# Fazit

---

- Sofern keine strukturierte Erfassung von Medikationsdaten, empfehlen sich Text Mining – Verfahren zur Extraktion der Medikamentennamen (ohne Dosierung und Darreichungsform), z.B. zur Prädiktion oder Qualitätssicherung von Diagnosencodes
- Erforderlich:
  - Thesaurus zur Abbildung von Medikamentennamen auf Wirkstoff-IDs
  - Wissensbasis: Wirkstoff – ICD – Matrix
  - Wissenschaftliche Fragestellung: manuell generierte vs. automatisch gelernte (aus PDV-Daten) Wissensbasis

## **Kontakt:**

**Stefan Schulz:** [stschulz@uni-freiburg.de](mailto:stschulz@uni-freiburg.de)

**Thorsten Seddig :** [seddig@averbis.de](mailto:seddig@averbis.de)

**Universitätsklinikum Freiburg, Institut für Medizinische Biometrie und Medizinische Informatik  
Averbis GmbH, Freiburg**

