

Informationsextraktion aus medizinischen Texten

K. Denecke

51. GMDS Jahrestagung Leipzig 2006

Inhalt

- Motivation
- Einsatzbereiche
- Erzeugung semantischer Strukturen
- Gezielte Extraktion von Schlüsselinformationen
- Evaluierung
- Zusammenfassung

Motivation

Studie der Uni-Klinik Freiburg:

Ungefähr 8,5% ihrer Arbeit
verbringen Krankenhaus-Mitarbeiter
täglich mit der Suche nach Daten

Daten sind zu 30% nicht dort, wo sie
benötigt werden. Sie sind nicht
auffindbar, nicht zugreifbar.

Ca. 5% aller ärztlichen Verrichtungen
erfolgen ohne Kenntnis der für die
Verrichtung notwendigen Informationen.

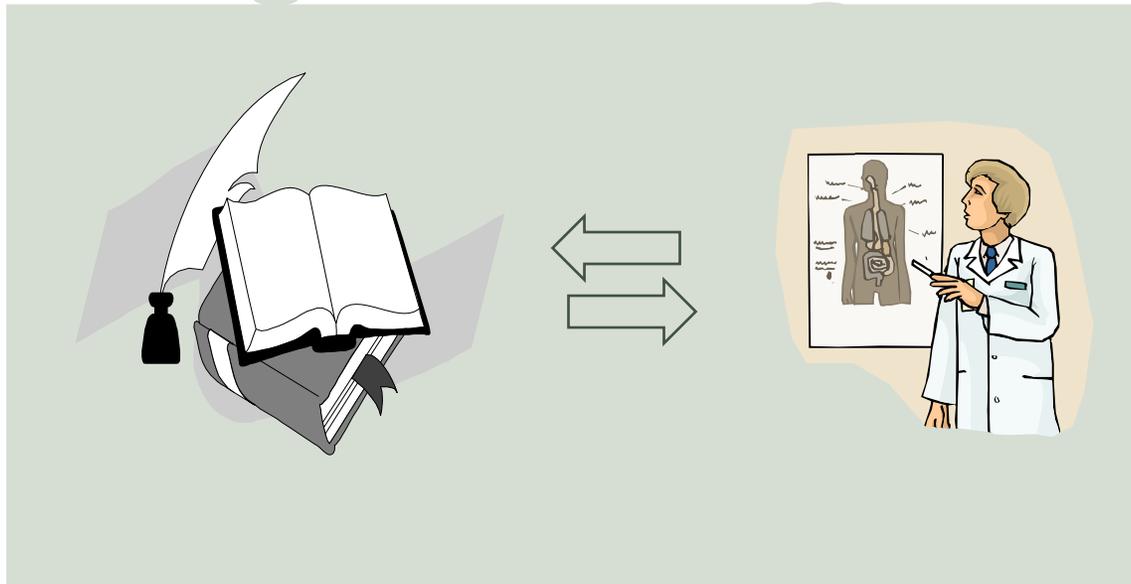
Medizinische Dokumente sind z.T.
nicht lesbar, unvollständig, ungeordnet
oder zu groß.

(Quelle: Prof. R. Trill, FH Flensburg, Hauptstadtkongress 2006)

Motivation

Wiederverwendung
von erfassten
Daten

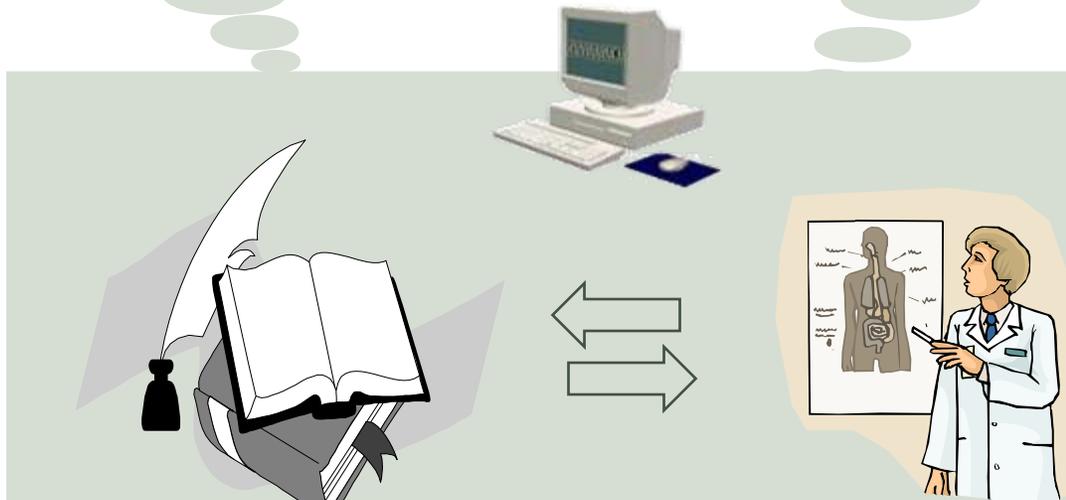
Leichter, effizienter
Zugang
zu Informationen



Motivation

Wiederverwendung
von erfassten
Daten

Leichter, effizienter
Zugang zu Informationen



ICD-10 Codes	Causes of Death	Number of Deaths			
		< 1 year	1-4	5-14	15-64
	All causes	2495	109	703	416
		1895	138	448	279
A00-B99	Infectious and parasitic diseases	95	0	1	9
A00-A09	Intestinal infectious diseases	93	0	2	9
A01	Whooping cough	1	0	0	0
E40-E44	Nutritional deficiencies	1	0	0	0
G00-G99	Diseases of the nervous system	0	0	8	0
G00-G09	Meningitis	37	0	2	18
J00-J99	Diseases of the respiratory system	734	0	3	66
J01-J18	Pneumonia	609	1	3	51
J10-J11	Influenza	9	0	0	0
Q00-Q99	Congenital anomalies	454	38	133	100
Q01-Q05	Spina bifida and hydrocephalus	375	35	94	61
Q20-Q29	Congenital anomalies of heart and circulatory system	222	8	47	17
P00-P99	Certain conditions originating in the perinatal period	879	127	559	187
P10-P15	Birth trauma	34	4	25	5

Generierung
Interpretation
Verarbeitung

Informationsextraktion
Information Retrieval



Einsatzbereiche: Intersektorale Kommunikation

Befund:
Der Tc- Uptake ist mit 2,09% grenzwertig hoch.
Schilddrüse in orthotoper Lage.
Relativ homogene Speicherung des linken Schilddrüsenlappens. Umschriebene Aktivitätsmehranreicherung in der Isthmusregion sowie maximale Speicherung im rechten Schilddrüsenlappen zentral.

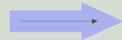
Sonographie:
Das Volumen des rechten Schilddrüsenlappens beträgt 9 ml, linksseitig 10 ml.
Im rechten Schilddrüsenlappen er
Volumen. In der Isthmusregion
Der linke Schilddrüsenlapp
10 mm.

Zusammenfassung:
Bei der Patientin liegt e
Die aktuellen Laborpar

Empfehlung:
Durchführung einer R
Im Vorfeld ist ein diagn
Bis dahin ist eine schil
Eine entsprechende ar

Satz (die aktuellen Laborparameter belegen eine latente hyperthyreose ohne schilddrüsenspez.)
(Diagnose) latente Hyperthyreose [D0015B1]
(Verfahren) Laborverfahren (V0005E1)

- Extraktion und Kodierung von Diagnosen, Prozeduren und Medikationen aus medizinischen Freitexten



Diagnosen,
Prozeduren,
Medikationen



ICD-10,
OPS,
ATC



Intersektorale Kommunikation

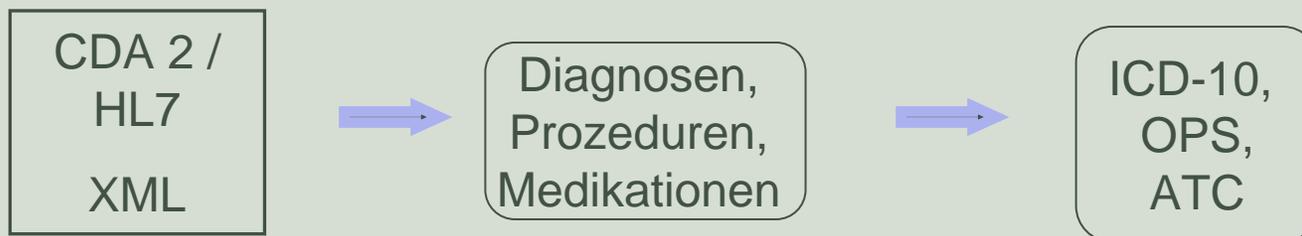
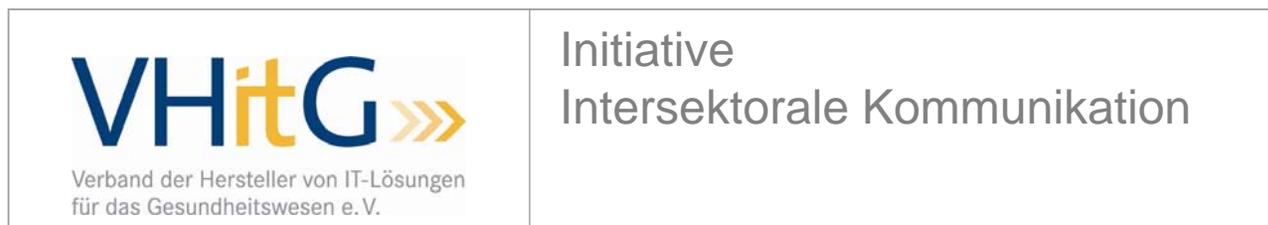
HL 7 CDA XML

Informationsextraktion aus med. Texten | GMDS Jahrestagung 10.-14. 09. 2006 | Folie 5

IV i003.0 hyperthyreose latent;
Ok

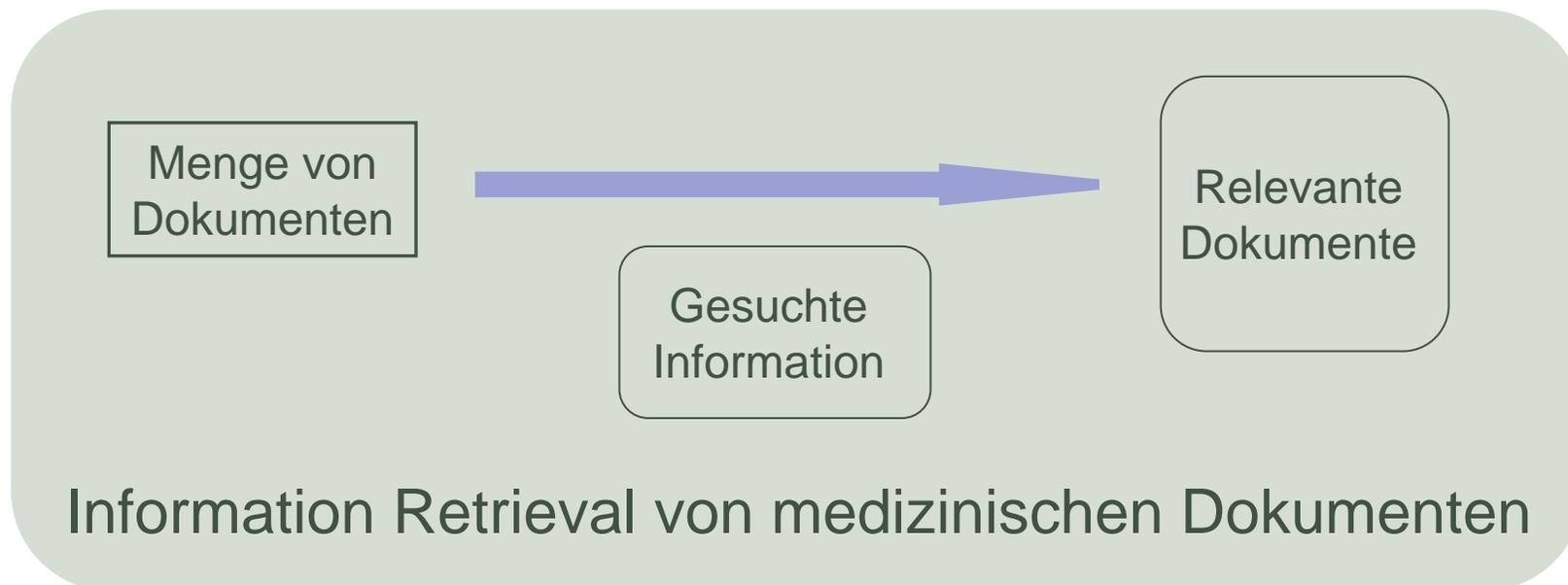
Codierung

Einsatzbereiche: VHitG-Arztbrief



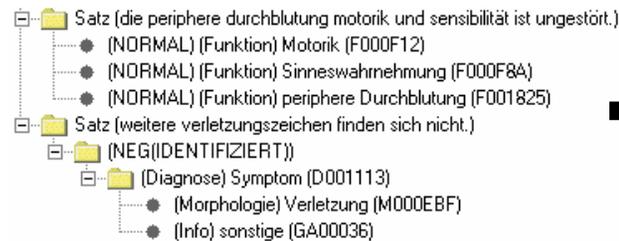
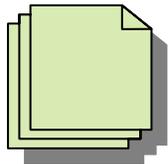
- Verarbeitung standardisierter Dokumente
- Konsistenzprüfungen

Einsatzbereiche: Intelligentes Information Retrieval



- Archivierung von Dokumenten
- Qualitative und quantitative Analyse

Material und Methoden



Informationen über die Aufnahme des Patienten:	
Aufnahmeart:	Erstaufnahme
Aufnahmearlass:	normal
<u>Aufnahmegrund</u>	
Diagnosebezeichnung:	<u>Skrotalhernie</u>
Topographie:	-
Seitenlokalisation:	rechts
<u>Diag.sicherheit:</u>	-
<u>Diag.typ:</u>	Aufnahmediagnose

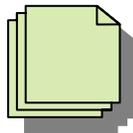
Informationen zur Entlassung:	
Entlassungszustand:	
Allgemeinzustand:	gut

Freitext

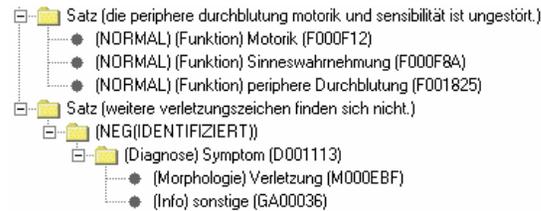
Semantische Struktur

gefüllter Frame

Material und Methoden



Freitext



Semantische Struktur



Informationen über die Aufnahme des Patienten:	
Aufnahmeart:	Erstaufnahme
Aufnahmeanlass:	normal
<u>Aufnahmegrund</u>	
Diagnosebezeichnung:	Skrotalhernie
Topographie:	-
Seitenlokalisation:	rechts
Diag.sicherheit:	-
Diag.typ:	Aufnahmediagnose
Informationen zur Entlassung:	
Entlassungszustand:	gut
Allgemeinzustand:	gut

gefüllter Frame

Achsen der Wingert-Nomenklatur

Diagnose (D)

Morphologie (M)

Funktion (F)

Prozedur (P)

Verfahren (V)

Topologie (T)

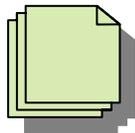
Ätiologie (E)

Stoffe (W)

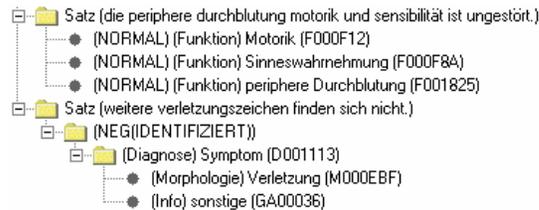
Job (J)

Info (G)

Material und Methoden



Freitext



Semantische Struktur

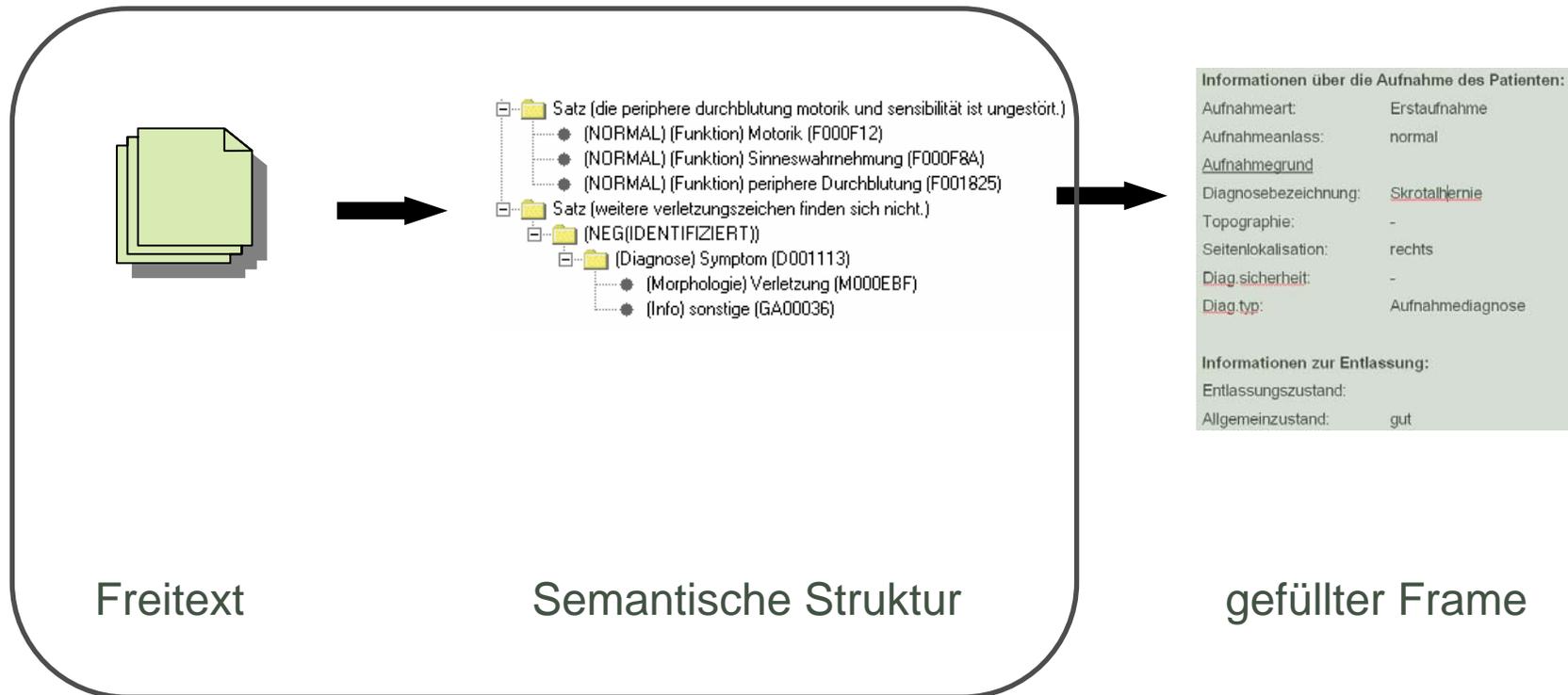


Informationen über die Aufnahme des Patienten:	
Aufnahmeart:	Erstaufnahme
Aufnahmehinweis:	normal
<u>Aufnahmegrund</u>	
Diagnosebezeichnung:	Skrotalhernie
Topographie:	-
Seitenlokalisation:	rechts
Diag.sicherheit:	-
Diag.typ:	Aufnahmediagnose
Informationen zur Entlassung:	
Entlassungszustand:	-
Allgemeinzustand:	gut

gefüllter Frame

- Wingert-Nomenklatur
- Methoden zur automatischen **Abbildung natürlicher Sprache** auf Konzepte der Nomenklatur
- Methoden zur **Erzeugung** einer semantischen Repräsentation
- Methoden zur eigentlichen **Extraktion**

Material und Methoden



Informationen über die Aufnahme des Patienten:	
Aufnahmeart:	Erstaufnahme
Aufhuan Anlass:	normal
<u>Aufnahmegrund</u> :	
Diagnosebezeichnung:	<u>Skrotalhernie</u>
Topographie:	-
Seitenlokalisation:	rechts
<u>Diag.sicherheit</u> :	-
<u>Diag.typ</u> :	Aufnahmediagnose
Informationen zur Entlassung:	
Entlassungszustand:	
Allgemeinzustand:	gut

Erzeugung semantischer Strukturen

Input:

Diagnose(n):

S42.21 Fraktur des proximalen Endes des Humerus,
Humeruskopf rechts
I10.00 Benigne essentielle Hypertonie
I25.0 Atherosklerotische Herz-Kreislauf-Krankheit

Therapie(n):

Mobilisierung nach Pöhlchen, Röntgen-Kontrolle, Analgesie

Anamnese:

Vorstellung in unserer Notaufnahme, nachdem die Patientin auf der Straße ausgerutscht und auf die rechte Schulter gefallen sei.

Aufnahmebefund:

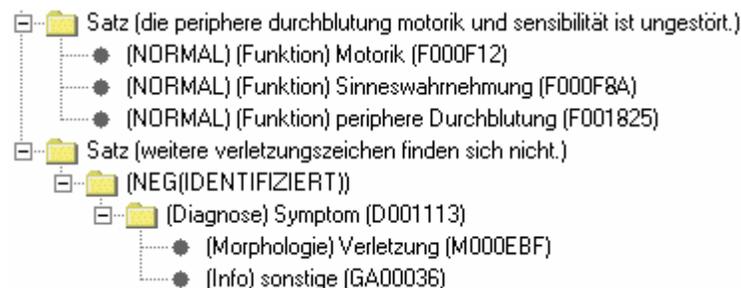
Rechte Schulter: Es besteht eine regelrechte Gelenkstellung. Neurologische Ausfälle liegen nicht vor. Eine Hautverletzung besteht nicht. Die Schulter ist druckdolent, eine Krepitation ist nicht nachweisbar. Motorik und Durchblutung sind peripher erhalten.

re.HG: Es findet sich keine Schwellung am rechten Handgelenk. leichte Druckdolenz über dem distalen Radius. Es besteht kein Tabatièrendruckschmerz. Keine Hautverletzung. Die periphere Durchblutung, Motorik und Sensibilität ist ungestört. Weitere Verletzungszeichen finden sich nicht.

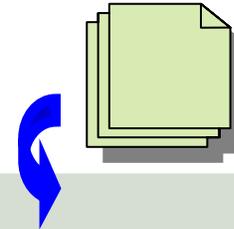
Verlauf:

Nach initialer Ruhigstellung im Gilchrist-Verband und Medikation mit Tramal und Novaminsulfon ließen die Schmerzen rasch nach, so daß wir mit einer krankengymnastischen Beübung nach Pöhlchen beginnen konnten. Diese wurde gut toleriert. Motorik, Sensibilität und Durchblutung waren jederzeit intakt. Die Röntgenkontrolle nach Belastung zeigte eine unverändert gering dislozierte Fraktur, so daß weiterhin keine OP-Indikation gegeben ist. Eine häusliche Versorgung ist aufgrund des Alters und der leichten Verwirrtheit zur Zeit noch nicht möglich.

Output:



Verarbeitungsmodulare



Vorverarbeitung

Parse Sätze
Bestimme Segmente
Identifiziere spezielle Ausdrücke
Ermittle Wortklassen
Indexiere

*[Reposition der Sonde] [init
Reflux der Sondenkost]*

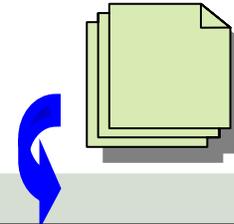
*{V00000B Reposition
W000008 Sonde}
{F00001F Reflux
E002656 Kost}*

Erzeugen semantischer Strukturen

Bestimme Hauptinformation
Mappe Indizes auf Konzepte
Erzeuge Relationen

*Reposition (V00000B)
|- Sonde (W000008)
|- (begleitet von) Reflux (F00001F)
|- Kost (E002656)*

Verarbeitungsmodulare



Vorverarbeitung

Parse S
Bestimm
Identifiz
Ermittle
Indexier

Erzeuge

Bestimm
Mappe
Erzeug

[Reposition der Sonde] [mit Reflux der Sondenkost]

[Reposition (V) der Sonde (W)] [mit Reflux (F) der
Sondenkost (E)]

[V W] [F E]

Rangfolge: D M F V P T W E J G

Rangfolge: D M F **V** P T **W** E J G

Rangfolge: D M **F** V P T W **E** J G

Beispiel einfach

leichte Hepatomegalie



{M0018BD, GA00B23}



Hepatomegalie (M0018BD)

|- (Schweregrad) leicht (GA00B23)

Rangordnung:
D **M** F V P T E J W **G**

Beispiel Negation

Kein Parenchymtumor im Bereich der Nieren

T0012D1, T002150, M000B65

Rangordnung:
D M F V P T E J W G

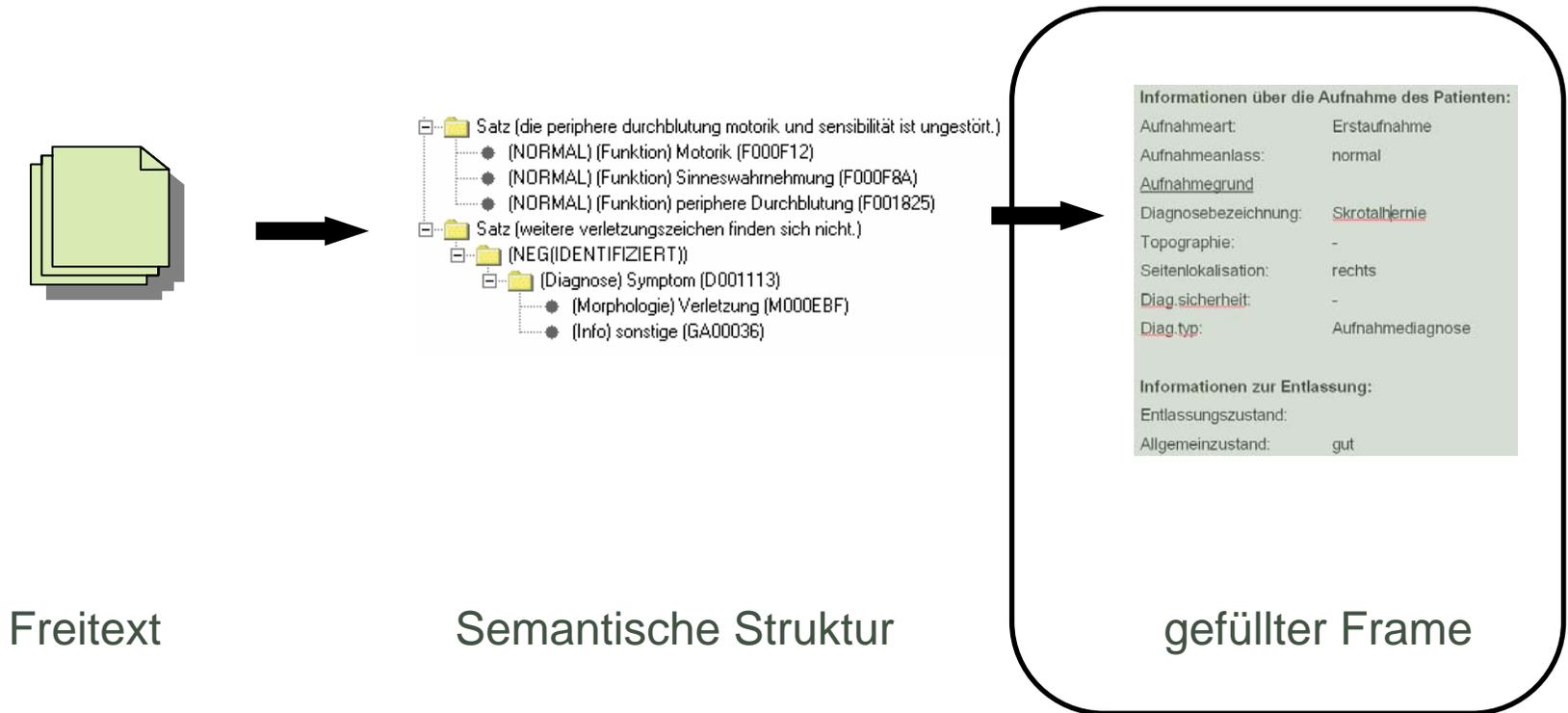
(NEG)

|- Neoplasma (M000B65)

|- (Topografie) Parenchym (T002150)

|- (im Bereich) Niere (T0012D1)

Verarbeitungsprozess



Gezielte Extraktion von Informationen

Diagnose: Skrotalhernie rechts K40.90

Therapie: laparoskopische Hernioplastik am 08.09.05 5-531.31

Die stationäre Aufnahme des Patienten erfolgte aufgrund einer großen, langjährig bestehenden, rechtsseitigen Skrotalhernie. Trotz des ausgedehnten Befundes entschied sich der Patient für ein laparoskopisches Verfahren.

Der Eingriff wurde am 08.09 durchgeführt. Nach Einführen der Optik fand sich eine große indirekte Hernie. Der Bruch enthielt vor allem ein großes Lipom. Nach entsprechender Präparation wurde ein titanisiertes Netz an der Bauchwand fixiert. Die Operation wurde vom Patienten gut toleriert.

Der postoperative Verlauf gestaltete sich komplikationslos. Die Wundheilung erfolgte primär. Während des stationären Aufenthalts erhielt der Patient eine Thromboembolieprophylaxe.

Wir entließen den Patienten in gutem Allgemeinzustand in Ihre Weiterbehandlung und bitten um eine Entfernung des Nahtmaterials am 15.09. Wir bedanken uns für die Einweisung.

Entlassungsmedikation:

Bisoprolol 5 1/2 - 0 - 1/2

Gezielte Extraktion von Informationen

Die stationäre Aufnahme des Patienten erfolgte aufgrund einer großen, langjährig bestehenden, rechtsseitigen Skrotalhernie. Trotz des ausgedehnten Befundes entschied sich der Patient für ein laparoskopisches Verfahren.

Informationen über die Aufnahme des Patienten:

Aufnahmeart: Erstaufnahme

Aufnahmearbeit: normal

Aufnahmegrund

Diagnosebezeichnung: Skrotalhernie

Topographie: -

Seitenlokalisation: rechts

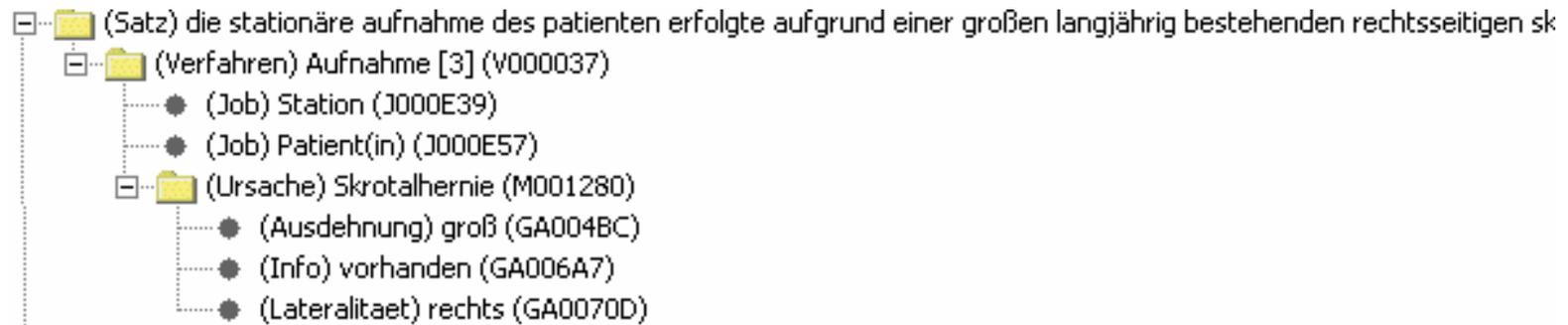
Diag.sicherheit: -

Diag.typ: Aufnahme-diagnose

Gezielte Extraktion von Informationen

Methoden:

Extraktion aus Konzeptgraphen



Aufnahmegrund

Diagnosebezeichnung: Skrotalhernie

Topographie: -

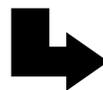
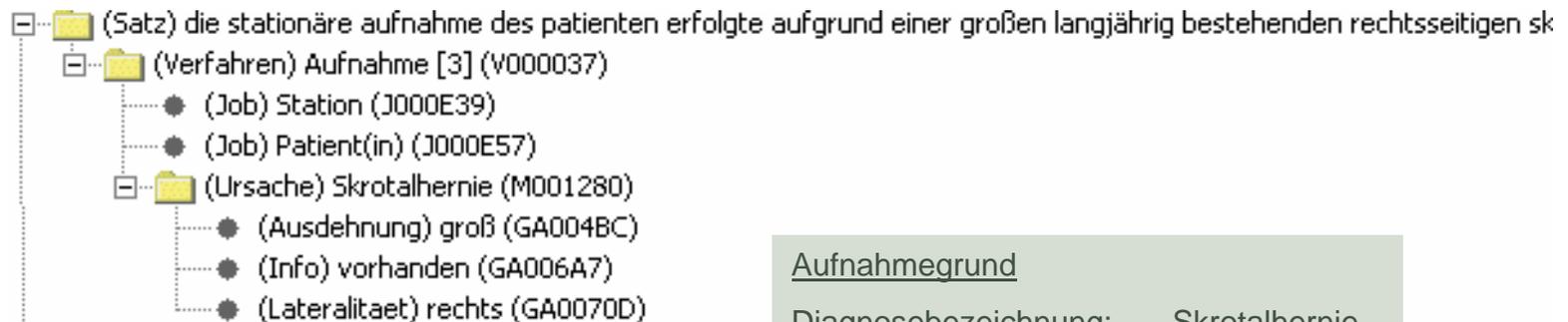
Seitenlokalisierung: rechts

Gezielte Extraktion von Informationen

Methoden:

Extraktion aus Konzeptgraphen

- => Indizes, die bestimmten Achsen angehören (z.B. für Diagnosen die Achsen Funktion, Morphologie, Diagnose)
- => Konzepte mit bestimmten semantischen Rollen (z.B. semant. Rolle = Ursache)
- => ganz bestimmte Indizes (z.B. GA0070D für "rechts")



Aufnahmegrund

Diagnosebezeichnung:	Skrotalhernie
Topographie:	-
Seitenlokalisation:	rechts

Extraktion von Informationen über die Einweisung aus Entlassbriefen

Material: 50 Entlassbriefe der Chirurgie der Oberhavelkliniken Hennigsdorf bei Berlin

Aufnahmeart:	Erstaufnahme
Aufnahmegrund:	Notfall
Aufnahmediagnose:	Appendizitis

- Genauigkeit (Precision): 84%
- Vollständigkeit (Recall): 92%

Extraktion von Informationen über die Medikationen

Auch der postoperative Verlauf gestaltete sich erfreulich positiv, die Operationsdrainagen konnten zeitgerecht entfernt werden, die Wundheilung verlief bis zum Entlassungsdatum problemlos. Der Kostaufbau wurde gut vertragen, so dass wir den Patienten am 16.11.05 aus der stationären Behandlung entlassen konnten. Wir baten um Entfernung des Nahtmaterials am 21.11.05.

Entlassungsmedikation:

Pantozol 40 0-0-1

Bisobeta 0 - 0 - 1/2



Bezeichnung:	Pantozol
Dosierung:	0-0-1
Wirkstoffmenge:	40 mg
Verabreichungszeitpunkt:	bei Entlassung

- Genauigkeit (Precision): 93,3%
- Vollständigkeit (Recall): 87,5%

Zusammenfassung

- **Verwendung existierender Technologien
(Indexierungsalgorithmus, Nomenklatur als Domänenwissen)**
- **Keine Einschränkung auf einen bestimmte medizinischen Bereich**
- **Erzeugen von semantischen Relationen unabhängig von syntaktischen Beziehungen**
- **Übertragbar auf andere Sprachen**
- **Gute Ergebnisse**

Vielen Dank für Ihre Aufmerksamkeit!

Denecke K, Kohlhof I, Bernauer J: *Use of Multiaxial Indexing for Information Extraction From Medical Texts*. In: Reichert A et al.: *Integrating Biomedical Information: From eCell to ePatient*. Proceedings of the European Federation for Medical Informatics. Special Topic Conference. Timisoara. Romania. April 2006

Denecke K, Kohlhof I, Bernauer J: *Information Extraction Based On Multiaxial Indexing And Phrase Structure Analysis*. 20th International Congress of the European Federation for Medical Informatics. Maastricht. August 2006

K.Denecke@id-berlin.de