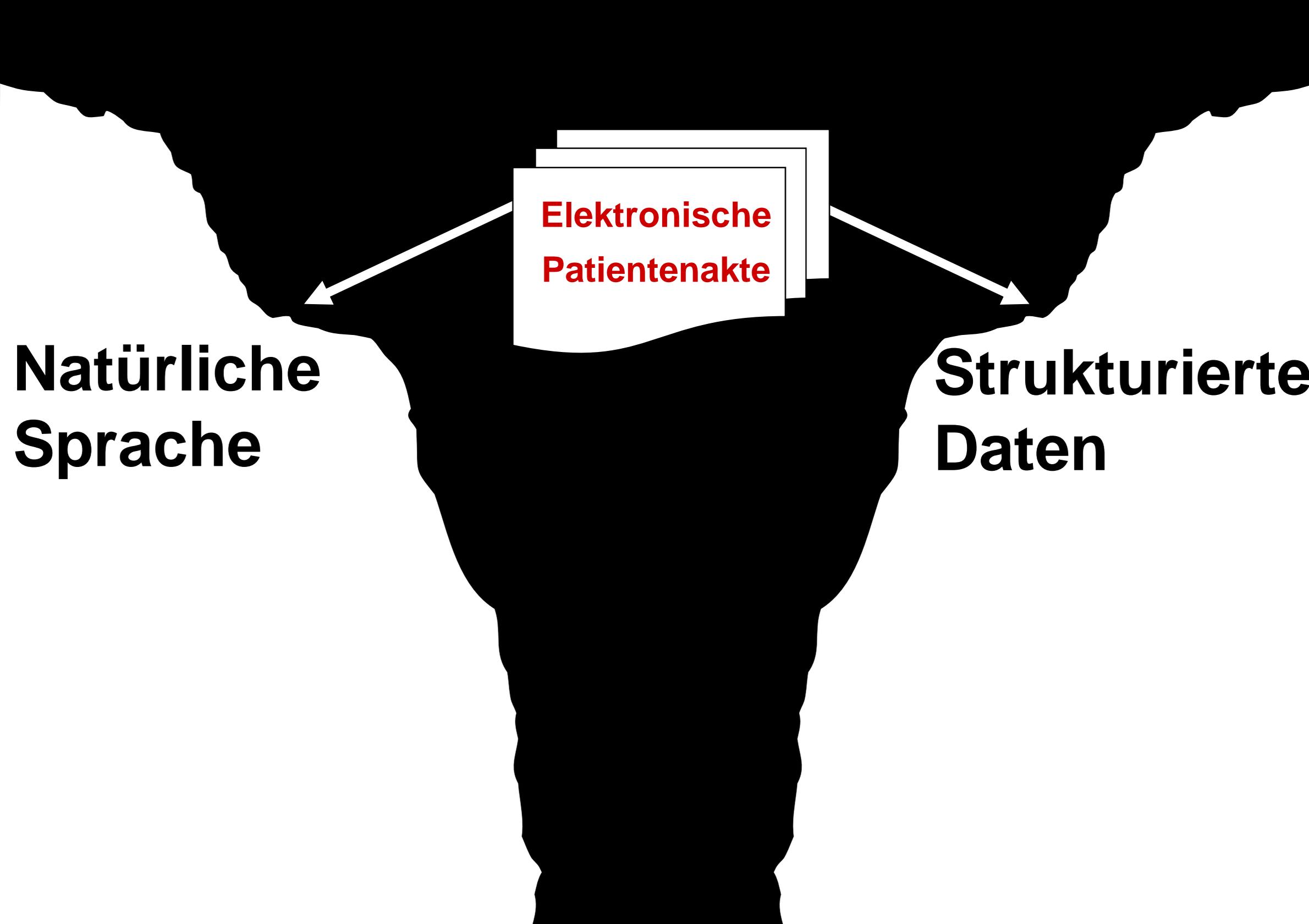




---

# Semantische Aufwertung der elektronischen Patientenakte durch wissensbasiertes Text-Mining

Priv. Doz. Dr. med. Stefan Schulz  
Abteilung Medizinische Informatik



**Elektronische  
Patientenakte**

**Natürliche  
Sprache**

**Strukturierte  
Daten**



**Natürliche  
Sprache**



Datum	Uhrz.	Pflegebericht – Verlaufsbeschreibung Krankenbeobachtung	Hz.
21.4.	13 <sup>30</sup>	Pat. kam mit einer Harnwegsinfektion, hat keine Schmerzen, was schon zur Toilette und hat sich selbstständig gewaschen	SH
22.4.	5 <sup>30</sup>	Pat. hat in d. Nacht keine Harnwegsinfektion geäußert, sie gab aber <del>keine</del> Beschwerden an	SH
22.4.	14 <sup>00</sup>	Pat. lt. Pflegeplan versorgt	SH
	21 <sup>00</sup>	Pat. lt. Plan versorgt; ist nach Betastung schmerzlos	SH
23.4.	8 <sup>00</sup>	Pat. saß fast die ganze Nacht im Bett und konnte nicht im Liegen schlafen	SH
	12 <sup>00</sup>	Pat. hatte 3x Breiig bis dünnflüssigen Stuhlgang (braun)	SH
	14 <sup>00</sup>	Versorgg. lt. Plan	SH
	21 <sup>00</sup>	Pat. lt. Plan versorgt, hält sich nicht an ihre Bettruhe	SH
24.4.	2 <sup>30</sup>	Pat. beim Toilettengang erwacht, hält sich nicht an Bettruhe (sieht herein vom davor)	SH
25.	3 <sup>30</sup>	Pat. wachte aus dem Lichte auf	SH

### **Familienanamnese:**

Vater verstorben an Bronchial-Karzinom, Mutter verstorben an den Folgen einer Pneumonie. Mutter Diabetes mellitus. 5 gesunde Kinder.

### **Systemanamnese:**

Derzeit Appetitlosigkeit, Trockengewicht um 75 Kg, derzeit 80 Kg. Miktio: gelegentlich Harn-verhalt, gehäuft Harnwegsinfekte, derzeit keine Algurie. Vor Dialyse keine Rest-Diurese. Stuhlgang obstipiert, benutzt regelmäßig Abführmittel. Vor NTX starker Juckreiz, Seit NTX deutlich rückläufig. Kein Husten/Auswurf. Noxen: Nichtraucherin, kein Alkohol.

### **Soziale Anamnese:**

Früher Arbeiterin in der Elektronikbranche, dann Hausfrau, verheiratet, lebt mit dem Ehemann zusammen.

Allergien. Keine bekannt.

### **Medikation bei Aufnahme:**

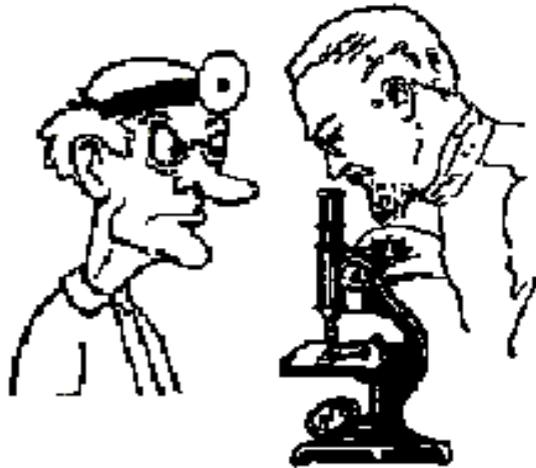
Ulcogant 1-1-1, Pepdul mit 0-0-0-1, Cellcept 2x1 g, Bayotensin 3 x 1, Cynt 0,2 1x1, Ludiomil 50 mg 1 x 1, Sandimmun 2 x 150 mg, Clexane 0,4 ml 1 x täglich s.c.

### **Status bei Übernahme:**

58-jährige Patientin in vorgealtertem, reduziertem Allgemein- und adipösem Ernährungszu-stand (80 Kg Gewicht bei 160 cm Körpergröße). RR 170/80 mm Hg, Puls 66/Minute, regelmäßig. Punktförmige Depigmentierungen an beiden Unterarmen bei Zustand nach heftigem Kratzen wegen Juckreiz. Keine zervikalen Lymphome. Mundschleimhaut trocken, Zunge weißlich belegt. Rachenschleimhaut reizlos, Tonsillen schlecht einsehbar. Schilddrüse nicht vergrößert. Pulmo: Sonorer Klopfeschall und vesikuläres Atemgeräusch. Cor: Spitzenstoß nicht tastbar, leise, reine Herztöne. 3/6. spindelförmiges Systolikum und 1-2/6. Decrescendo-Sofort-dialstolikum über der Aorta mit Fortleitung in die Karotis. Kein abdominales und inguinales Strömungsgeräusch. Abdomen: Bei Adipositas Organgrenzen schlecht beurteilbar, Leber/Milz nicht vergrößert. Reizlose Narbe im Bereich des rechten Unterbauches bei Zustand nach NTX. Dort leichte Druckdolenz.

Wirbelsäule nicht klopfeschmerzhaft. Bds. Unterschenkelödeme. Feinschlägiger Tremor beim Arm-Vorhalte-Versuch. Pupillen isokor, Lichtreaktion prompt. Finger-Nase-Versuch bds. unsicher, ataktisch. Reflexe seitengleich.

# Natürliche Sprache

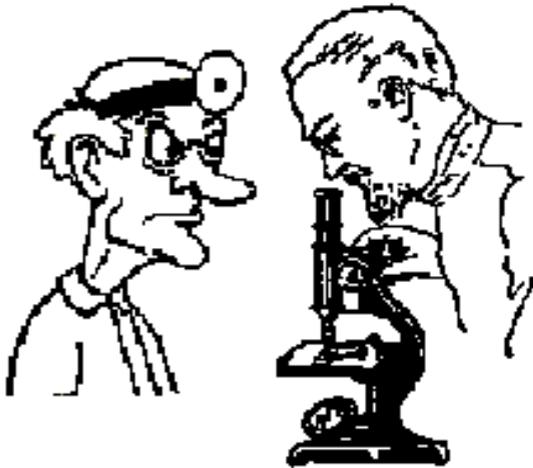


... unverzichtbar für  
Dokumentation und  
Kommunikation

- Krankenversorgung
- Wissenschaft
- Arzt-Patient
- Wissensvermittlung

*nur für das Universitätsklinikum  
Freiburg (p.a.)*

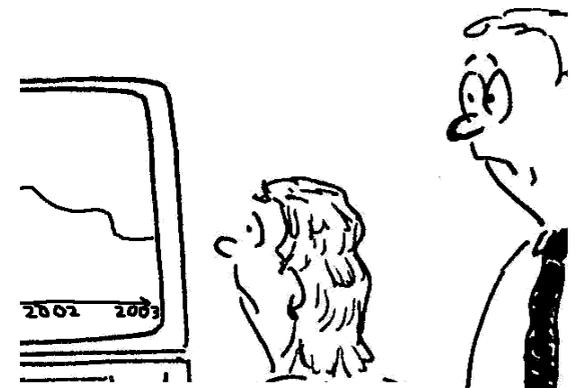
# Natürliche Sprache



280.000	Arztbriefe
140.000	Radiologiebefunde
55.000	Pathologiebefunde
40.000	Operationsberichte
70.000	sonstige Texte

600.000

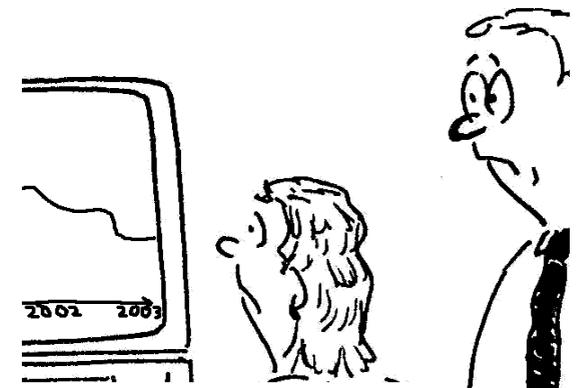
# Strukturierte Daten



# ...unverzichtbar für Klassierung und Abstraktion:

- Kodierung von Diagnosen und Prozeduren, DRGs
- Leistungsabrechnung
- Qualitätssicherung
- Controlling
- Medizinische Forschung
- Gesundheitsberichterstattung (Register)

## Strukturierte Daten



# Epidemiologie (Morbidität, Mortalität)

Causes of Death			Number of Deaths	Number of Deaths at ages (in days)			
ICD 10 Codes	Cause Groupings	Sex	< 1 year	< 1	1-6	7-27	28-364
-	All causes	M	2485	168	703	416	1198
		F	1885	138	448	279	1020
		U	0	0	0	0	0
A00-B99	Infectious and parasitic diseases	M	95	0	1	9	85
		F	93	0	2	9	82
A00-A09	Intestinal infectious diseases	M	65	0	0	6	59
		F	63	0	0	6	57
A37	Whooping cough	M	1	0	0	0	1
		F	0	0	0	0	0
E40-E64	Nutritional deficiencies	M	1	0	0	0	1
		F	0	0	0	0	0
G00-G98	Diseases of the nervous system	M	92	0	2	26	64
		F	61	0	4	16	41
G00,G03	Meningitis	M	37	0	2	18	17
		F	30	0	4	12	14
J00-J98	Diseases of the respiratory system	M	734	0	3	66	665
		F	609	1	3	51	554
J12-J18	Pneumonia	M	705	0	3	65	637
		F	587	1	3	50	533
J10,J11	Influenza	M	0	0	0	0	0
		F	0	0	0	0	0

# Pauschalisierte Entgeltsysteme

## L XIII. Herzchirurgie

---

### 3050

Operative Maßnahmen in Verbindung mit der Herz-Lungen-Maschine zur Herstellung einer extrakorporalen Zirkulation

1850	107,83€	248,01€
------	---------	---------

### 3051

Perfusion der Hirnarterien, zusätzlich zur Leistung nach Nummer [3050](#)

1290	75,19€	172,94€
------	--------	---------

### 3052

Perfusion der Koronararterien, zusätzlich zur Leistung nach Nummer [3050](#)

1110	64,70€	148,81€
------	--------	---------

### 3053

Perfusion von Arterien eines anderen Organs, zusätzlich zur Leistung nach Nummer [3050](#)

1110	64,70€	148,81€
------	--------	---------

### 3054

Operative extrathorakale Anlage einer assistierenden Zirkulation

1850	107,83€	248,01€
------	---------	---------

### 3055

Überwachung einer assistierenden Zirkulation, je angefangene Stunde

554	32,29€	74,27€
-----	--------	--------

Die Leistung nach Nummer 3055 ist nur während einer Operation berechnungsfähig.

### 3060

Intraoperative Funktionsmessungen am und/oder im Herzen

554	32,29€	74,27€
-----	--------	--------

### 3065

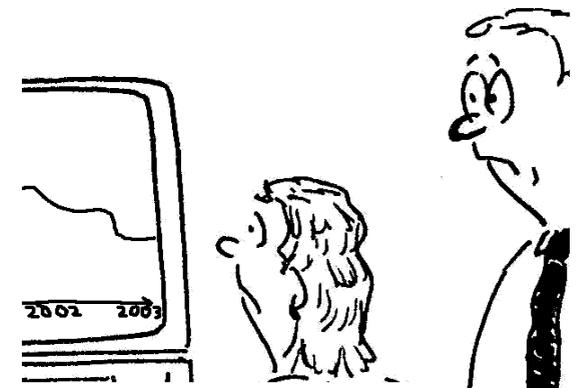
Operation am Perikard, als selbständige Leistung

2000	116,57€	268,13€
------	---------	---------

# ...erfordern medizinische Terminologiesysteme

- ICD
- OPS 301
- DRG
- ICF
- LOINC
- SNOMED
- etc., etc.

**Strukturierte  
Daten**



# Natürliche Sprache



**Anamnese:**  
 Vater verstorben an Bronchial-Karzinom, Mütter verstorben an den Folgen einer Pneumonie, Mütter Diabetes mellitus, 5 gesunde Kinder.

**Systemanamnese:**  
 Derzeit Appetitlosigkeit, Trockengewicht um 75 Kg, derzeit 80 Kg. Miktio: gelegentlich Harn-verhalt, gehäuft Harnwegsinfekte, derzeit keine Algurie. Vor Dialyse keine Rest-Diurese. Stuhlgang obstipiert, benutzt regelmäßig Abführmittel. Vor NTX starker Juckreiz, Seit NTX deutlich rückläufig. Kein Husten/Auswurf. Noxen: Nichtraucherin, kein Alkohol.

**Soziale Anamnese:**  
 Früher Arbeiterin in der Elektronikbranche, dann Hausfrau, verheiratet, lebt mit dem Ehemann zusammen.

**Allergien:** Keine bekannt.

**Medikation bei Aufnahme:**  
 Uroogant 1-1-1, Peptid mit 0-0-0-1, Celcept 2x1 g, Bayolensin 3 x 1, Cyt 0.2 1x1, Ludomil 50 mg 1 x 1, Sandimmun 2 x 150 mg, Clexane 0.4 ml 1 x täglich s.c.

**Status bei Übernahme:**  
 58-jährige Patientin in vorgealtertem, reduziertem Allgemein- und adipösem Ernährungsstatus (80 Kg Gewicht bei 160 cm Körpergröße), RR 170/80 mm Hg, Puls 66/Minute, regelmäßig, punktförmige Depigmentierungen an beiden Unterarmen bei Zustand nach heftigem Kratzen wegen Juckreiz. Keine zervikalen Lymphome. Mundschleimhaut trocken, Zunge weißlich belegt, Rachenschleimhaut reizlos, Tonsillen schlecht einsehbar, Schilddrüse nicht vergrößert. Pulm: Sonorer Klopfeschall und vesikuläres Atemgeräusch, Cor: Spitzenstoß nicht tastbar, leise, reine Herzklappe. 3/6 spindelförmiges Systolikum und 1-2/6. Decrescendo-Sofort-diastolikum über der Aorta mit Fortleitung in die Karotiden. Kein abdominales und inguinales Strömungsgeräusch. Abdomen: Bei Adipositas Organgrenzen schlecht beurteilbar. Leber/Milz nicht vergrößert. Reizlose Narbe im Bereich des rechten Unterbauches bei Zustand nach NTX. Dort leichte Druckdolenz. Wirbelsäule nicht klopfmerthaft. Bds. Unterschenkelödeme. Feinschlägiger Tremor beim Arm-Vorhalte-Versuch. Pupillen isokor, Lichtreaktion prompt. Finger-Nase-Versuch bds. unsicher, ataktisch, Reflexe seitengleich.

# Strukturierte Daten



ICD 10 Codes	Cause of Death	Sex	Number of Deaths				
			< 1 year	1-4	5-9	10-14	15-19
	All causes	M	2495	168	703	416	1119
		F	1055	139	448	279	1020
		M	8	0	0	0	0
		F	95	0	1	9	85
A00-B99	Infectious and parasitic diseases	M	93	0	2	9	82
		F	63	0	0	6	59
A00-A09	Intestinal infectious diseases	M	63	0	0	6	57
		F	1	0	0	0	1
A37	Whooping cough	M	8	0	0	0	8
		F	0	0	0	0	0
E40-E64	Nutritional deficiencies	M	0	0	0	0	0
		F	0	0	0	0	0
G00-G08	Diseases of the nervous system	M	92	0	2	26	64
		F	61	0	4	16	41
G08_G09	Meningitis	M	37	0	2	18	17
		F	30	0	4	12	14
J00-J99	Diseases of the respiratory system	M	724	0	3	66	655
		F	609	1	3	51	564
J12-J18	Pneumonia	M	705	0	3	65	613
		F	587	1	3	50	531
J10-J11	Influenza	M	8	0	0	0	8
		F	8	0	0	0	8
Q00-Q99	Congenital anomalies	M	454	38	133	100	183
		F	376	35	94	61	185
Q01-Q05	Spina bifida and hydrocephalus	M	71	5	16	5	45
		F	74	2	8	17	47
Q20-Q28	Congenital anomalies of heart and circulatory system	M	222	8	67	67	90
		F	149	9	42	23	75
P00-P96	Certain conditions originating in the perinatal period	M	679	127	559	167	6
		F	571	97	340	131	3
P10-P15	Birth trauma	M	36	13	75	19	0
		F	34	4	25	5	0

# Datenerfassung

**Natürliche  
Sprache**



**Strukturierte  
Daten**



**+**

**-**

**Qualität  
Kosten**

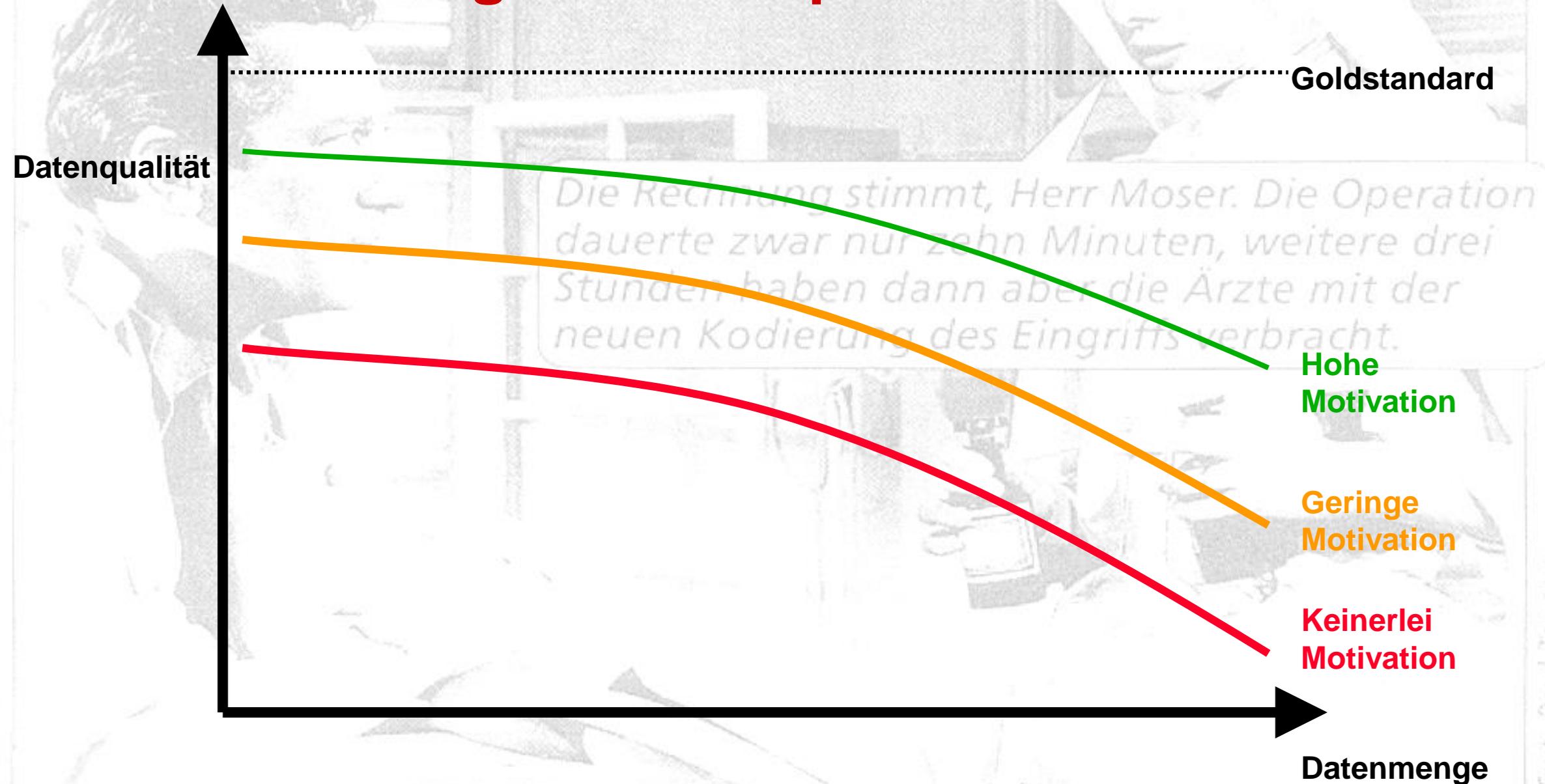
**-**

**+**



*Die Rechnung stimmt, Herr Moser. Die Operation dauerte zwar nur zehn Minuten, weitere drei Stunden haben dann aber die Ärzte mit der neuen Kodierung des Eingriffs verbracht.*

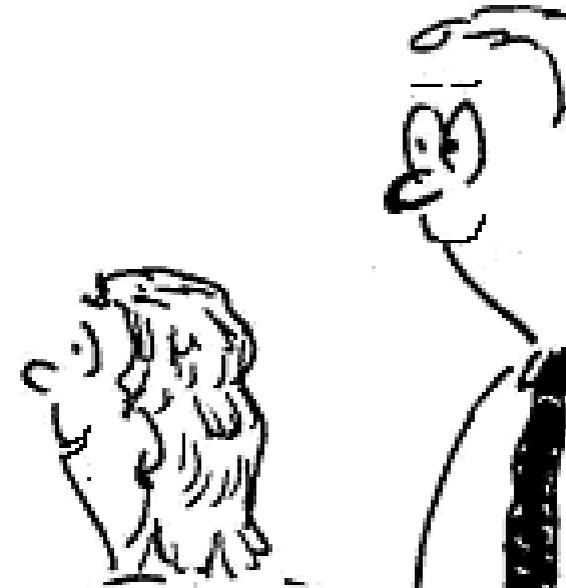
# Erfassung strukturierter Daten: Abhängigkeit: Datenmenge – Datenqualität - Motivation



# Datenauswertung

## Natürliche Sprache

## Strukturierte Daten



# Beispiele Datenauswertung



*„Ich brauche die Arztbriefe der Patienten mit den Merkmalen A und B, die von Dr. C mit D behandelt wurden“*

# Beispiele Datenauswertung



**Dokumentenrecherche am Medizinischen Arbeitsplatz**

# Beispiele Datenauswertung



## Dokumentenrecherche am Medizinischen Arbeitsplatz

*„Wir möchten Kostenprofile unterschiedlicher Patientengruppen miteinander vergleichen“*

# Beispiele Datenauswertung



**Dokumentenrecherche am Medizinischen Arbeitsplatz**



**Medizinisches Controlling**

# Beispiele Datenauswertung



**Dokumentenrecherche am Medizinischen Arbeitsplatz**



**Medizinisches Controlling**



*„Wie haben sich bestimmte Qualitätsindikatoren in meiner Abteilung entwickelt?“*

# Beispiele Datenauswertung



**Dokumentenrecherche am Medizinischen Arbeitsplatz**



**Medizinisches Controlling**



**Qualitätssicherung**

# Beispiele Datenauswertung



**Dokumentenrecherche am Medizinischen Arbeitsplatz**



**Medizinisches Controlling**



**Qualitätssicherung**



*„Ich brauche verlässliche Daten aus klinischen Registern“*

# Beispiele Datenauswertung



**Dokumentenrecherche am Medizinischen Arbeitsplatz**



**Medizinisches Controlling**



**Qualitätssicherung**



**Klinische Epidemiologie**

# Beispiele Datenauswertung



**Dokumentenrecherche am Medizinischen Arbeitsplatz**



**Medizinisches Controlling**



**Qualitätssicherung**



**Klinische Epidemiologie**



*„Ich möchte Patienten für eine Studie rekrutieren“*

# Beispiele Datenauswertung



**Dokumentenrecherche am Medizinischen Arbeitsplatz**



**Medizinisches Controlling**



**Qualitätssicherung**



**Klinische Epidemiologie**



**Klinische Forschung**

# Beispiele Datenauswertung



**Dokumentenrecherche am Medizinischen Arbeitsplatz**



**Medizinisches Controlling**



**Qualitätssicherung**



**Klinische Epidemiologie**



**Klinische Forschung**



*Ich will Phänotypdaten, um meine Hypothesen zu belegen*

# Beispiele Datenauswertung



**Dokumentenrecherche am Medizinischen Arbeitsplatz**



**Medizinisches Controlling**



**Qualitätssicherung**



**Klinische Epidemiologie**



**Klinische Forschung**



**Biomedizinische Grundlagenforschung**

# Auswertungsproblem

---

Routinemäßig  
erhobene  
strukturierte Daten  
(ICD, OPS):

- unzureichend
- zu grob
- unsicherer  
Qualität



## Typische Anwendungsszenarien

---



Dokumentenrecherche am Medizinischen Arbeitsplatz



Medizinisches Controlling



Qualitätssicherung



Klinische Epidemiologie

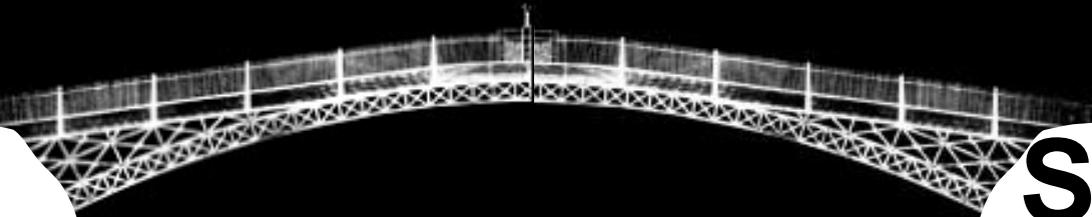


Klinische Forschung



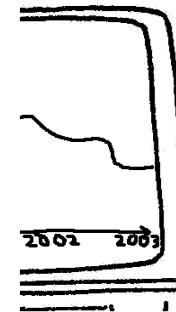
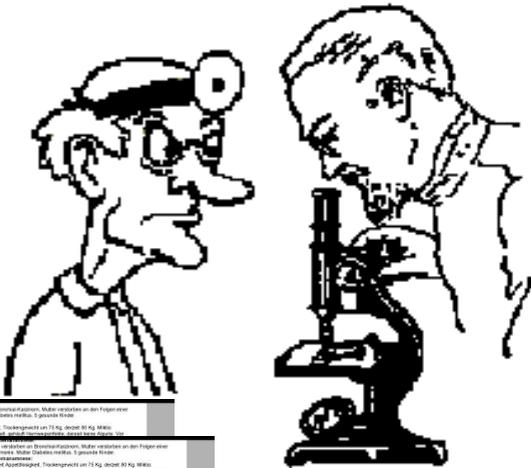
Biomedizinische Grundlagenforschung

# Ziel: Überbrückung



## Natürliche Sprache

## Strukturierte Daten

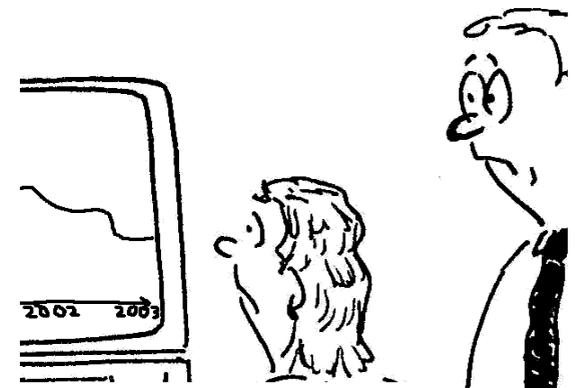
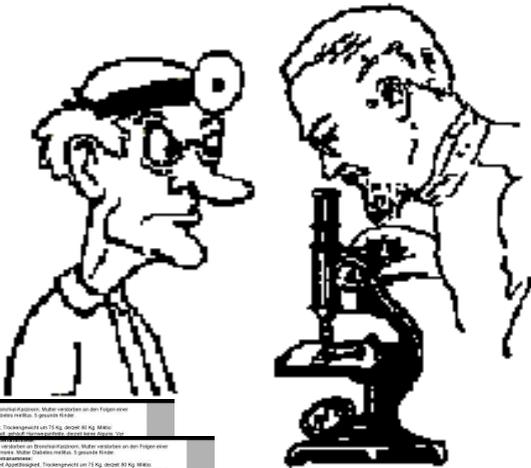


# Methode: Text Mining



## Natürliche Sprache

## Strukturierte Daten



# Text Mining: Im Wesentlichen zwei Hauptszenarien

---

- **Informationsextraktion:**

Automatische Extraktion von für eine Fragestellung relevanter Fakten in unstrukturierten Texten

- **Dokumentenretrieval:**

gezieltes Suchen nach relevanten Dokumenten in großen Dokumentensammlungen

# Text Mining: Im Wesentlichen zwei Hauptszenarien

---

- **Informationsextraktion:**

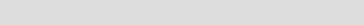
Automatische Extraktion von für eine Fragestellung relevanter Fakten in unstrukturierten Texten

- **Dokumentenretrieval:**

gezieltes Suchen nach relevanten Dokumenten in großen Dokumentensammlungen

# Anwendungsbeispiel Informationsextraktion: Tumordokumentation

shadow was pointed out on a routine chest X-ray film, but she had no further examination. Physical examination on admission revealed purpura of the upper and lower extremities, swelling of the gums and tonsils, but no symptoms showing the complication of myasthenia gravis. Hematological tests revealed leucocytosis: WBC count 68 700/ $\mu$ l (blasts 11.5%, myelocytes 0.5%, bands 2.0%, segments 16.0%, monocytes 65.5%, lymphocytes 4.0%, atypical lymphocytes 0.5%), Hb 7.1 g/dl (reticulocytes 12%) and a platelet count of  $9.1 \times 10^4$ / $\mu$ l. Further laboratory examination revealed elevated serum lactic dehydrogenase (589 U/l), vitamin B<sub>12</sub> (2010 pg/ml) and ferritin (650.0 ng/ml). Human chorionic gonadotropin and  $\alpha$ -fetoprotein levels were normal. A bone marrow aspiration revealed hypercellular bone marrow with a decreased number of erythroblasts and megakaryocytes and an increased number of monoblasts that were positive for staining by  $\alpha$ -naphthyl butyrate esterase and negative for staining by naphthol ASD chloroacetate esterase. Chest X-ray upon admission revealed a mediastinal mass and an elevated left diaphragm. Computed tomography (CT) of the chest showed a left anterior mediastinal mass. Based on these findings, the patient was diagnosed with a mediastinal tumor accompanied by AMoL. First, in June 1991, the patient was treated with DCMP therapy: daunorubicin (DNR) (25 mg/m<sup>2</sup>, days 1, 2, 3, 4, 6 and 8), cytosine arabinoside (Ara-C) (100 mg/m<sup>2</sup>, days 1-9), 6MP-riboside (6-MP) (70 mg/m<sup>2</sup>, days 1-9) and prednisolone (PSL) (20 mg/m<sup>2</sup>, days 1-9), followed by five courses of consolidation chemotherapy [1, DCMP; 2, ID-Ara-C:adriacin (ADR), vincristine (VCR), Ara-C, PSL; 3, DCMP; 4, ID-Ara-C; 5, A-triple V: Ara-C, VP-16, VCR, vinblastine (VBL)]. After induction chemotherapy, a hematological examination and bone marrow findings had improved to normal, and complete remission was attained. Chest CT scan after chemotherapy in November 1991 revealed regression of the mediastinal tumor. An invasive thymic tumor was suspected and surgery was undertaken in January 1992. The tumor (50  $\times$  45  $\times$  45 mm), located mainly in the anterior mediastinum, was strongly adhered to the adjacent tissues. Resection of the tumor included the left upper lobe of the lung, the phrenic nerve and pericardium. The histological finding was that the tumor cells have large, vesicular nuclei and prominent nucleoli, but keratinization was unclear. The results of immunohistochemical finding of anti-TdT was negative. From these findings, we diagnosed poorly or moderately differentiated squamous cell carcinoma of the thymus. The postoperative course was uneventful. The patient underwent radiation therapy of the mediastinum and left hilum at doses of 4000 cGy delivered over 4 weeks. She was discharged in March 1992. After the first AMoL remission, the patient suffered a relapse six times and was repeatedly admitted for chemotherapy. During these periods, chest X-ray and CT revealed no recurrence of the mediastinal tumor. During her tenth admission, the patient developed pneumonia during chemotherapy and died in October 1996. No autopsy was performed.

Tumorregister - Template	
Datum	
Erstdiagnose	
Primärlokalisation	
Grading	
Staging	
Morphologie	
Datum	
Ersttherapie	
Chemotherapie	
Bestrahlung	

# Text Mining: Im Wesentlichen zwei Hauptszenarien

---

- **Informationsextraktion:**

Automatische Extraktion von für eine Fragestellung relevanter Fakten in unstrukturierten Texten

- **Dokumentenretrieval:**

gezieltes Suchen nach relevanten Dokumenten in großen Dokumentensammlungen

# Anwendungsbeispiel Dokumentenrecherche: Prototyp Hautklinik Freiburg

PIZ:  Nachname:  Vorname:  Geschlecht:

Ergebnisse in 3 msec gefunden. 100% Genauigkeit. -- Zeige Ergebnisse 1 bis 4:

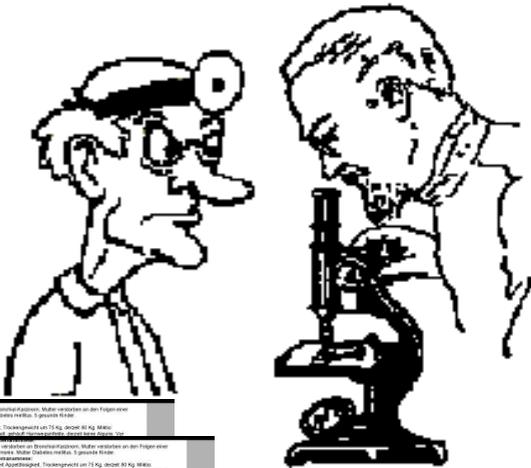
	22990373	Presley	Elis	08.01.1935	♂
... Anamnese und Befund: Seit 01.03.06 kam es bei Herrn Presley zur Ausbildung von <b>Erythemen</b> mit teilweise papulösen Hautveränderungen am linken Unterschenkel welche mit starkem <b>Juckreiz</b> und Brennen vergesellschaftet waren. Im Laufe der letzten 2 Tage sei es dann zu einer ...					
Ersteller: Dr. Albert Schweitzer    Erstellung: 06.03.2006    Übermittlung: 09.03.2006					
	25724135	Hendrix	Jimi	27.11.1942	♂
... unserer stationären Behandlung befand. Diagnosen: 1 <b>Pruriginöses Ekzem</b> 2 Demenz 3 Harninkontinenz 4 Eingeschränkte Mobilität bei Hüftkopfnekrose rechts 5 Schwerhörigkeit Anamnese und Befund: Herr Hendrix lebt in einem Pflegeheim und klagte dort seit ca. 4-6 Monaten über starken <b>Juckreiz</b> am ...					
Ersteller: Dr. Albert Schweitzer    Erstellung: 09.03.2006    Übermittlung: 09.03.2006					
	22118404	Joplin	Janis	19.01.1943	♀
... am unteren Rücken. Ebenso sieht man am Decolletée zwei ca. 3 mal 2 cm große <b>erythematöse</b> und nässende Erosionen bei Zustand nach Lasertherapie. Am Hals und prästernal sowie am rechten Unterschenkel weiterhin <b>erythematöse</b> z.T. erosive z. T. hyperkeratotische Hautveränderungen Die ...					
Ersteller: Dr. Albert Schweitzer    Erstellung: 01.02.2006    Übermittlung: 06.03.2006					
	21965936	Kelly	Grace	12.11.1929	♀
... Ecural Lösung). Zusätzlich erhielt die Patientin Antihistaminika gegen den <b>Juckreiz</b> . Bei Aufnahme finden sich streckseitenbetont am gesamten Integument multiple <b>erythemasquamöse</b> Plaques. Die Kopfhaut ist gerötet und weist eine feine weißliche Schuppung auf. Axillär beidseits inguinal und in der Rima ani ...					
Ersteller: Dr. Albert Schweitzer    Erstellung: 02.03.2006    Übermittlung: 02.03.2006					

# Werkzeuge (I): Natürlichsprachliche Systeme



## Natürliche Sprache

## Strukturierte Daten



# Natürlichsprachliche Systeme – Forschungsgegenstand der Computerlinguistik

---

- Grundprinzipien:

- Analyse der Struktur von geschriebenem Text (Syntax)

- Analyse der Phrasenstruktur

- Abbildung auf semantische Einheiten (Einträge in Wörterbüchern)

- Methodische Ansätze

- Symbolische Ansätze (Regelsysteme)

- Probabilistische Ansätze (statistisch basierend auf echten Daten)

- Problematik:

- Vieldeutigkeit

- Limitierte Verfügbarkeit von Ressourcen (Grammatiken, Lexika, Ontologien, annotierte Korpora)

- Inhärente Komplexität sprachverarbeitender Algorithmen

Derzeitiger  
Forschungs- und  
Anwendungsschwerpunkt:

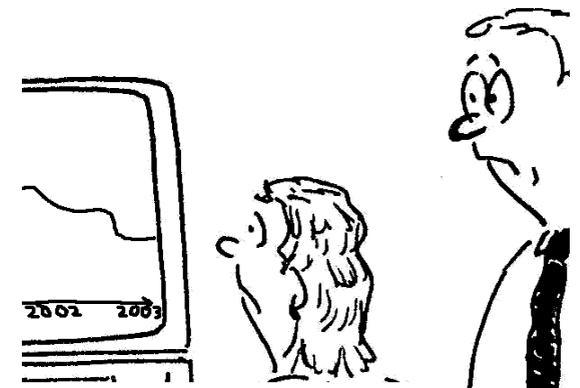
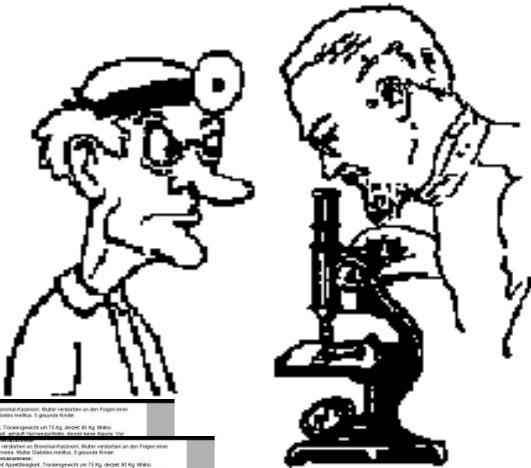
Bio-Text Mining  
(Faktenextraktion aus  
Literaturabstracts)

# Werkzeuge (II): Terminologiesysteme, Ontologien



## Natürliche Sprache

## Strukturierte Daten



# Ontologien / Terminologiesysteme: Interdisziplinärer Forschungsgegenstand

---

- Grundprinzip:
  - Beschreibung der Eigenschaften (Umwelt) Gegenstände und Prozesse einschließlich Zusammenhänge
- Methodische
  - Logikbasiert (Axiomensystem)
  - Berechenbar (algorithmisch) deduzierbar / deduktionell folgerbar.
- Problematik:
  - Konsensfindung
  - Handhabung
- Beispiel:
  - SNOMED CT (mouse anatomy, mouse development, zebrafish, C. elegans, drosophila, ...)
  - I,w,S. auch ICD, OPS, MeSH, ...
  - Open Biological Ontologies

Derzeitiger  
Forschungs- und  
Anwendungs-  
schwerpunkt:

**Bio-Ontologien**

(Cell, cell component, molecular function,  
biological process, gene / protein  
sequence, chemicals of biological interest,

# Beispiel: ontologisches Schließen durch Wissensbasierung

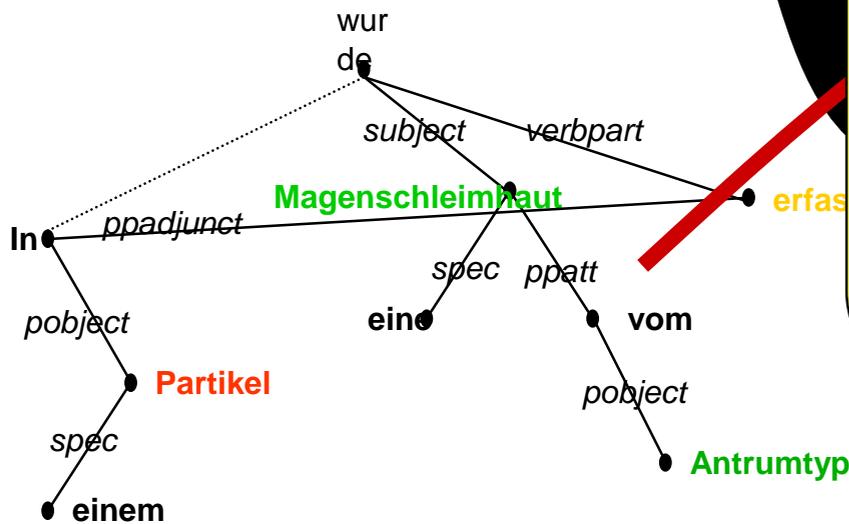
---

- Eine offene Tibiaschaftfraktur ist
  1. eine Fraktur des Unterschenkels, weil Tibiaschaft Teil des Schienbeins, welches Teil von Unterschenkel
  2. eine offene Verletzung, weil allen Frakturen Verletzungen sind
  3. eine den Bewegungsapparat betreffende Verletzung, da Tibia Teil von Bewegungsapparat und Fraktur eine Art von Verletzung
  4. eine Verletzung die zur Funktionseinschränkung führt, da Stabilität der u. Extremität abhängig von Stabilität der Tibia und Schaftfrakturen immer die Knochenstabilität beeinträchtigen
  5. Ein Infektionsrisiko, da (2.) und offene Verletzungen immer infektionsgefährdet und

# Kombination: Natürlichsprachliche Systeme – Ontologien "Sprachverstehen"

Weitgehende Erhaltung Information

## syntaktische Repräsentation



Bisher  
Laborprototypen  
vorbehalten und nicht  
in ausreichendem  
Maße in  
Realdomänen  
skalierbar

## semantische Repräsentation (Instanziierung einer Ontologie)



# Kombination: Natürlichsprachliche Systeme – Ontologien "Morphosemantische Indexierung"

Starke Abstraktion

**syntaktische  
Repräsentation**

**semantische  
Repräsentation**

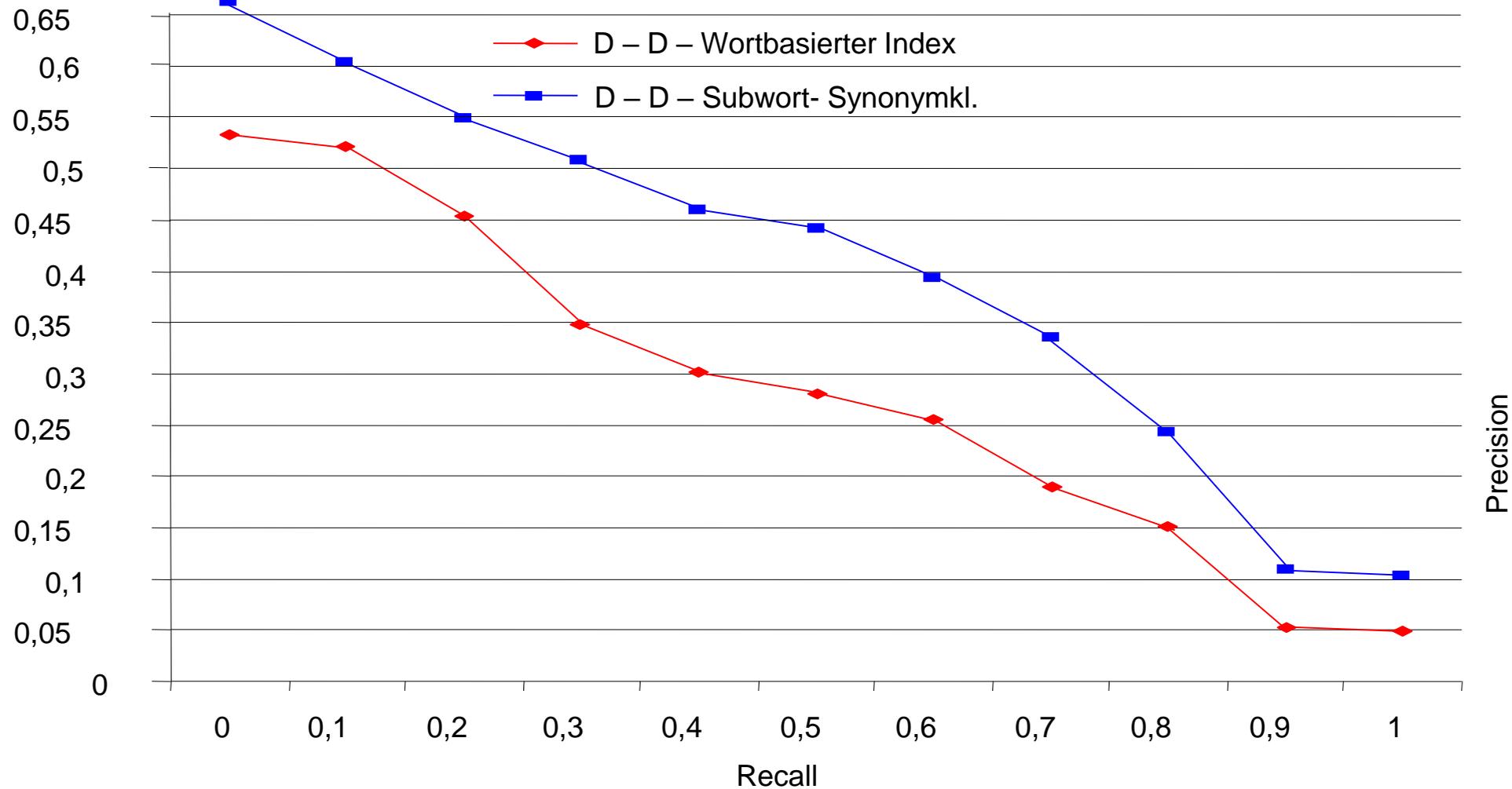
High TSH values suggest the diagnosis of primary hypothyroidism ...

Erhöhte TSH-Werte erlauben die Diagnose einer primären Schilddrüsenunterfunktion ...

#up tsh #value #suggest  
#diagnost #primar #hypo  
#thyre

#up tsh #value #permit  
#diagnost #primar #thyre  
#hypo #function

# MorphoSaurus: semantisches medizinisches Dokumentenretrieval



# Ausblick: Semantisch angereicherte Patientendokumentation

Medizinischer  
Arbeitsplatz:  
Recherche in  
elektronischen  
Patientendokumenten  
(„med. Google“)

**Quelle:  
Befund-  
berichte,  
Arztbriefe  
(freier Text)**

Controlling-  
Arbeitsplatz /  
Falldaten - Data-  
Warehouse

Text-Mining-  
Engine

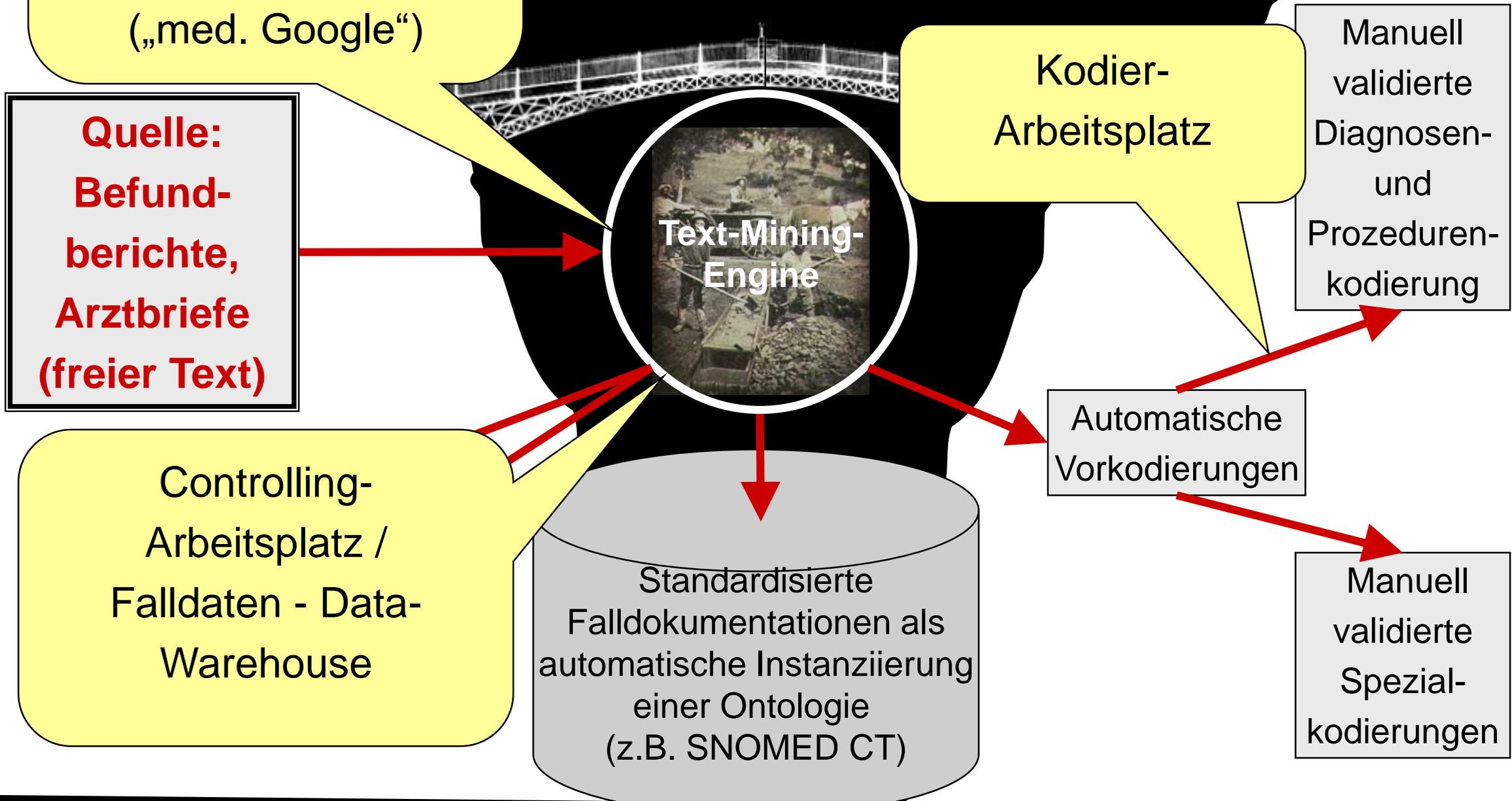
Kodier-  
Arbeitsplatz

Manuell  
validierte  
Diagnosen-  
und  
Prozeduren-  
kodierung

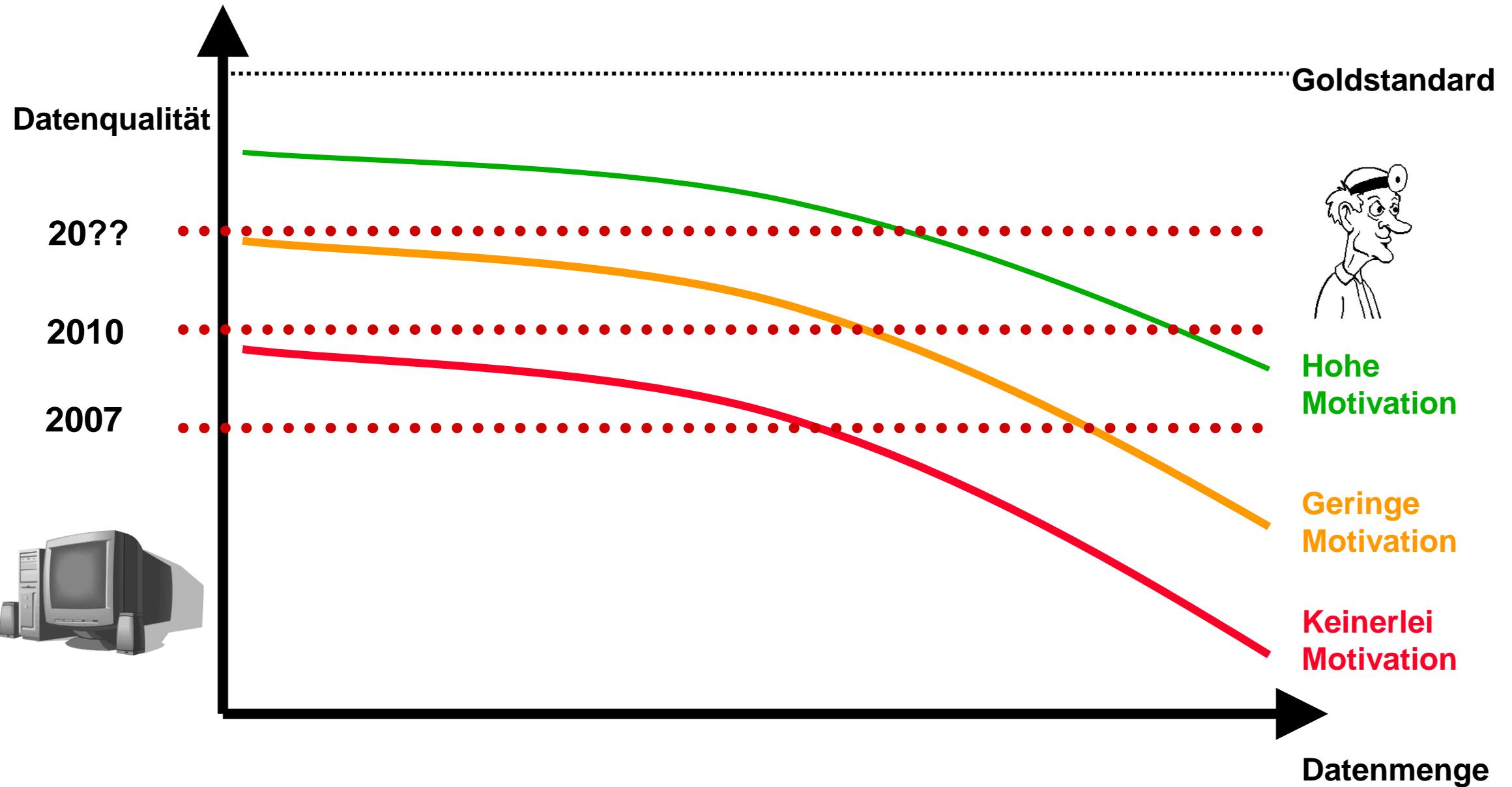
Automatische  
Vorkodierungen

Standardisierte  
Falldokumentationen als  
automatische Instanziierung  
einer Ontologie  
(z.B. SNOMED CT)

Manuell  
validierte  
Spezial-  
kodierungen



# Ausblick: Menschliche vs. Maschinelle Sprachverarbeitung



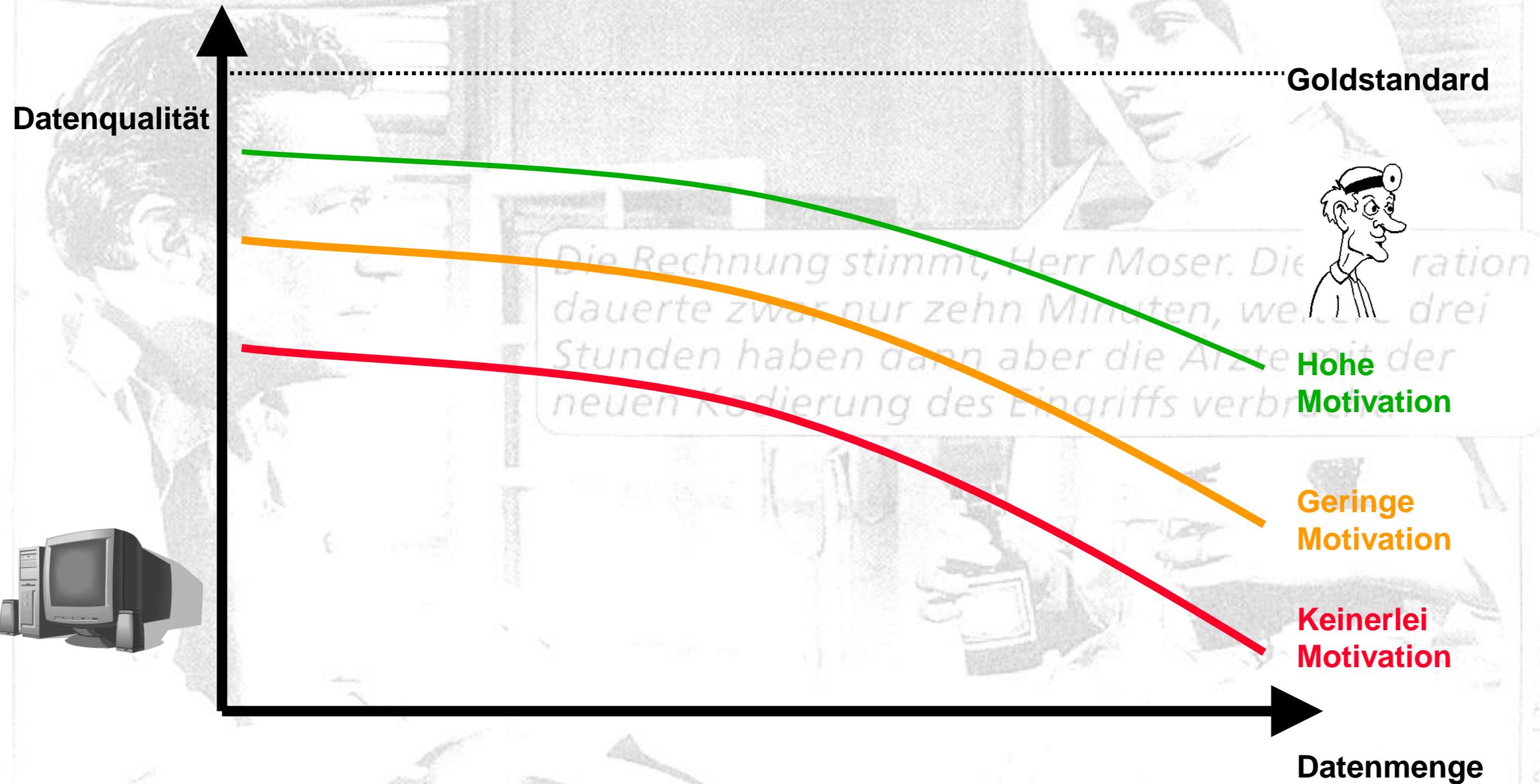


# Herausforderung für medizinische Sprachverarbeitung

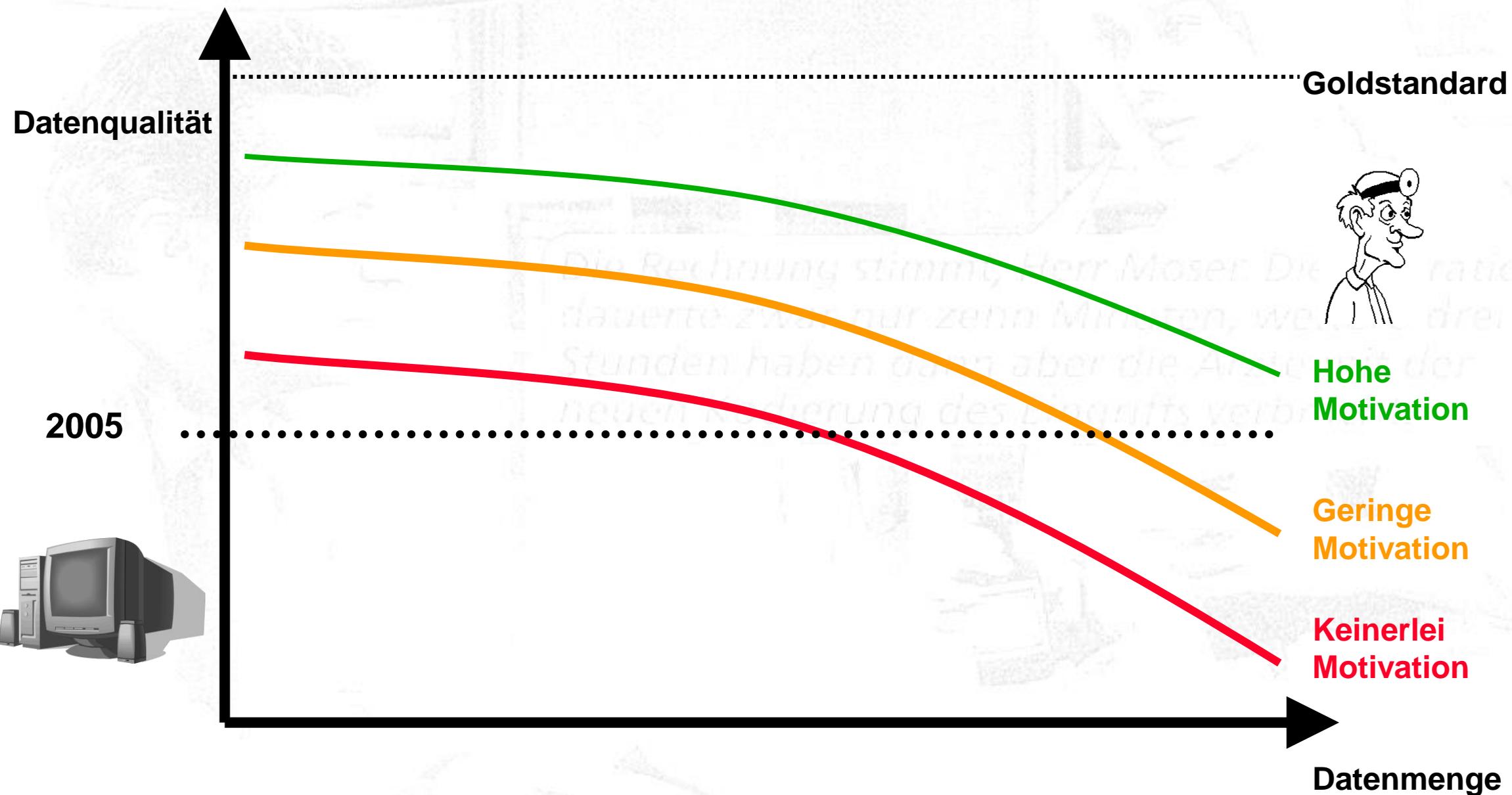
---

- Sehr große Textmengen
- Hohe Anforderungen an Datenschutz
- Lexikalische Produktivität
- Ambiguität
  - Lexikalisch: **Bruch (Hernie)** vs. **Bruch (Fraktur)**
  - Syntaktisch: z.B. Anbindung von PPs
    - “[es wurde] [ Magenschleimhaut vom Antrumtyp ] [erfasst]”
    - “[es wurde] [ Magenschleimhaut] [vom Antrumtyp erfasst]”
  - Semantisch, z.B. Skopus von Quantoren, Negationen, Koordinationen, Gradaussagen
    - *Jede Probe wies einen erhöhten PH-Wert auf*
- Komplexität, Berechenbarkeit, z.B.
  - Abhängenzgrammatiken: **NP-vollständig**
  - Prädikatenlogik höherer Ordnung, Modallogiken: **unentscheidbar**
- Kombination mit medizinischen Terminologiesystem
- Integration von symbolischen und stochastischen Ansätzen

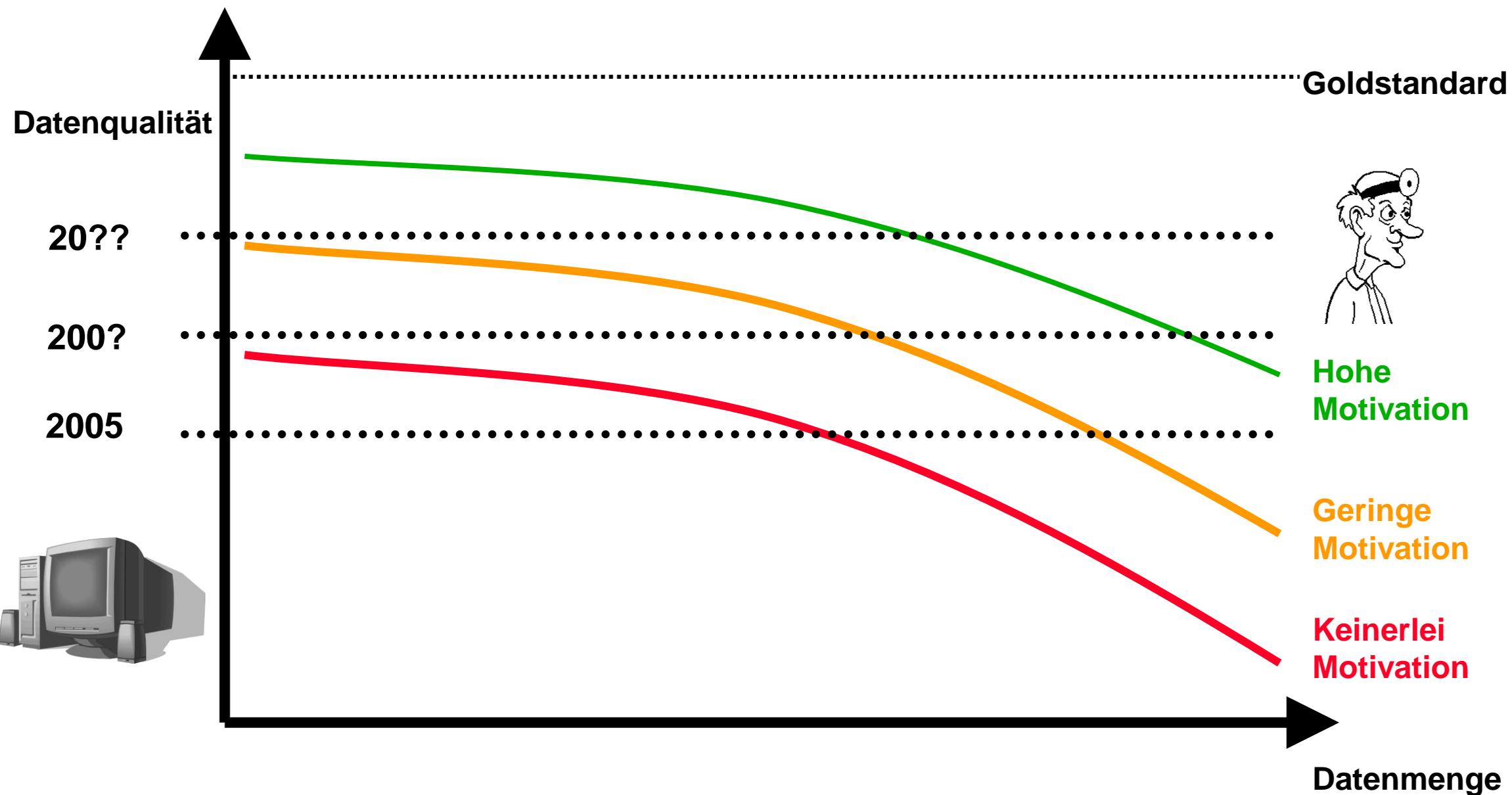
# Ausblick: Menschliche vs. Maschinelle Sprachverarbeitung



# Ausblick: Menschliche vs. Maschinelle Sprachverarbeitung



# Ausblick: Menschliche vs. Maschinelle Sprachverarbeitung



---

# Natürlichsprachliche Systeme als Mehrwert für die Elektronische Krankenakte

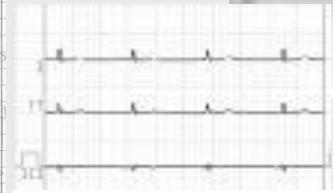
Stefan Schulz

Abteilung Medizinische Informatik

Universitätsklinikum Freiburg

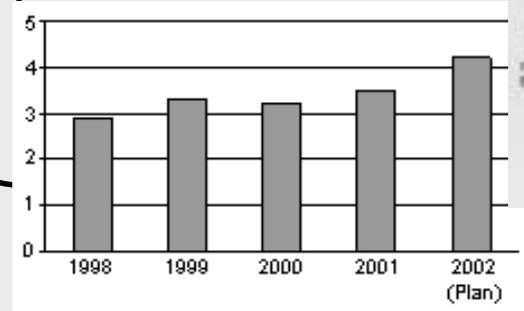
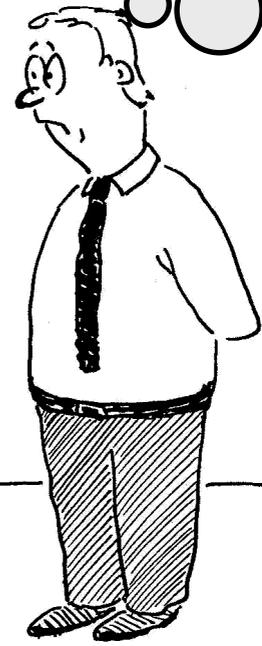
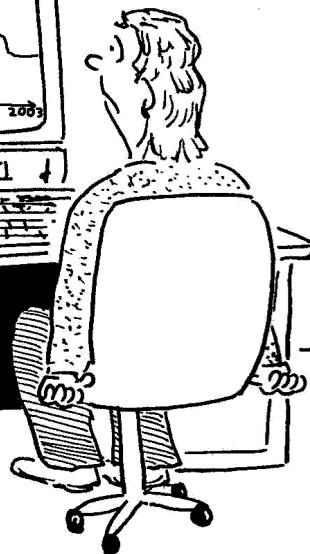
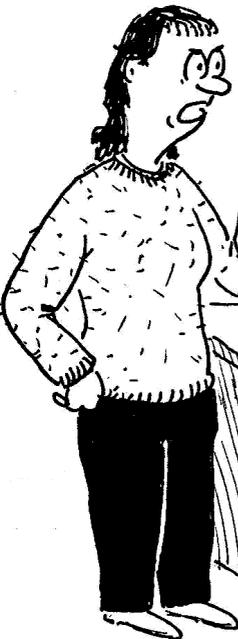


Handwritten medical notes on a grid background. The text is in German and includes patient history and examination findings.



Blutbild, groß		
Leukocyten	DZ	6.5
Erythrocyten	DZ	4.42
Hämoglobin	SP	13.5
Hämatokrit	RP	41
MCV		97
MCH		

**ICD DRG \$301**



**L4**