



54. gmds-Jahrestagung Essen, 7.-10. September 2009

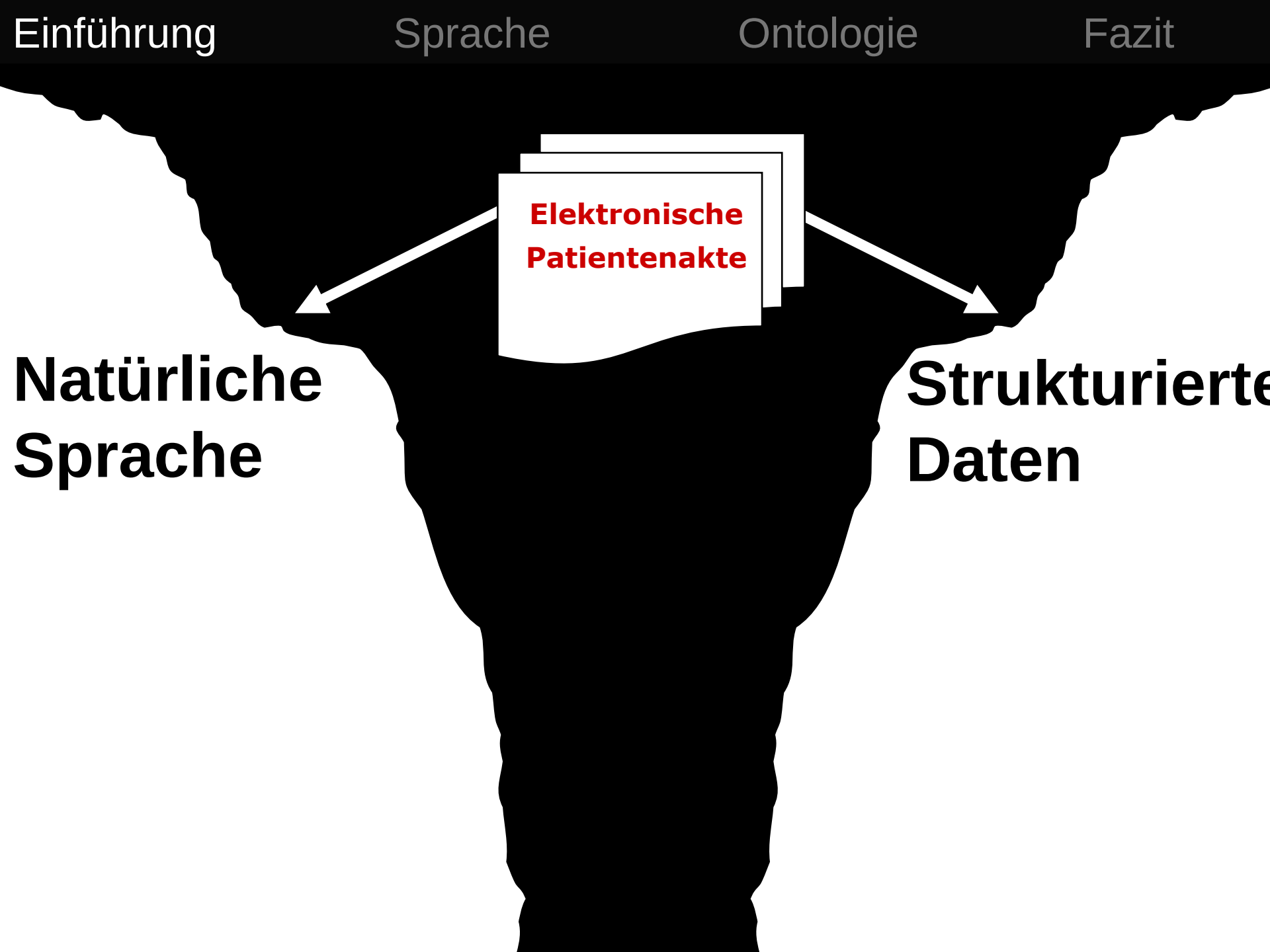
Workshop: Semantische Interoperabilität
durch ontologiegetriebene EHR-Architektur

Semantische Anreicherung natürlichsprachlicher Inhalte in der elektronischen Gesundheitsakte

Stefan SCHULZ

Universitätsklinikum Freiburg

Institut für Medizinische Biometrie und Medizinische Informatik



Einführung

Sprache

Ontologie

Fazit

**Natürliche
Sprache**

[illegible]

Datum	Uhrz.	Pflegebericht – Verlaufsbeschreibung Krankenbeobachtung	Hz.
21.4.	13 ³⁰	Pat. kam mit einer Harnstriege, hat keine Blasen, war schon zur Toilette und hat sich selbstständig gewaschen	ky 6
22.4.	5 ³⁰	Pat. hat in d. Nacht kein Wasser getrunken, sie gab aber ein ^{ein} Becken ^{Becken} mit ^{mit} Wasser ^{Wasser}	sh
22.4.	14 ⁰⁰	Pat. lt. Pflegeplan versorgt	tt
	21 ⁰⁰	Pat. lt. Plan versorgt; ist nach Betastung selbsterregt	AP
23.4.	8 ⁰⁰	Pat. saß fast die ganze Nacht im Bett und konnte nicht im Bett schlafen	sh
	12 ⁰⁰	Pat. hatte 3x Breiig bis dünnflüssigen Stuhlgang (braun)	ky
	14 ⁰⁰	Versorgg. lt. Plan	1
	21 ⁰⁰	Pat. lt. Plan versorgt, hält sich nicht an ihre Bettruhe	AP
24.4.	2 ³⁰	Pat. beim Toilettengang erwacht, hält sich nicht an Bettruhe (sieht besser aus damit)	7

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

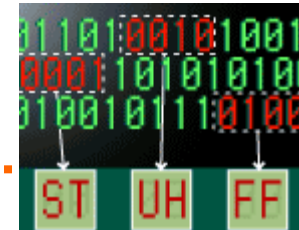
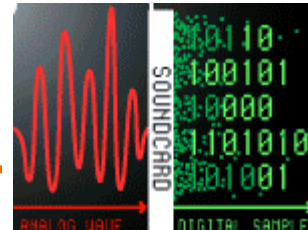
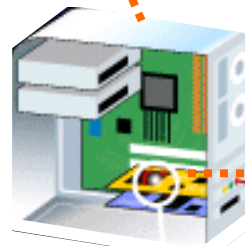
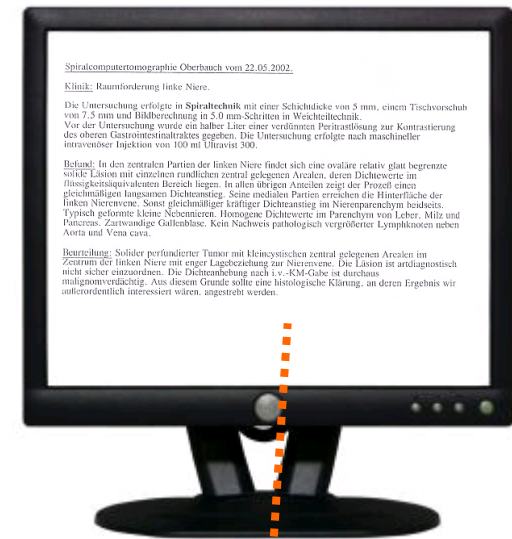
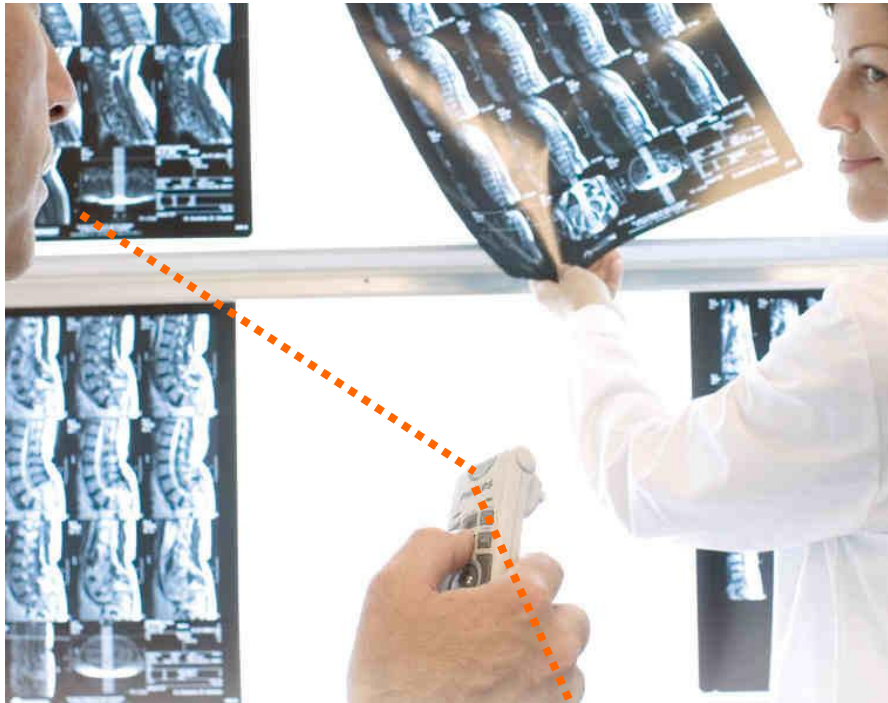
Discharge summary, typed by MD, not proofread

Erotilde, 58 anos # HAS # Obesidade Pcte interna com hist de edema e dor em MID há 3 semanas. Refere que no inicio do quadro apresnetava hiperemia local importante, nega febre. Foi avaliada no posto de saude e iniciado amoxicilina. Fez uso do ATB por 6 dias, com piora da dor, do edema, e surgimento de lesões arredondadas, planas, com bordos bem definido e pequeno ponto escurecido central. Consultou novamente no posto de saude, snedo trocado o ATB para eritromicina, o qual usou por mais 6 dias, com piora das lesões e da dor. Na avaliação inical a pcte apresentava edema e hiperemia imporante de MID, bem como lesoes ulceradas, necróticas com bordos bem definidos, sem secreção. Foi realizado ECO doppler que confirmou TVP em MID sendo então iniciada anticoagulação com enoxaheparina. Solicitada consultoria da dermatolo que realizou biopsia das lesões

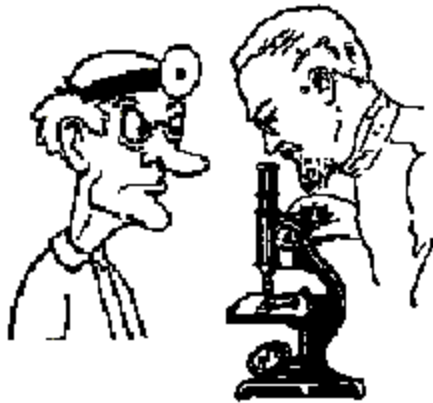
AP: vasculite leucocitoclastica. A pcte recebeu ciprofloxacina por 5 dias e após 2 dias de oxacilina. Recebe alta em bom estado geral, com diminuição importante do edema e da dor em MID. Lesões em fase de cicatrização.

Revisada a literatura: existe associação de vasculite em areas de estase, bem como associação com farmacodermia. Em uso de: Varfarin 5mg 1cp por dia Captopril 25mg TID Amitrptilina 50mg Fluoxetina 40mg HCTZ 25 mg Paciente submetida a fundoplicatura videolaparoscópica. Recebe alta aceitando bem a via oral, sinais vitais estáveis, sem intercorrências ao longo da internação. Plano de retorno ambulatorial.

Text produziert durch Spracherkennung



Natürliche Sprache



... unverzichtbar für Dokumentation und Kommunikation

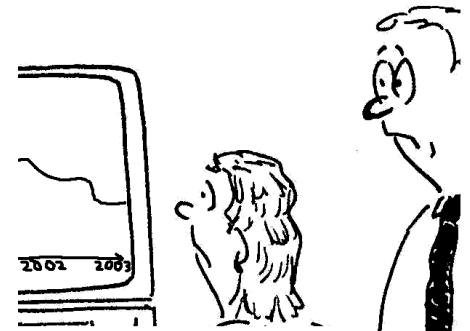
- Krankenversorgung
- Wissenschaft
- Arzt-Patient
- Wissensvermittlung

Natürliche Sprache



600.000

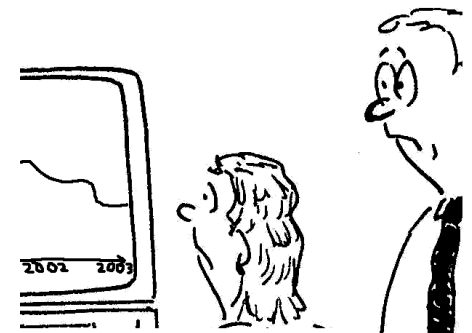
Strukturierte Daten



...unverzichtbar für Klassierung und Abstraktion:

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

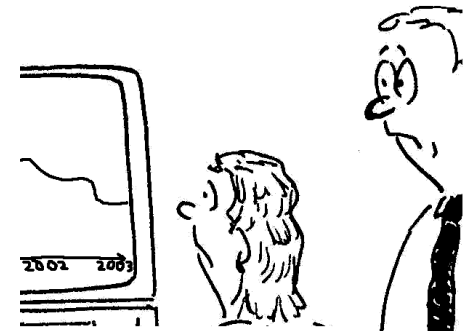
Strukturierte Daten



...erfordern medizinische Terminologien / Ontologien

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

Strukturierte Daten



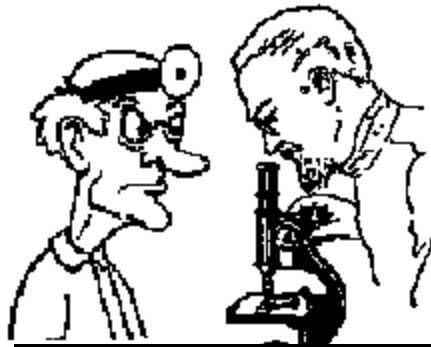
Einführung

Sprache

Ontologie

Fazit

Natürliche Sprache

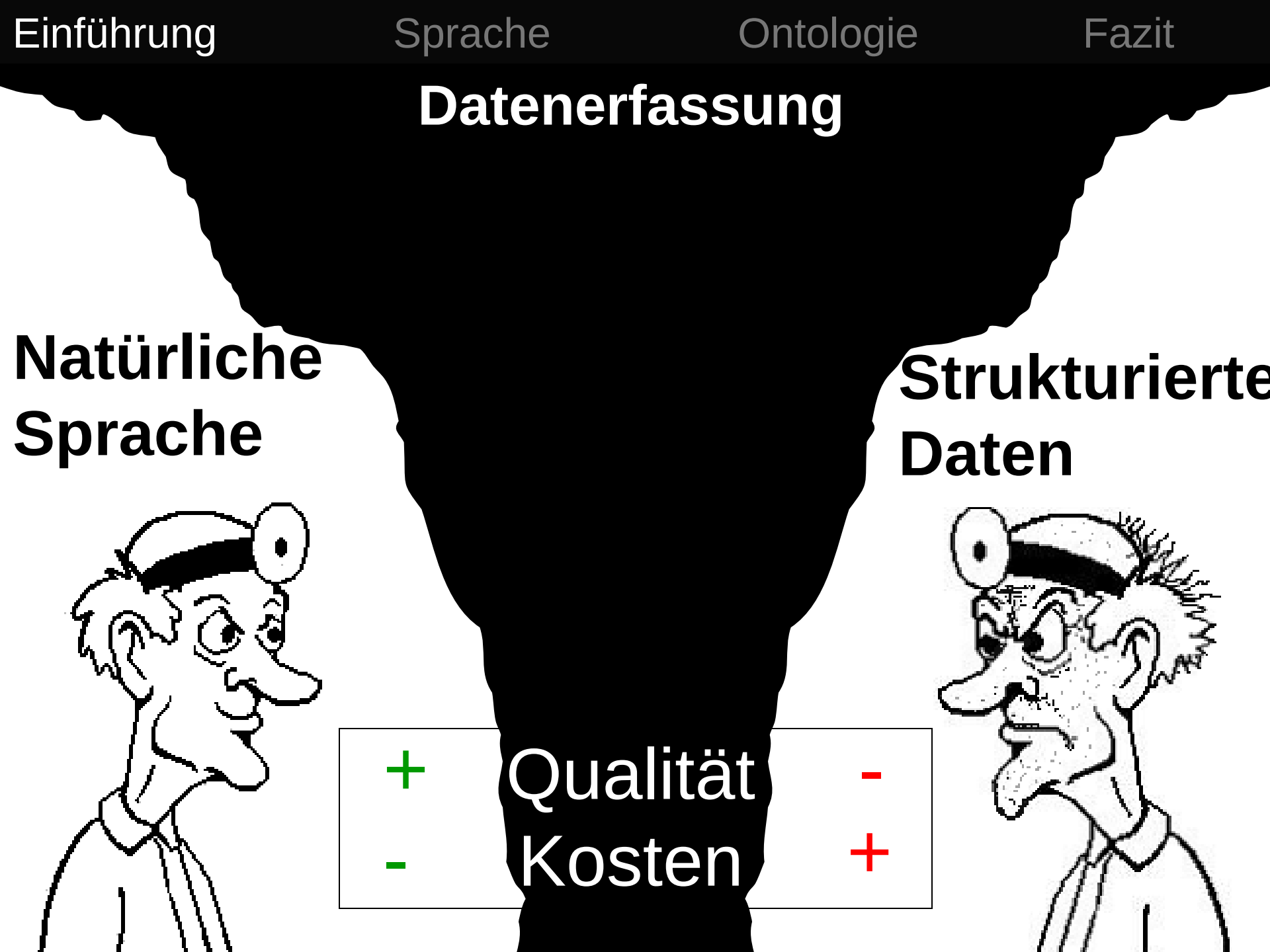


Patientenanamnese:
Vater verstorben an Bronchial-Karzinom. Mutter verstorben an den Folgen einer Pneumonie. Mutter Diabetes mellitus. 5 gesunde Kinder.
Systemanamnese:
Derzeit Appetitlosigkeit, Trockengewicht um 75 Kg, derzeit 80 Kg. Mälig; gelegentlich Häm-hematurie, gelblich Harnwegsinfekte, derzeit keine Ätiologie. Vor Dialyse keine Res-Diurese. Stuhlgang obstipiert, benutzt regelmäßig Abführmittel. Vor NTX starker Juckreiz. Seit NTX deutlich rückläufig. Kein Husten/Auswurf. Nasser, Nichtraucherin, kein Alkohol.
Soziale Anamnese:
Früher Angestellte in der Elektronikbranche, dann Hausfrau, verheiratet, lebt mit dem Ehemann zusammen.
Allergien. Keine bekannt.
Medikation bei Aufnahme:
Ulcogant 1-1-1, Pepud mit 0-0-0-1, Celcept 2x1 g, Beyotensin 3 x 1, Cytel 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 Aufnahme:
58-jährige Patientin in vorgealtertem, reduziertem Allgemein- und adiposum Ernährungszustand (80 Kg Gewicht bei 160 cm Körpergröße), RR 170/90 mm Hg. Puls 66/min, regelmäßig, funktionelle Dysregulationen an beiden Unterarmen bei Zustand nach heftigem Kratzen wegen Juckreiz. Keine zervikalen Lymphome. Mundschleimhaut trocken, Zunge weißlich belegt. Rachenschleimhaut rotlich, Tonsillen schlecht einsehbar. Schilddrüse nicht vergrößert. Pulmo: Sonorer Knöchel und vesikuläres Atemgeräusch. Cor: Spitzenstoß nicht tastbar, leise, reine Herzton. 3/6: spindelförmiges Systolikum und 1-2/6: Decrescendo-Systolikum über der Aorta mit Fortleitung in die Karotis. Kein abdominales und inguinale Strömungsgeräusch. Abdomen: Bei Adipositas Organengrenzen schlecht beurteilbar. Leber/Milz nicht vergrößert. Feinziehe Narbe im Bereich des rechten Unterbaues bei Zustand nach NTX. Dort leichte Druckschmerz. Wirtelsäule nicht knöchelhaft. Bds. Unterschenkelödeme. Feinschlägiger Tremor beim Am-Vorhute-Versuch. Pupillen isokor, Lichtreaktion prompt. Finger-Nase-Versuch bds. unsicher, statisch. Reflexe selektisch.

Strukturierte Daten

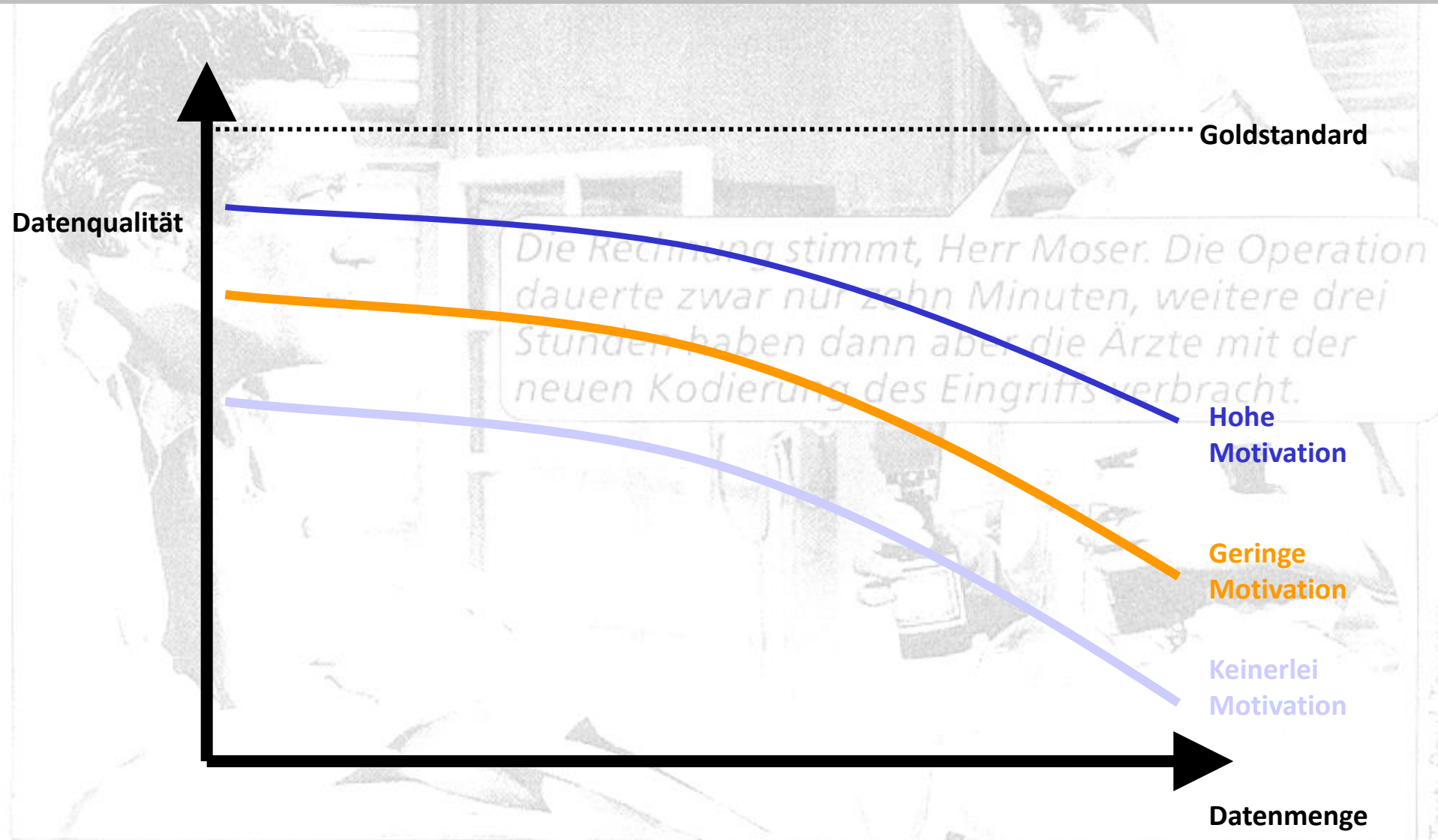


Causes of Death		Number of Deaths (Number of Deaths at age 10 days)				
ICD-10 Code	Cause Groupings	Sex	< 1 year	1-4	5-9	10-14
	All causes	M	1000	100	100	100
		F	1000	100	100	100
A00-B99	Infectious and parasitic diseases	M	100	10	10	10
		F	100	10	10	10
A00-B99	Infectious and parasitic diseases	M	100	10	10	10
		F	100	10	10	10
A17	Whooping cough	M	10	10	10	10
		F	10	10	10	10
E40-E44	Nutritional deficiencies	M	10	10	10	10
		F	10	10	10	10
G00-G09	Diseases of the nervous system	M	100	10	10	10
		F	100	10	10	10
I00-I09	Heart diseases	M	100	10	10	10
		F	100	10	10	10
J00-J09	Diseases of the respiratory system	M	100	10	10	10
		F	100	10	10	10
J10-J18	Pneumonia	M	100	10	10	10
		F	100	10	10	10
J20-J28	Influenza	M	100	10	10	10
		F	100	10	10	10
Q00-Q99	Congenital anomalies	M	100	10	10	10
		F	100	10	10	10
Q00-Q99	Sex Males and heterosexuals	M	100	10	10	10
		F	100	10	10	10
Q00-Q99	Congenital anomalies of heart and circulatory system	M	100	10	10	10
		F	100	10	10	10
P00-P06	Certain conditions originating in the perinatal period	M	100	10	10	10
		F	100	10	10	10
P10-P15	Birth trauma	M	100	10	10	10
		F	100	10	10	10





Erfassung strukturierter Daten: Abhängigkeit: Datenmenge – Datenqualität – Motivation



Electronic
Health
Record

Narrative
Content

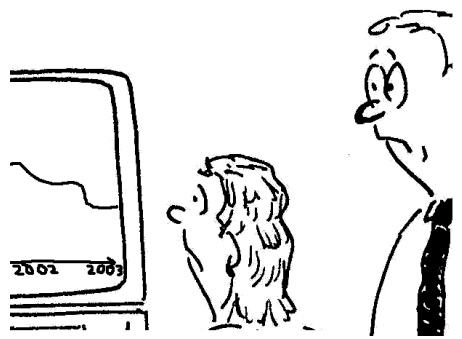
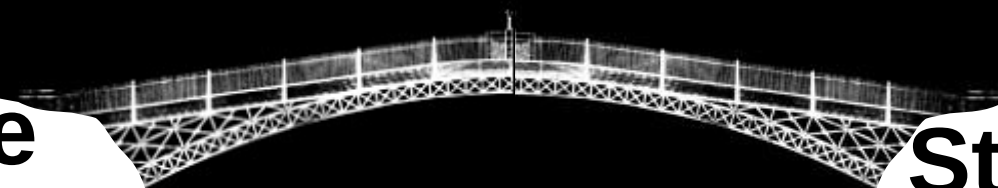
Structured
Content

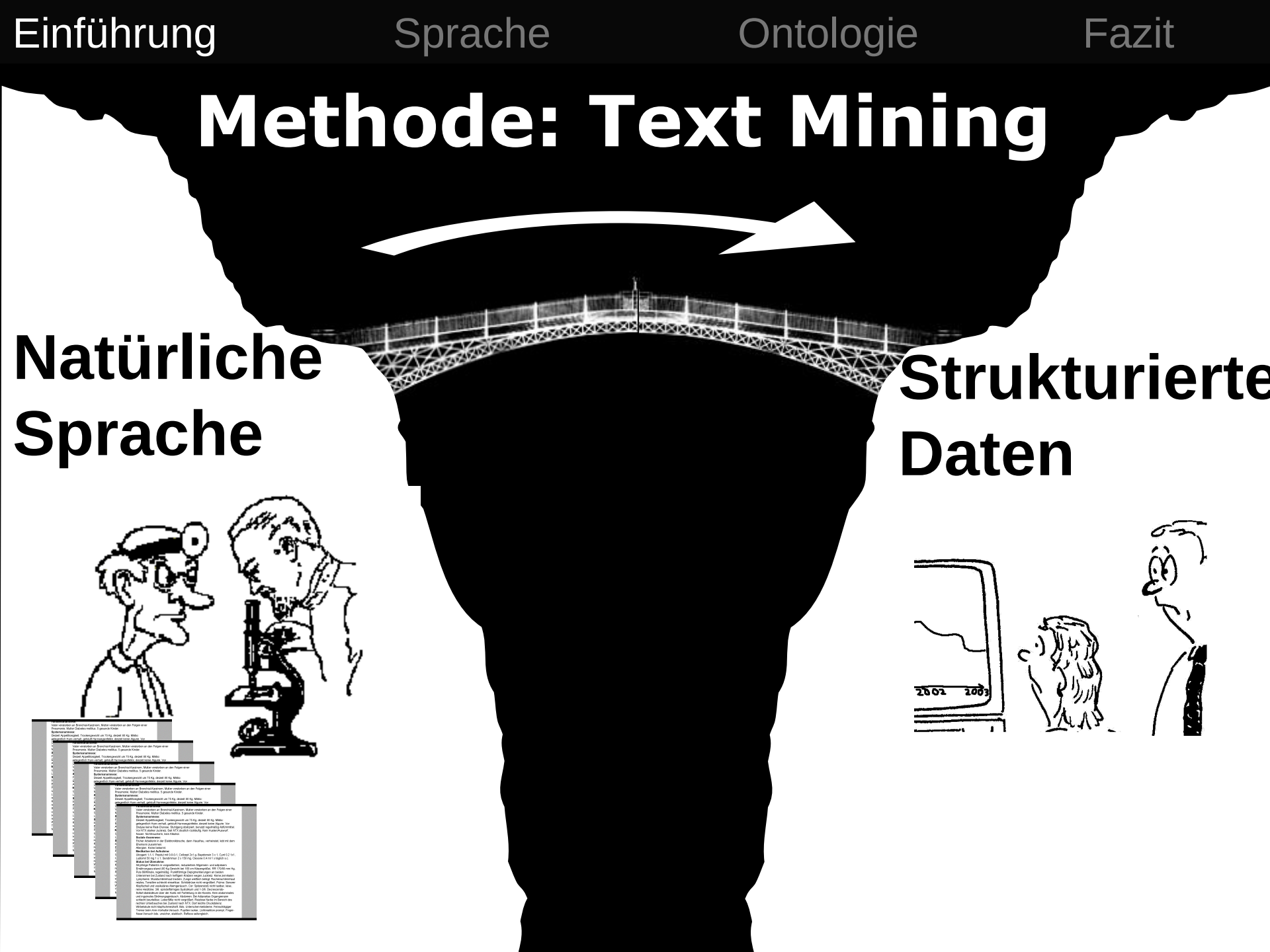
+	Ease of production	-
+	Quality	-
+	human communication	-
-	machine communication	+
-	data analysis	+

Ziel: Überbrückung

Natürliche
Sprache

Strukturierte
Daten

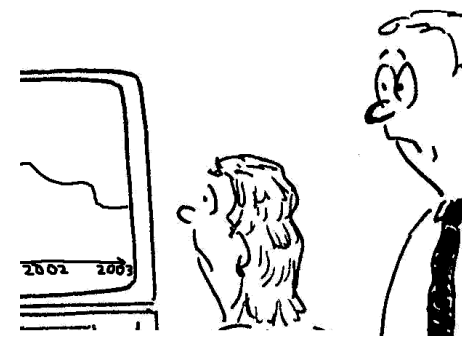




Methode: Text Mining

Natürliche Sprache

Strukturierte Daten



Semantische Anreicherung unter Verwendung von Sprachtechnologie

**Natürliche
Sprache**

Wie soll
strukturierter Inhalt
aus Texten
extrahiert werden?

**Strukturierte
Daten**

Wie kann ein
interoperabler
Repräsentations-
mechanismus
aussehen?

Semantische Anreicherung unter Verwendung von Sprachtechnologie

**Natürliche
Sprache**

Wie soll
strukturierter Inhalt
aus Texten
extrahiert werden?

**Strukturierte
Daten**

Wie kann ein
interoperabler
Repräsentations-
mechanismus
aussehen?

Text Mining

- Verwendet Technologien zur Verarbeitung natürlicher Sprache
- Zwei Paradigmen
 - Dokumentenretrieval
Spezialfall: Term Retrieval
 - Informationsextraktion



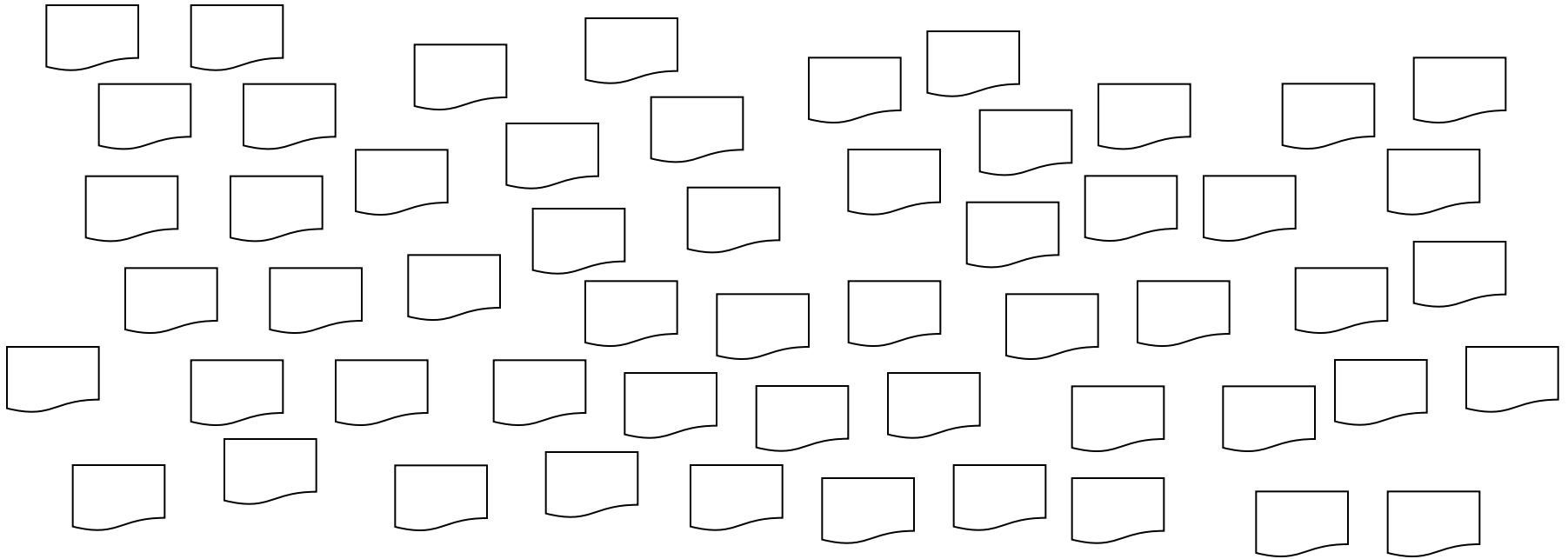
Erotide, 58 anos # HAS # Obesidade Pcte interna com hist de edema e dor em MID há 3 semanas. Refere que no início do quadro apresnetava hiperemia local importante, nega febre. Foi avaliada no posto de saúde e iniciado amoxicilina. Fez uso do ATB por 6 dias, com piora da dor, do edema, e surgimento de lesões arredondadas, planas, com bordos bem definido e pequeno ponto escurecido central. Consultou novamente no posto de saúde, snedo trocado o ATB para eritromicina, o qual usou por mais 6 dias, com piora das lesões e da dor. Na avaliação inicial a pcte apresentava edema e hiperemia importante de MID, bem como lesões ulceradas, necróticas com bordos bem definidos, sem secreção. Foi realizado ECO doppler que confirmou TVP em MID sendo então iniciada anticoagulação com enoxaheparina. Solicitada consultoria da dermato que realizou biopsia das lesões

AP: vasculite leucocitoclastica. A pcte recebeu ciprofloxacina por 5 dias e após 2 dias de oxacilina. Recebe alta em bom estado geral, com diminuição importante do edema e da dor em MID. Lesões em fase de cicatrização. Revisada a literatura: existe associação de vasculite em áreas de estase, bem como associação com farmacodermia. Em uso de: Varfarin 5mg 1cp por dia Captopril 25mg TID Amitriptilina 50mg Fluoxetina 40mg HCTZ 25 mg

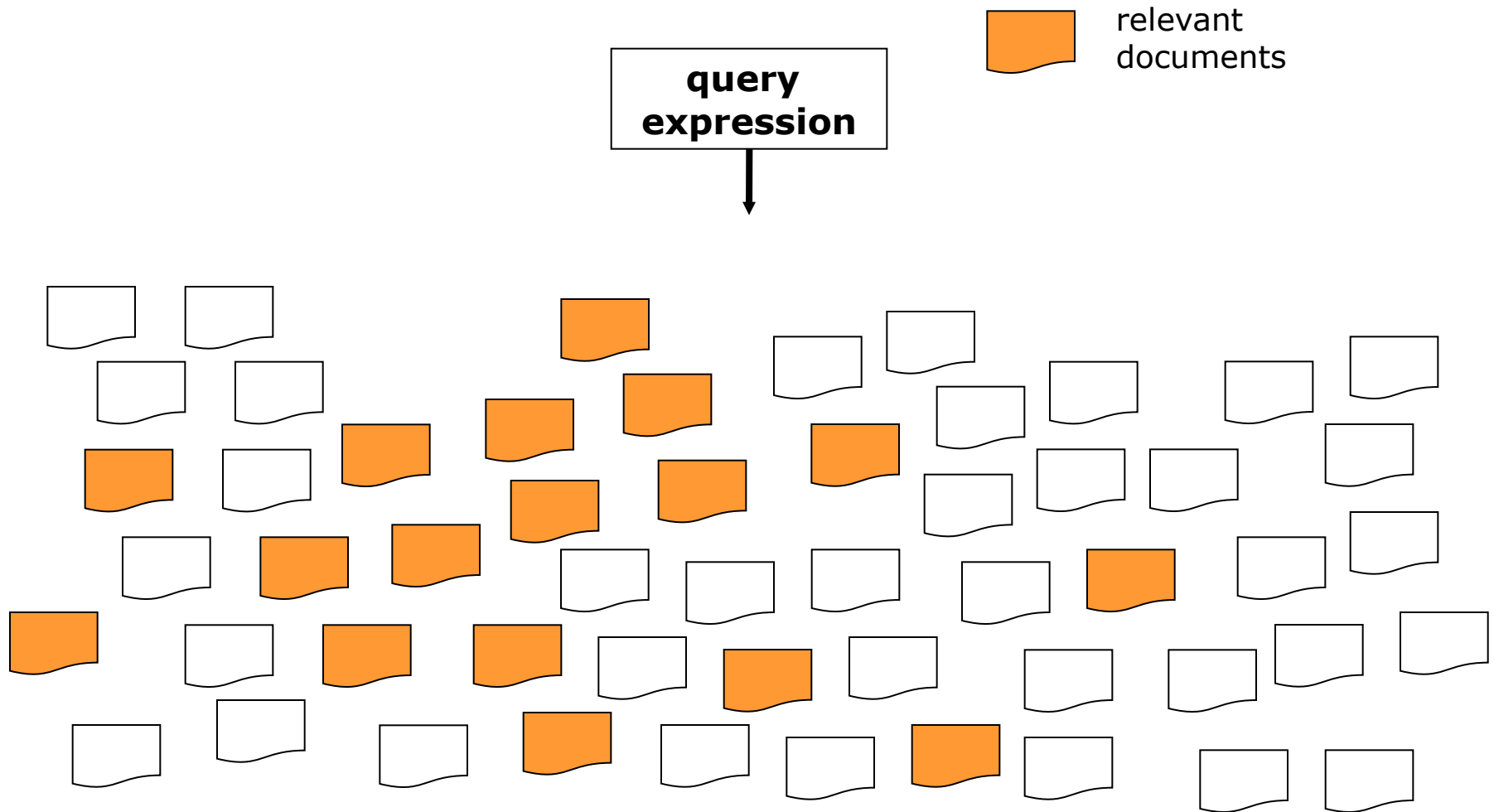
Paciente submetida a fundoplicatura videolaparoscópica. Recebe alta aceitando bem a via oral, sinais vitais estáveis, sem intercorrências ao longo da internação. Plano de retorno ambulatorial.

Dokumentenretrieval

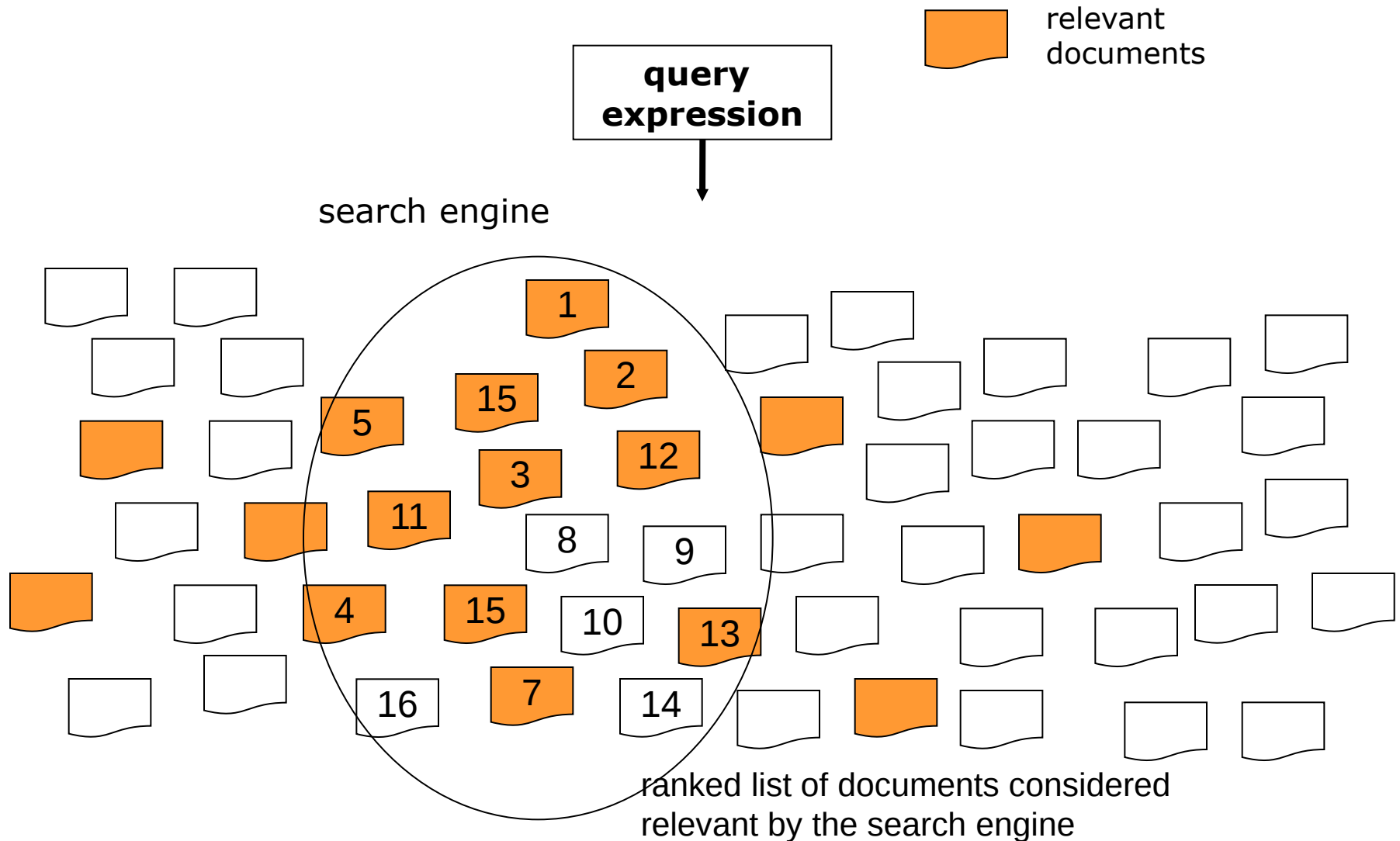
**query
expression**



Dokumentenretrieval



Dokumentenretrieval



AP: **vasculite leucocitoclastica**. A pcte recebeu ciprofloxacina por 5 dias e após 2 dias de oxacilina. Recebe alta em bom estado geral, com diminuição importante do edema e da dor em MID. Lesões em fase de cicatrização. Revisada a literatura: existe associação de vasculite em áreas de estase, bem como associação com

Termretrieval als Spezialfall des Dokumentenretrievals

AP: **vasculite leucocitoclastica**. A pcte recebeu ciprofloxacina por 5 dias e após 2 dias de oxacilina. Recebe alta em bom estado geral, com diminuição importante do edema e da dor em MID. Lesões em fase de cicatrização. Revisada a literatura: existe associação de vasculite em áreas de estase, bem como associação com

Terminologie / Ontologie

			1				
			5	2			
			3	4			
			7				

Information extraction: example

Anwendungsbeispiel Informationsextraktion: Tumordokumentation

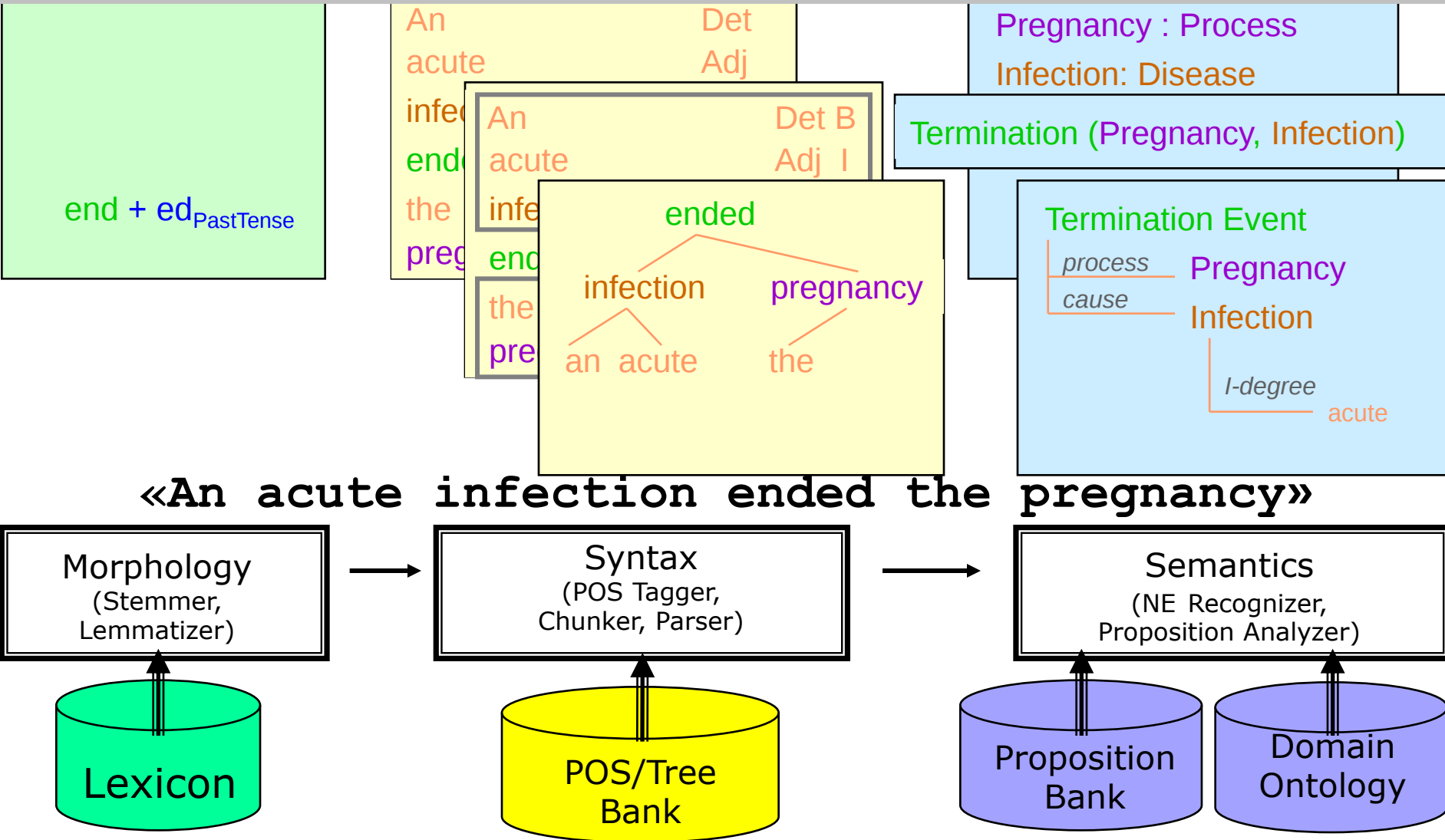
shadow was pointed out on a routine chest X-ray film, but she had no further examination. Physical examination on admission revealed purpura of the upper and lower extremities, swelling of the gums and tonsils, but no symptoms showing the complication of myasthenia gravis. Hematological tests revealed leucocytosis: WBC count 68 700/ μ 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 9.1×10^4 / μ l. Further laboratory examination revealed elevated serum lactic dehydrogenase (589 U/l), vitamin B₁₂ (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) (20 mg/m², days 1, 2, 3, 4, 6 and 8), cytosine arabinoside (Ara-C) (100 mg/m², days 1-9), 6MP-riboseide (6-MP) (70 mg/m², days 1-9) and prednisolone (PSL) (20 mg/m², 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 x 45 x 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 keratinazation 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		

Sprachtechnologien

- von einfachem Pattern Matching...
"do*" → "do", "dog", "done", "doctor", etc.
- zu aufwendigeren Methoden, unter Verwendung von
 - lexicales Wissen
 - grammatical Wissen
 - domain Wissen
 - Empirisches Wissen (e.g. **annotated corpora**)
 - Probabilistische Methoden
 - Methoden der künstlichen Intelligenz

NLP-“Pipeline”



Output sprachanalytischer Verfahren

- Nominalphrasen (Termkandidaten)
- Prädikat-Argument-Strukturen
- Erkennung und Klassifikation von Namen
- Anbindung von Präpositionalphrasen
- Skopus von Negationen
- Anaphernauflösung
- Diskursanalyse
- etc...

Echtes “Textverstehen” ist nach wie vor the “Heilige Gral”
der Computerlinguistik

Medizinsprachliche Herausforderungen

- Hohe lexikalische Produktivität
 - Nominalkomposita: "Penicillinallergieverdacht"
 - Akronyme und Abürzungen
- Syntaktische Eigenheiten
 - Telegrammstil
 - Aufzählungslisten (z.B. Laborwerte), Tabellen
- Qualitätsprobleme
 - Fortbestehende Fehler (Rechtschreibung, Zeichensetzung), meist durch Flüchtigkeit
- Aussagen durch lokale und Dokumentenkontexte geprägt
 - Ambiguitäten
 - implizite Annahmen

Semantische Anreicherung unter Verwendung von Sprachtechnologie

**Natürliche
Sprache**

Wie soll
strukturierter Inhalt
aus Texten
extrahiert werden?

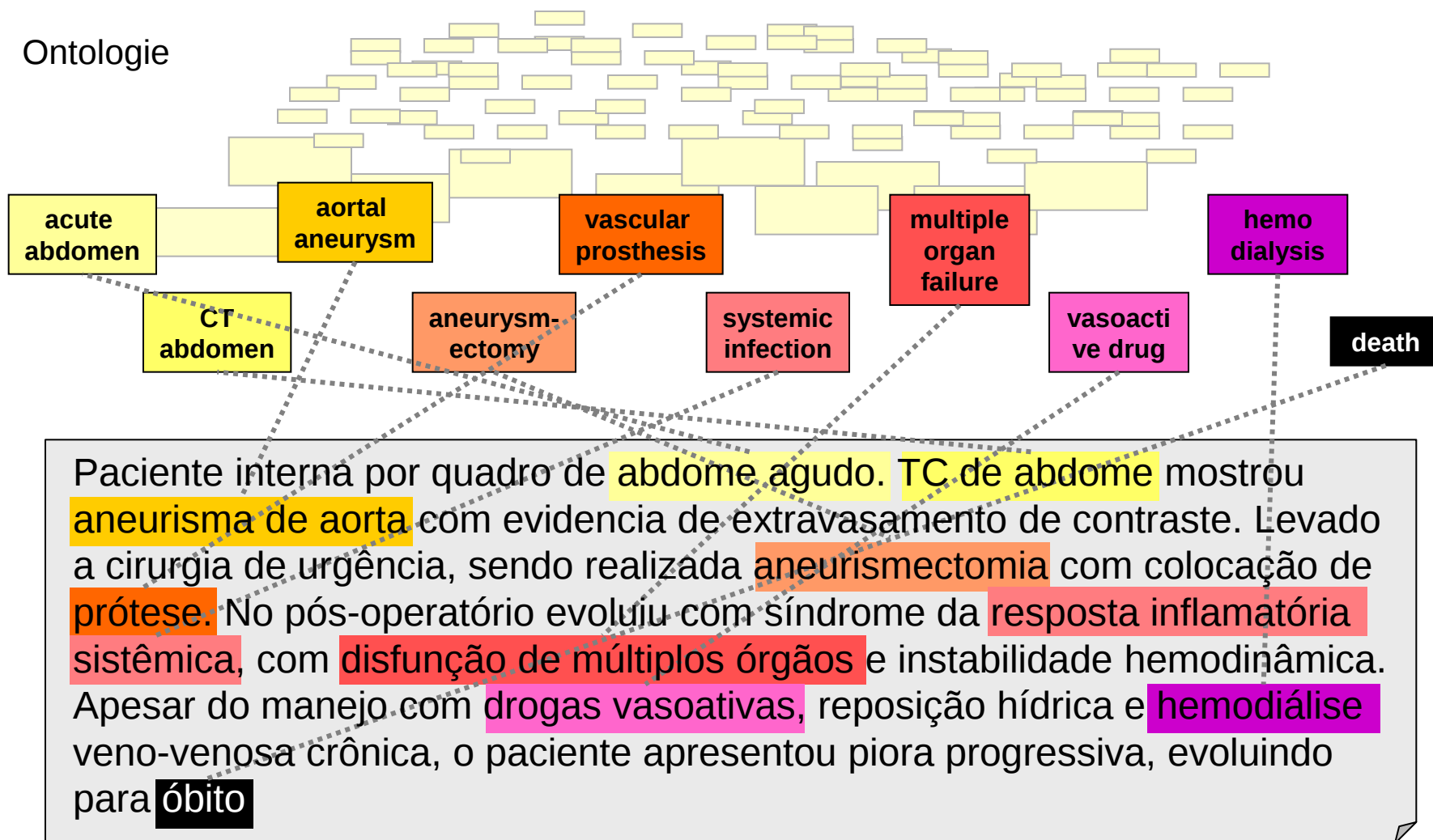
**Strukturierte
Daten**

Wie kann ein
interoperabler
Repräsentations-
mechanismus
aussehen?

Naiver Ansatz: Diskursobjekte entsprechen Ontologieinstanzen

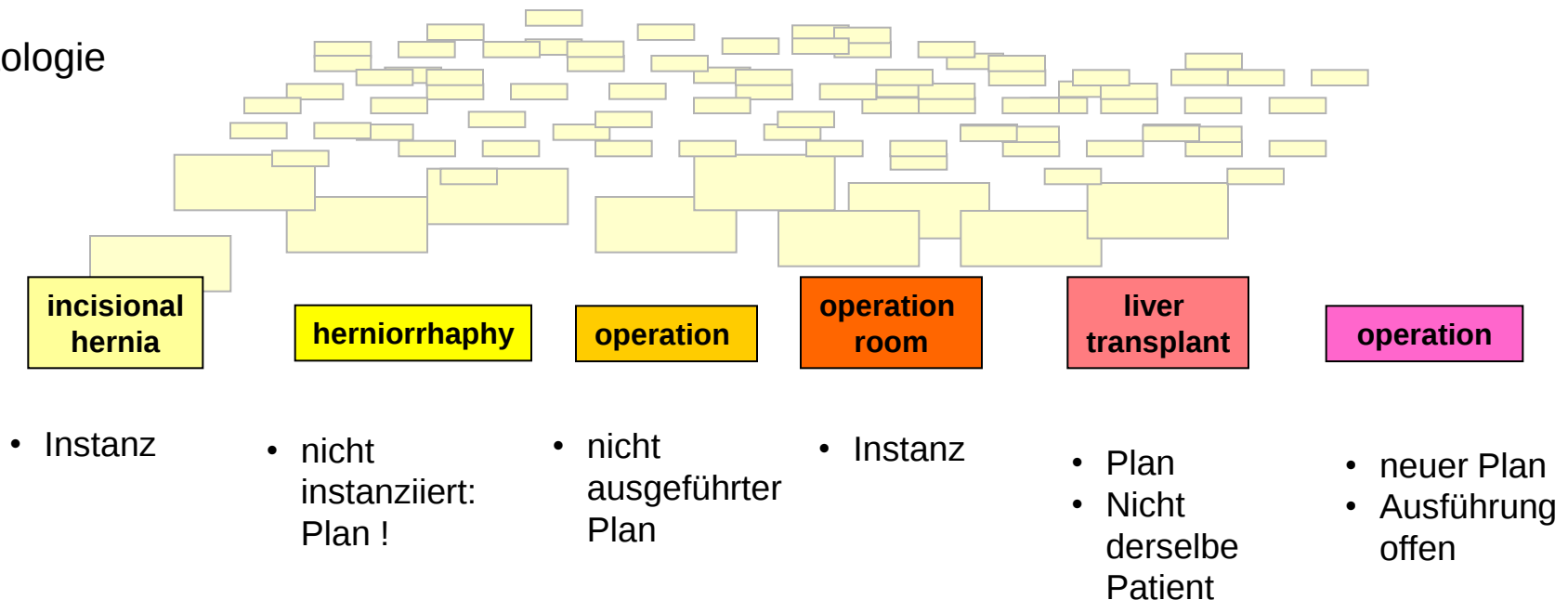
- Term T_i wird in einem Text identifiziert: Finde eine passende Ontologiekategorie mittels Termretrieval und Disambiguierung
- Interpretiere die Erwähnung von T_i als Referenz zu einer Instanz dieser Klasse
- d.h. für jedes T_i existiert ein reales Objekt

Beispiel 1



Beispiel 2

Ontologie



Patient with **incisional hernia** admitted for **herniorrhaphy**, but **operation** was suspended because **operation room** was urgently needed for **liver transplant**. Discharged with orientation and rescheduled **operation**

Warum Ontologien allein nicht ausreichen

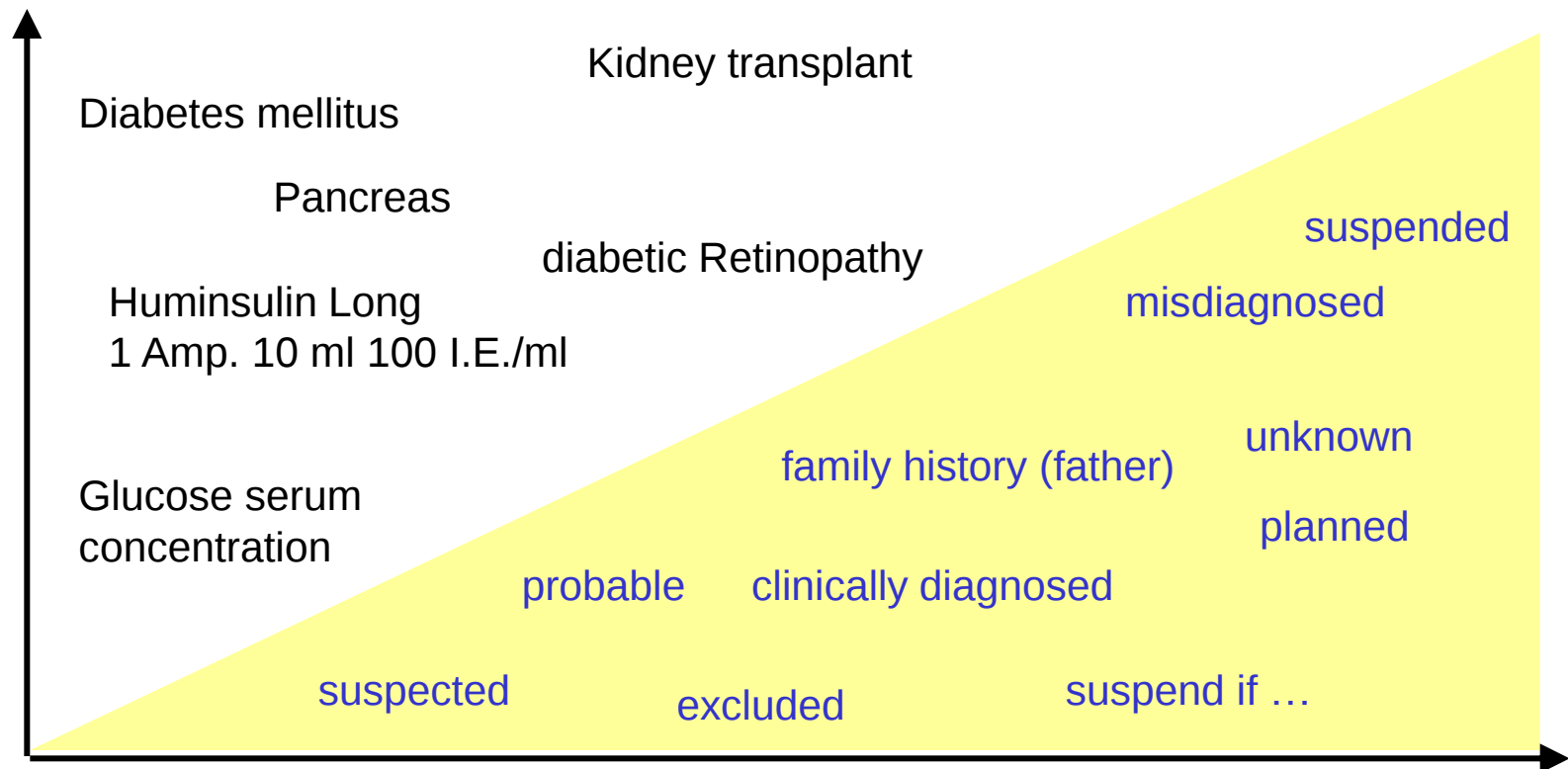
Unterschiedliche Kontexte:

- Bezugsperson:
 - “Vater war Diabetiker”
- Unsicherheit / Verdacht
 - “...kam zur Diabetesabklärung”
- Negation
 - “kein Anhalt eine Diabetes mellitus”
- Plan
 - “falls weiterhin ... , sollte ein Diabetes mellitus ausgeschlossen werden”

Grenzproblematik

Ontologie

“was ist”
Typen von
Gegenständen
beschrieben
anhand ihrer
generischen
Eigenschaften

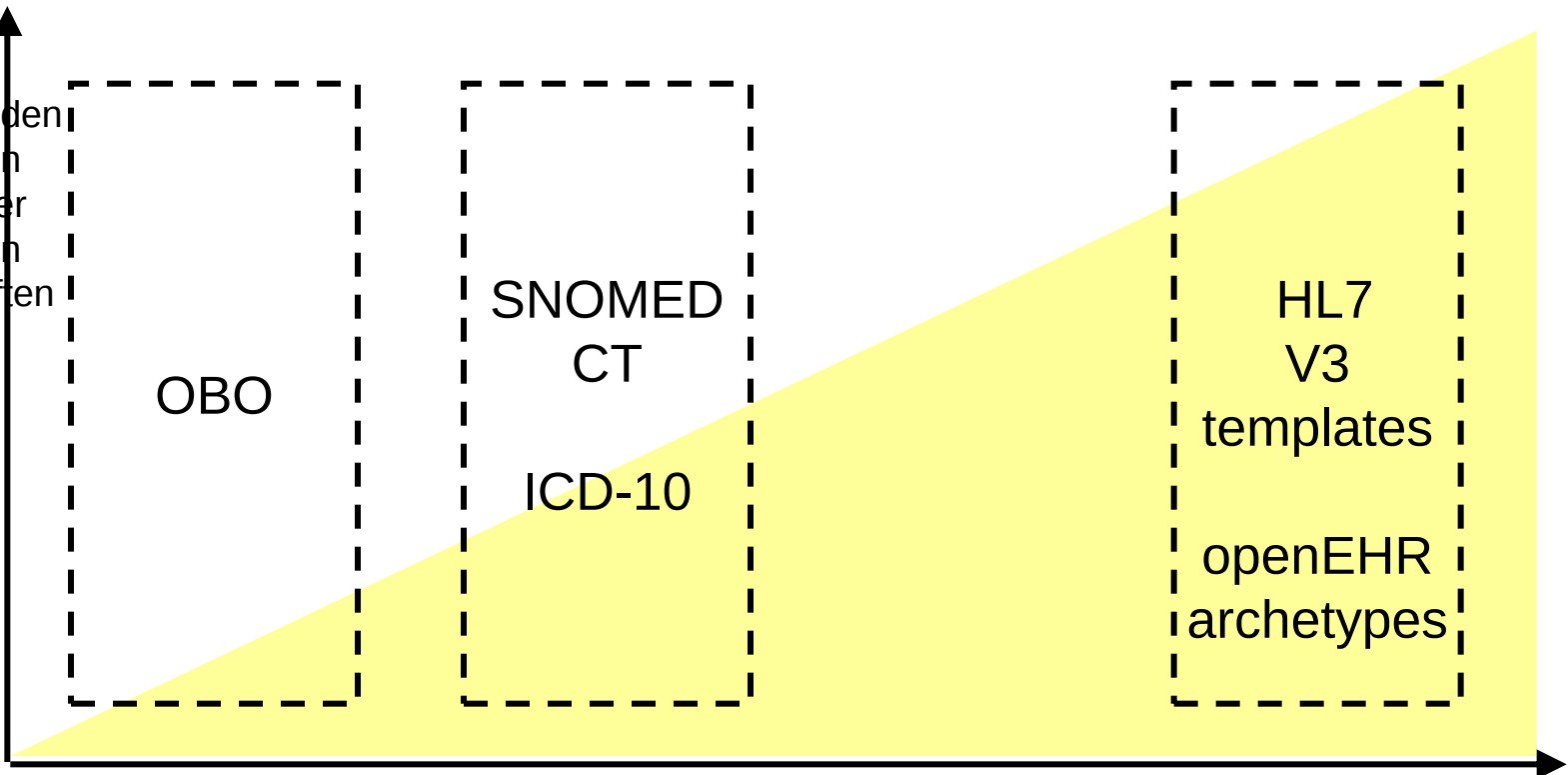


Epistemische Kontexte
Wissen, Verdacht, Vermutung,
Ausschluss...

Ontologie vs. Epistemologie

Ontologie

“was ist”
Typen von
Gegenständen
beschrieben
anhand ihrer
generischen
Eigenschaften



Epistemische Kontexte

Wissen, Verdacht, Vermutung,
Ausschluss...

Informationsmodelle

vermengen ontologische mit epistemischer Information

Diabetes Mellitus - Familienanamnese

ja nein unbekannt

Großvater

☐☐☐

Großmutter

☐☐☐

Vater

☐☐☐

Mutter

☐☐☐

Bruder /

☐☐☐

Schwester

Fazit - Repräsentation

- Klinische Freitexte kombinieren
 - Bezüge zu Typen / Klassen von Objekten (Ontology)
 - Bezüge zur Wertigkeit der Aussagen (Epistemologie)
- Wird der epistemische Kontext ignoriert: falsch positive
- Wird die Ontologie ignoriert: falsch negative
- “Bindung” von Ontologien an Informationmodelle:
 - Gegenstand aktueller Forschung
 - Pragmatische Vorschläge (TermInfo:
<http://www.hl7.org/v3ballot/html/welcome/environment/index.htm>)
- Weiterhin finden Terminologiesysteme Verwendung, wenn beides miteinander kombinieren

ICD

Version 2007

A16

Respiratory tuberculosis, not confirmed bacteriologically or histologically

Fazit - Textanalyse

- Für die semantische Anreicherung klinischer Freitexte insbesondere notwendig:
 - (fehlertolerante) Termerkennung, -matching, -disambiguierung
 - Kontexterkennung
- Zur Qualitätssteigerung
 - Evaluationsstandards
 - Clinical Coding Challenges

Fazit - Textanalyse

Computational Medicine Center

Medical NLP Challenge

[Home](#)[Details \(PDF\)](#)[Registration](#)[Results](#)[Organizers](#)[FAQ](#)

International Challenge: Classifying Clinical Free Text Using Natural Language Processing

Purpose

To challenge the international Natural Language Processing (NLP) research community to create and train computational intelligence algorithms that automate the assignment of [ICD-9-CM](#) codes to clinical free text.

Introduction

It is surprisingly hard for computers to handle free text as smoothly and effectively as humans do. So far, the results of the numerous efforts to achieve this have been mixed. Indeed, at times it has appeared that the complexities of free text are such as to render the effort futile. Not so; in fact, successive attempts to address the problem of converting free text into actionable knowledge have advanced the science of natural language processing and led to demand for software that simulates and complements what people are able to do.

We are sponsoring an international challenge task on the automated processing of clinical free text. Even with advances in structured vocabularies, many hospitals continue to electronically store some patient data as free text. This practice produces terabytes of information that, beyond the clinical visit, has limited utility because of its volume and accessibility. Natural language processing can potentially uncover implicit structure in this data, rendering it accessible to targeted search engines as well as special purpose systems dedicated to billing, quality assurance and discovery. This challenge offers participants an opportunity to test their untested algorithms or apply existing ones. Additionally, the Challenge provides full access to a carefully anonymised body of clinical data suitable for training and testing.

Competition Process

All participants will be required to [register](#). On 1 Feb 2007, participants will be given access to a training data set, which they will use to develop their algorithms.

The test data set will be made available on 1 Mar 2007. Participants will use their algorithms to process the test data, and will submit their results in XML format, along with a brief description of their methods. For complete details about the data formats and evaluation process, download the [Challenge Details](#) document.

Download Data Sets

With the challenge concluded, all data sets are now available for download at the [CMC Resource Catalog](#).

Important Dates

- 1 Dec 2006
 - Web site launched
- 22 Jan 2007
 - Registration begins
- 1 Feb 2007
 - Training data set released
- 28 Feb 2007
 - Registration ends
- 1 Mar 2007
 - Test data set released
- 18 Mar 2007
 - Last day to submit results
 - Due by midnight EDT
- 1 Apr 2007
 - Competition results announced

