Einführung in die Computerlinguistik und Sprachtechnologie

Vorlesung im WiSe 2018/19 (B-GSW-12)

Prof. Dr. Udo Hahn

Lehrstuhl für Computerlinguistik Institut für Germanistische Sprachwissenschaft Friedrich-Schiller-Universität Jena

http://www.julielab.de

Allgemeine Hinweise

- Vorlesung: Mi, 10-12h (Humboldt 8, SR 1)
- Übung zV: Fr, 8-10h (Fürstengrab. 1, SR 275)
 - beginnt am 19.10.
- Vorlesungsmaterialien im Netz
 - http://www.julielab.de/ ⇒ "Students"
- B-GSW-12 besteht aus VL+ÜB und Seminar!
- Sprechstunde: Mi, 12-13h (nA) (FG 30, R 004)
- Email: udo.hahn@uni-jena.de
- URL: http://www.julielab.de
- Fachliteratur ist überwiegend in Englisch

Bitte ...

• ... Handys/Smartphones ausschalten

 ... 90 Minuten ohne Mailund Tweet-Check <u>sind</u> möglich

"Digital detox"

... kein Picknick



Institut für Germanistische Sprachwissenschaft der FSU Jena

- Lehrstuhl für Theoretische Linguistik Grammatiktheorie
 - Prof. Dr. Peter Gallmann bzw. n.n.
- Lehrstuhl für Angewandte Linguistik Computerlinguistik
 - Prof. Dr. Udo Hahn
- Professur für Pragmatik
 - Prof. Dr. Pia Bergmann
- Professur für Phonetik & Sprechwissenschaft
 - Prof. Dr. Adrian Simpson
- Professur für Geschichte der deutschen Sprache
 - Prof. Dr. Eckhard Meineke

Computerlinguistik in Jena (1/2)

- Institutionell: Teil der Germanistischen Sprachwissenschaft
 - aber einzelsprachübergreifende Methodik
 - besondere Anwendungsdomänen:
 - Naturwissenschaften: Biologie + Medizin
 - Sozial- und Wirtschaftswissenschaft
 - Digital Humanities
- Integration in die Informatik:
 - Neben- bzw. Anwendungsfach für
 - B.Sc.: Informatik, Angewandte Informatik
 - M.Sc.: Informatik, Computational Science

Computerlinguistik in Jena (2/2)

- Aktive Forschergruppe
 - Lehrstuhl für Computerlinguistik = Jena University Language & Information Engineering (JULIE) Lab
 - Hohe internationale Visibilität (Publikationsdichte)
 - Deutsche Forschungsgemeinschaft (DFG)
 - Aktuell: (1/5) SFB 1076 AquaDiva Biodiversität in der Critical Zone
 - Aktuell: Graduiertenkolleg Modell ,Romantik' [Digital Humanities]
 - Bundesministerium f
 ür Bildung & Forschung (BMBF)
 - Aktuell: Nationale Förderinitiative "Systemmedizin" (J L AC)
 - Frühere Projekte: Forschungs-Cluster JenAge Nationaler Forschungskern, StemNet
 - Förderinitiativen der Europäischen Union
 - Frühere Projekte: MANTRA (SA), CALBC (SA), BOOTStrep (STREP), ..
- Ausgründung von Start-up-Firmen
 - Averbis, TexKnowlogy
- Jobs, Jobs, Jobs ... etwa als studentische Hilfskraft
- Themen, Themen, Themen ... BA- oder MA-Arbeit, Dissertation

Weitere Veranstaltungen

- Seminar zu B-GSW-12
 - SoSe 2019
- Vorlesung/Übung ASQ-DH
 - Einführung in Digital Humanities: Grundlagen der Informatisierung der Geisteswissenschaften
 - Di, 17-19, Humboldt 8, SR 3

Computer (und Menschen!) tun sich schwer mit Sprache(n) ...

Die pykka Sprache

 Güney pykka-i tassas pel Criftek ut pykka-e coggy pons Criftek

```
(1)
- coggy
Criftek
               (2)
               (1)
Günny
               (1)
- pel
               (1)
- pons
- pykka-i
               (1)
pykka-e
               (1)
- tassas
                (1)
                (1)
ut
```

Lexikografische Ordnung

> Häufigkeitszählung

Die pykka Sprache

- Günny pykka-i tassas pel Criftek ut pykka-e coggy pons Criftek
 - Perspektive des Computers/Menschen auf diese Äußerung:
 - <u>uninterpretierbare</u> Buchstaben-/Lautsequenz
 - Fehlt: Spezifikation von Wortbedeutung (Lexikon)
 - Fehlt: Regeln für Wortverknüpfung (Syntax)
 - Fehlt: Regeln für die Verbindung Syntax/Semantik
- Günny pykka-i tassas aus Criftek und pykka-e coggy nach Criftek
 - Pel → aus, ut → und, pons → nach
 - Lediglich ein Syntaxskelett

Die pykka Sprache

- Günny pykka-i tassas aus Criftek und pykkae coggy nach Criftek
- Deutsche Wortäquivalente:
 - { Deutschland, Costa-Rica }
 - { exportieren, importieren }
 - { Optoelektronik, Banane }
- Deutschland importier<u>t</u> Banane<u>n</u> aus Costa-Rica und exportier<u>t</u> Optoelektronik nach Costa-Rica

Von pykka ins Deutsche I

- Günny pykka-i tassas aus Criftek und pykkae coggy nach Criftek
- Deutsche Wortäquivalente:
 - [Deutschland = Günny, Costa-Rica = Criftek]
 - [importieren = pykka-i, exportieren = pykka-e]
 - [Banane(n) = tassa(s), Optoelektronik = coggy]
- Standard-Interpretation:

Deutschland importiert Bananen aus Costa-Rica und exportiert Optoelektronik nach Costa-Rica

Von pykka ins Deutsche II

- Günny pykka-i tassas aus Criftek und pykkae coggy nach Criftek
- Deutsche Wortäquivalente:
 - [Costa-Rica = Günny, Deutschland = Criftek]
 - [importieren = pykka-i, exportieren = pykka-e]
 - [Banane = tassas, Optoelektronik = coggy]
- Non-Standard-Interpretation:

Costa-Rica importier<u>t</u> Banane<u>n</u> aus Deutschland und exportier<u>t</u> Optoelektronik nach Deutschland

Konstituenten der Analyse/ Produktion natürlicher Sprache

- Inventar von Wörtern (Lexikon) und ihrer Bedeutungen (lexikalische Semantik)
- Verknüpfungsregeln für Wörter (Syntax)
- Kompositionelle Ableitung der Bedeutung eines Satzes (Satz-Semantik) aus den lexikalischen Bedeutungen der Wörter und der Syntaxstruktur (semantische Interpretation)
- Evaluation der semantischen Interpretation auf der Basis von Hintergrundwissen (Enzyklopädie, Alltagswissen usw.)

Computerlinguistik I

- Linguistik: Gegenstandsbereich sind (überwiegend) natürliche Sprachen
 - Deutsch, Englisch, Französisch, ...
- Beispiele für formale Sprachen

```
    L = {a<sup>n</sup>b<sup>n</sup>, n∈N}
    = {ab, aabb, aaabbb, aaaabbbb, ... }
```

- jede Programmiersprache, Auszeichnungssprache
 - JAVA, C++, ..., XML, HTML, ...
- jede Logik
 - Aussagenlogik, Prädikatenlogik, Typenlogik, ...
- Differentialgleichungen, Integrale, Vektoren, ...

Computerlinguistik II

- Beschreibungen und Formalisierungen entsprechen den Anforderungen, die sich aus der Verarbeitung durch Computer ergeben
 - keine natürlichsprachige Beschreibung (à la Duden oder Grammatik für Fremdsprachenerwerb), sondern formalisiert und damit explizit
 - explizite Spezifikation von Verfahrensbeschreibungen (Algorithmen), die von einer (abstrakten)
 Maschine ausgeführt werden können
 - Beachtung formaler (komplexitätstheoretischer)
 Eigenschaften der Beschreibung: Berechenbarkeit,
 Entscheidbarkeit, "Rechen-Kosten" (Zeit, Speicher)

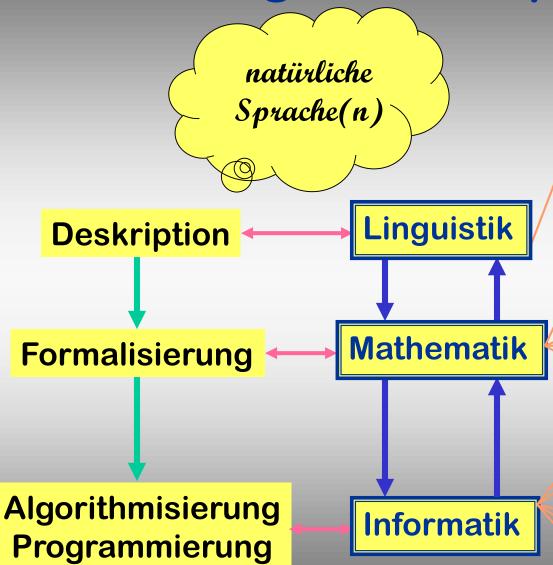
Computerlinguistik III

- Fundierung computerlinguistischer Beschreibungen durch Bezug auf theoretische und methodische Prinzipien der Linguistik und Informatik
 - Linguistische Grammatikmodelle vs. formale Grammatikmodelle der Informatik
 - Automatenmodelle der Informatik als Grundlage des Parsings natürlicher Sprache
 - Lexikonmodelle und Suchverfahren in Lexika
 - Semantische Repräsentationsformalismen vs.
 Wissensrepräsentationssprachen
 (Beschreibungslogik)

Computerlinguistik IV

- Realisierung dieser Beschreibungen durch ihre Implementation in einem natürlichsprachlichen (Teil-)System entsprechend informatischer Standards
 - Computerlinguistik ist keine naiv "programmierte" Linguistik
 - Programmiertechnologien (z.B. objekt-orientiert)
 - Daten(bank)technik (Speicher- und Zugriffsmethoden)
 - Software Engineering
 - Portierbarkeit (Domänenwechsel)
 - Wiederverwendbarkeit (Middleware: UIMA usw.)
 - Robustheit (NL ist ein sehr komplexes, nach wie vor nur partiell beschriebenes System)

Verortung der Computerlinauistik



Theoretische Linguistik

Phrasenstruktur-Grammatik
Dependenzgrammatik
Unifikationsgrammatik
Konstruktionsgrammatik
modelltheoretische Semantik
strukturelle Semantik
Frame-Semantik

Algebra

Formale Grammatiken & Sprachen Automatentheorie

Graphentheorie

Logik

Wahrscheinlichkeitstheorie Analysis, Numerik

Algorithmen &
Datenstrukturen
Programmierung
Informationssysteme
Künstliche Intelligenz

Maschinelles Lernen, Deduktionssysteme Wissensrepräsentation

Computerlinguistik-Standorte

www.ims.uni-stuttgart.de/info/SitesEurope.html#Germany



Computerlinguistik-Starbrücken

www.ims.uni-stuttgart.de/info/SitesEurope.html#Germany



U Stuttgart (3)

U Heidelberg (5)

RWTH Aachen

U München (2)

TU Darmstadt (4)

U Jena

U Tübingen (3)

U Bielefeld (4)

U Potsdam (2)

U Bremen

U Bochum (2)

U Erlangen-Nbg.

U Osnabrück (2)

U Hamburg (3)

KIT Karlsruhe

U Duisburg-Essen

U Leipzig

U Magdeburg

U Düsseldorf

U Gießen

U Hildesheim

U Koblenz

Computerlinguistik-Standorte

www.ims.uni-stuttgart.de/info/SitesEurope.html#Germany SCHLESWIG POMMERN Texttechnologie -Osnabrück **Digital** NORDRHEIN-WESTFALEN **Humanities** Kassel O HESSEN Informations-Wissenschaft / Information Retrieval Konstanz

U Stuttgart (3)

U Heidelberg (5)

RWTH Aachen

U München (2)

TU Darmstadt (4)

U Jena

U Tübingen (3)

U Bielefeld (4)

U Potsdam (2)

U Bremen

U Bochum (2)

U Erlangen-Nbg.

U Osnabrück (2)

U Hamburg (3)

KIT Karlsruhe

U Duisburg-Essen

U Leipzig

U Magdeburg

U Düsseldorf

U Gießen

U Hildesheim

U Koblenz

TU Darmstadt (2) U Frankfurt/M. (2)

U Leipzig

U Bamberg

U Köln

U Passau

U Jena

HU Berlin

U Stuttgart

U Konstanz

U Dortmund

U Kassel

U Würzburg

U Göttingen

U Münster

U Regensburg

J Hildesheim

U Düsseldorf

U Dortmund

BU Weimar

U Bamberg

U Kaiserslautern

TU Dresden

http://www.dig-hum.de/

Computerlinguistik-Starbrücken (6 37 (23 [25]

www.ims.uni-stuttgart.de/info/SitesEurope.html#Germany SCHLESWIG POMMERN Texttechnologie -Osnabrück **Digital** NORDRHEIN-WESTFALEN **Humanities** Kassel 0 Informations-Wissenschaft / Information Retrieval Konstanz

U Stuttgart (3)

U Heidelberg (5)

RWTH Aachen

U München (2)

TU Darmstadt (4)

U Jena

U Tübingen (3)

U Bielefeld (4)

U Potsdam (2)

U Bremen

U Bochum (2)

U Erlangen-Nbg.

U Osnabrück (2)

U Hamburg (3)

KIT Karlsruhe

U Duisburg-Essen

U Leipzig

U Magdeburg

U Düsseldorf

U Gießen

U Hildesheim

U Koblenz

TU Darmstadt (2) U Frankfurt/M. (2)

U Leipzig

U Bamberg

U Köln

U Passau

U Jena

HU Berlin

U Stuttgart

U Konstanz

U Dortmund

U Kassel

U Würzburg

U Göttingen

U Münster

BU Weimar

U Bamberg

U Kaiserslautern

TU Dresden

Warum Compute Malufe?

(auch mit einer Antwor

A section of the control of the cont

Wirtschaft

Industrie zeigt leichte Ermüdungs

Unternehmen können 2006 die guts Wachstumsnate wohl nicht halt

em RERLIN 30 Dezenber Die drug
dag samwinstandelab Wechsten, der
daß desenvolle auf zu benannteln
son zu erweit er auf 15 heren vone
Kantinusfahren geleichte
Auf 20 dezen der der
Auf 20 dezen der der der der
daß der Verhandere der
Kantinusfahren
Kantinusfahren
Auf 20 dezen die der des desen
ger den dewen felle undelen vorlie der
delen vorlie sieden der der
delen vorlie sieden der der
deren
deren

Joseph Met. (18) Joseph Operationsbericht

**resident, De Wei

**r

0190,0000,rIm Januar 1994 Sympathektomie rechtsseitig w PROMETER IN rechten Großzeh in Form eines ulcus perforans

And the state of t

Behaarung vor. Eine Fistelöffnung ist nicht auz rechten Großzehe abgeheiltes ulcus perforans.

Banking wat was Das Pedigewebe uninagert uch Duransack und se state was das Pedigewebe uninagert uch Duransack und se state was das Pedigewebes mit dem CUSA, E strang se state was de pedige abore abore abore consistenz, es wirkt fester. Die Nervenwurzeln sind fest mit dem Gewebe verbacken. Im intraduralen Abschnitt was de pedige se pe weiter disseziert. Duranaht, schichtweiser Wundverschluß.

Diagnose: Spina bifida occulta mit großflächigem Lipom epifascial lumbosakral, das sich nach intraspinal, intradural

Bewerbung als Speditionskaufmann - Nachricht Ed Senden
□ Senden< Datei Bearbeiten Ansicht Einfügen Format Extras Verfassen ? Nachricht Diese Nachricht wurde noch nicht gesendet. Zur Nachverfolgung vor oder am Mittwoch, 6. Februar 2002 17:00. wallner@holzhauer.de mruf@gmx.net Bewerbung als Speditionskaufmann Sehr geehrte Frau Wallner, in der "Welt" vom vergangenen Wochenende suchen Sie einen s lich mächte mich mich auf die Stelle bewerben und mich Ihnen v

Form vorstellen:

Nach meinem Wehrdienst machte ich bei der Firma Schneller ei Speditionskaufmann. Zur Zeit arbeite ich als Vertriebsmitarbeiter Innendienst bei Schreibwaren, möchte aber aus privaten Gründen gerne in diesem Jahr nach Be Da ich bereits weitreichende Erfahrungen in der Logistik sammeln konnte, würde Position in Ihrer Firma sehr reizen.

lich hoffe, damit ihr Interesse geweckt zu haben. Meine ausführliche Bewerbungs Sie in den nächsten Tagen per Post

Mit freundlichen Grüßen

Michael Ruf

lebenslauf_michael_ruf,

Michael Ruf

Waldstraße 16 60357 Frankfurt

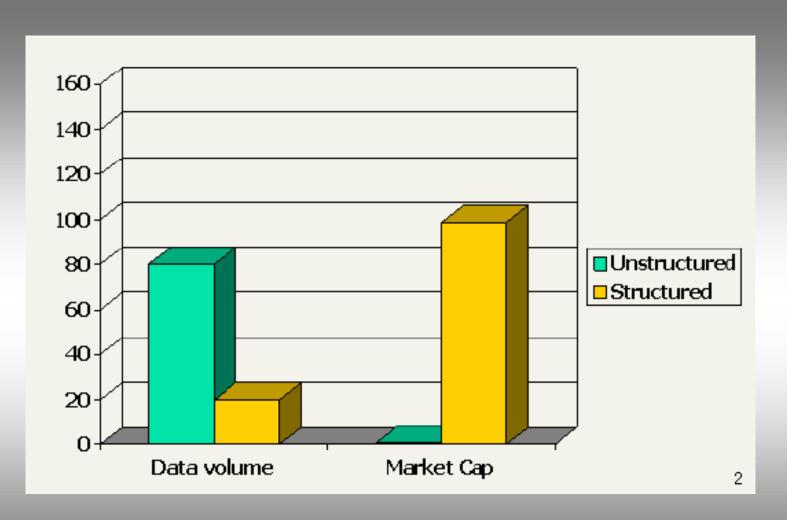
Tel: 069 - 13 45 87 55



Klug & Cummings

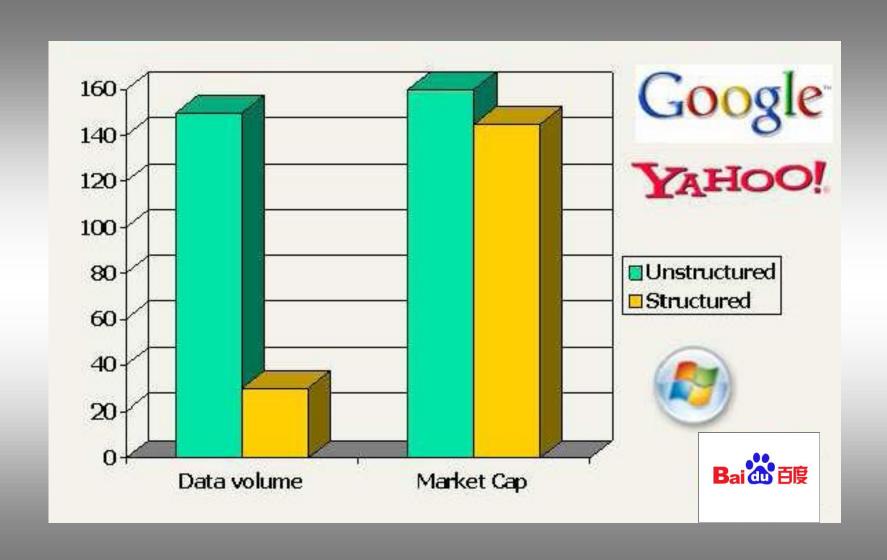
8

Structured vs. Unstructured Data (1996)



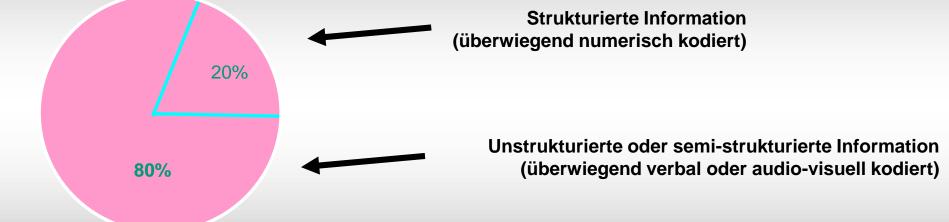
Source: Prabhakar & Raghavan, Verity (2002)

Structured vs. Unstructured Data (2006)



Die zunehmende Bedeutung von Sprachtechnologie

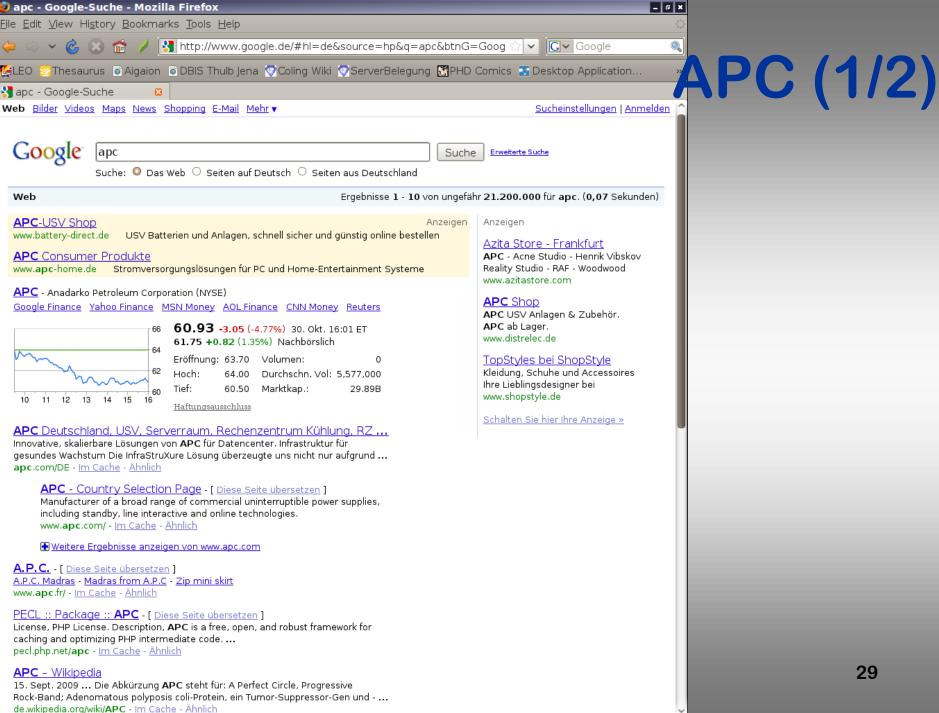
80:20 Argument



Quelle: Oracle Corp.

Sprachtechnologische Meilensteine

- Suchmaschinen
 - Google (Yahoo!, Bing, Baidu ...)
- Natürlichsprachliche Assistenten
 - Alltagsplanung: Siri, Alexa
 - Fragebeantwortung: Watson, Ask (Jeeves)
- Speech-to-Text-Systeme
 - Dragon Dictate, ...
- Übersetzungssysteme
 - Google Translate, Microsoft Translator, Language Weaver, Lernout & Houspie, DeepL DeepL ...
- Linguistisch sensitive Schreibprogramme
 - · Word, ...





Twease helps you find precisely what you're looking for, quickly. Search Medline™ abstracts in new ways and find the information you need. Type your search phrase here and click "Search" to get started, or, take a tour of Twease with our Tutorial. Search apc Search Order results by C Relevance Publication Date C Oldest First

Activated protein C ligation of apoer2 (LRP8) causes Dab1 dependent signaling in U937 cells. (2009/01/07)

of activated protein C (APC) to cells triggers multiple ... breakdown . One paradigm for APC s signaling emphasizes its binding ... apoer 2) dependent signaling by APC and found that APC initiated ... by APC and found that APC initiated rapid phosphorylation of Tyr ... Ser 473 in Akt . APC also induced phosphorylation of Ser ... receptor family members , inhibited APC induced 19116273 phosphorylation of Dab1 and ... apoer2 by using sirna ablated APC induced Dab1 phosphorylation , suggesting ... , suggesting that RAP sensitive APC induced signaling requires apper 2 resonance equilibrium binding studies, APC bound with high affinity to ... lipoprotein receptor. RAP blocked APC binding to sapper but not ... of U937 cells to immobilized APC . RAP also blocked APC ... APC , RAP also blocked APC s ability to inhibit endotoxin ... that ligation of apoer2 by APC signals via Dab1 phosphorylation and ... gsk3beta , thereby contributing to APC s beneficial effects on cells

Kinetics of major histocompatibility class I antigen presentation in acute infection, (2009/01/06)

between the size of the APC population and duration of Ag In both infections , APC production 19124733 was calculated to occur ... rate of dendritic cells . APC production was found to closely ... , in influenza infection, APC production ceased between the third ... self limiting infections generate the APC necessary to drive T cell **Related Articles**

[APC = antigen-presenting cells]

Argon plasma coagulation therapy, (2009/01/05)

in argon plasma coagulation (APC) therapy in gastroenterology gastroenterology . recent findings: APC has gained importance in gastrointestinal ... the last decade, and APC technology 19122503 has improved over time ... series , have shown that APC can be used effectively in ... are low . summary: APC can be considered as a ... trials including a comparison of APC to other treatment modalities as ... long term follow up after APC treatment are required to clarify ... clarify the relative value of APC in gastroenterology .

Related Articles

Related Articles

19121182 Sterility screening of platelet concentrates : questioning the optimal test strategy. (2009/01/05) of apheresis platelet concentrates (APC) and pooled platelet concentrates **Related Articles**

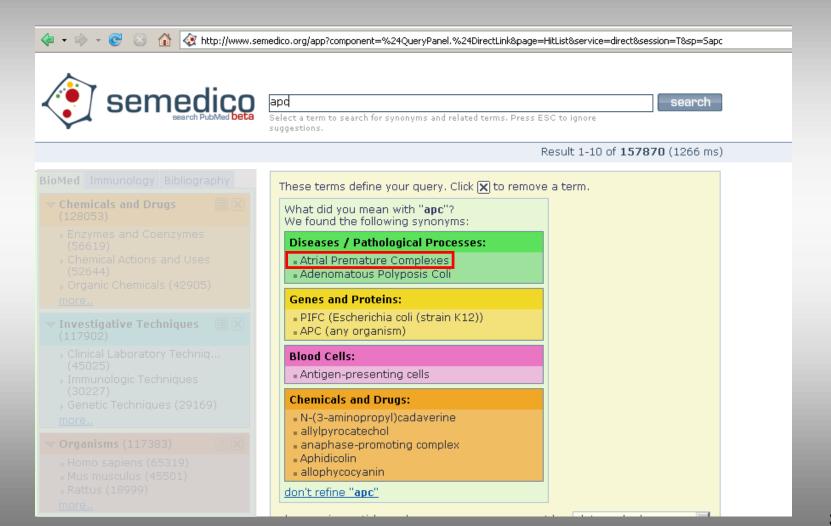
-3∙

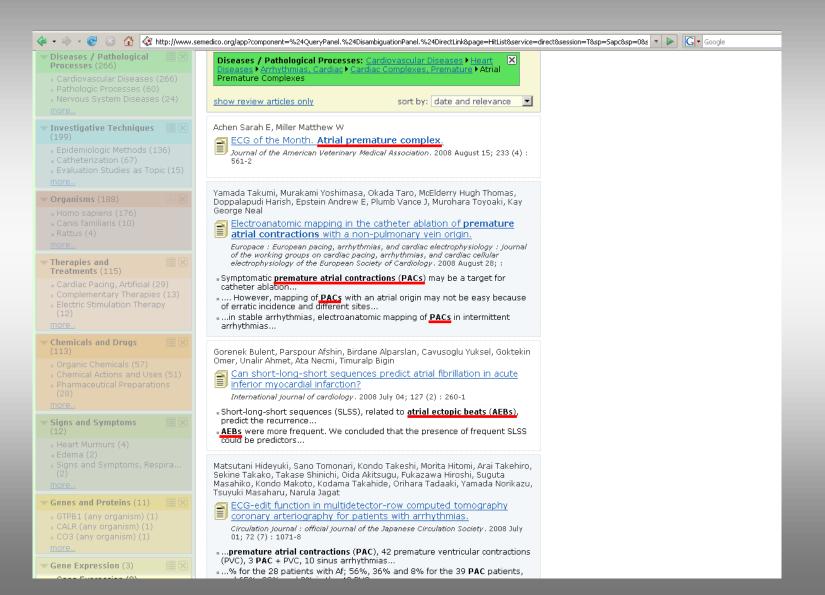
Semantische Suchmaschine Semedico

- Ermöglicht einen tieferen, semantischontologisch strukturierten Zugang zu biomedizinischen Fachtexten
- semantisch
 - Automatische Erkennung von Entitäten (=Begriffe, Konzepte)
 - Abbildung von Synonymen und Abkürzungen auf Konzept-Identifikatoren

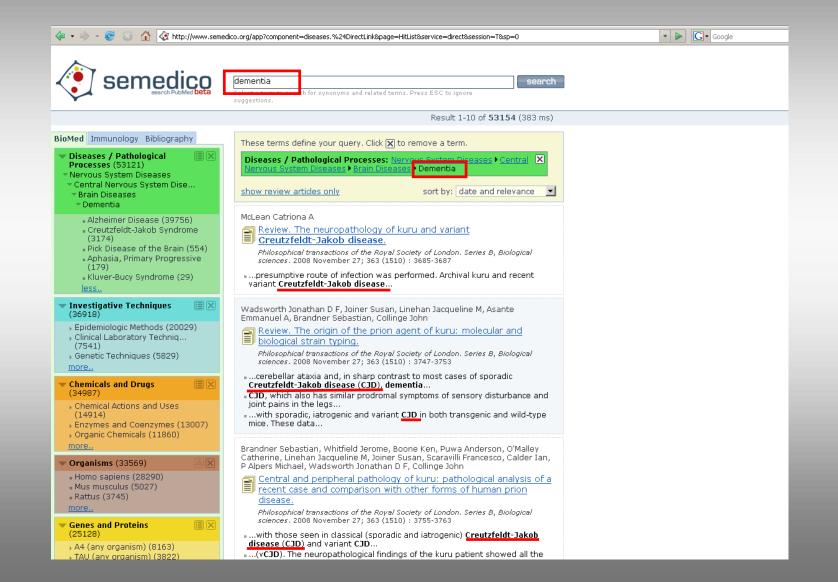




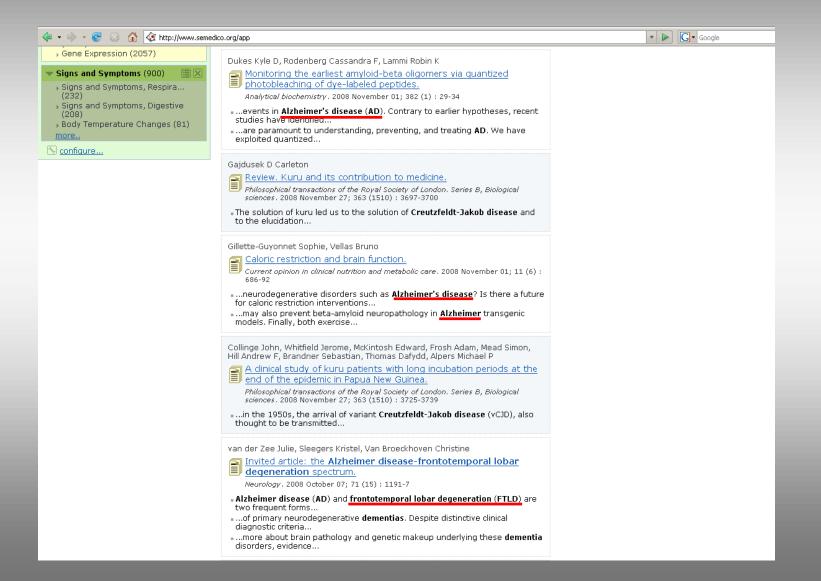


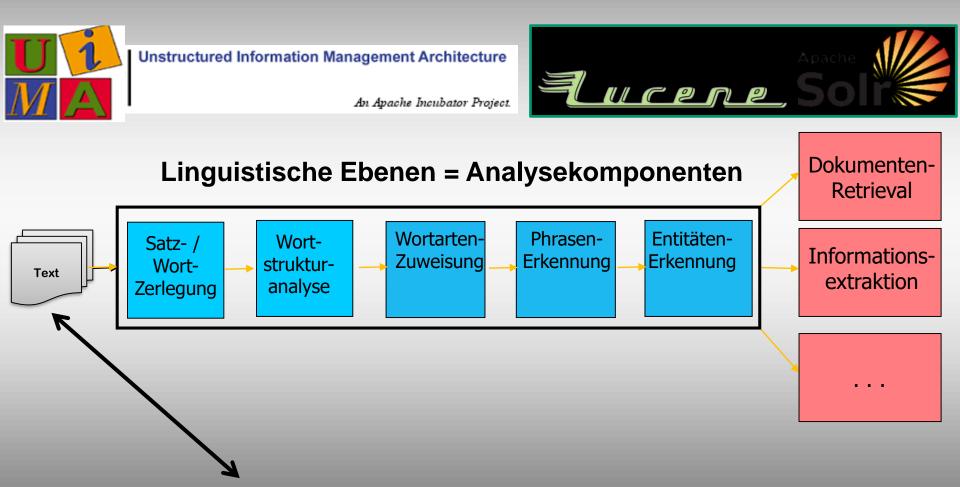


- Ermöglicht einen tieferen, semantischontologisch strukturierten Zugang zu Medline-Texten
- semantisch
 - Automatische Erkennung von Entitäten (=Begriffe, Konzepte)
 - Abbildung von Synonymen und Abkürzungen auf Konzept-Identifikatoren
- ontologisch
 - Taxonomien werden bei der Suche mit einbezogen
 - Ontologisch motivierte Facetten ermöglichen Browsing bei Suche

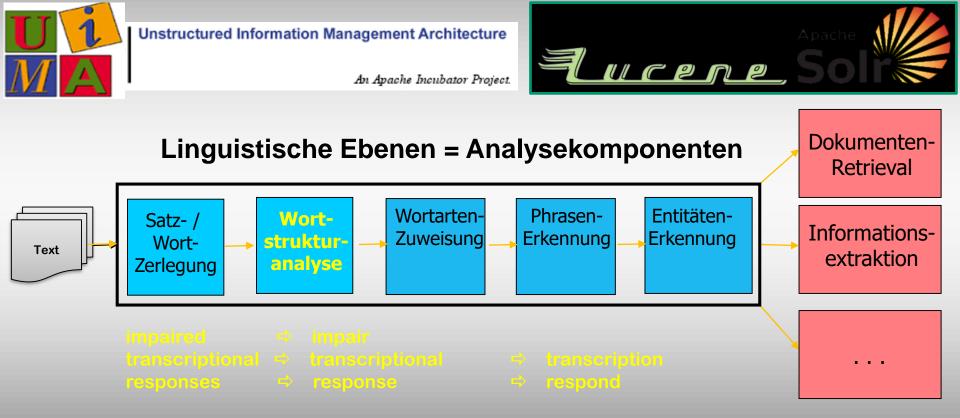




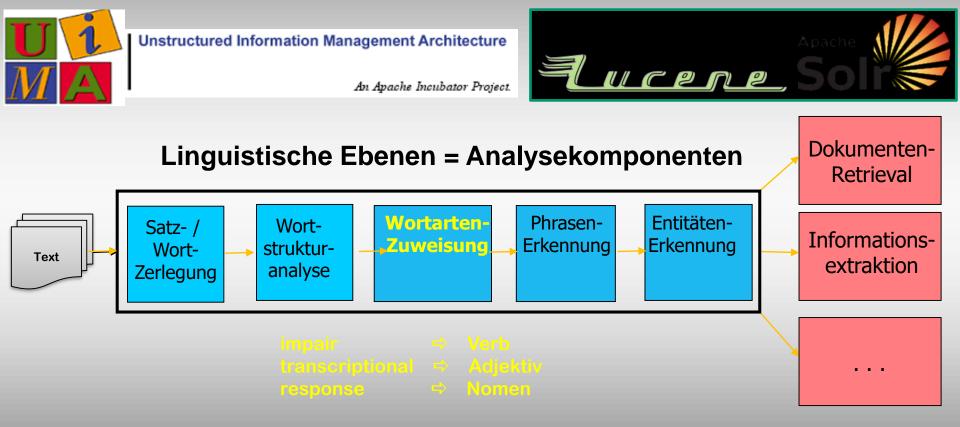




... may explain impaired transcriptional responses in leishmania-infected cells



... may explain impaired transcriptional responses in leishmania-infected cells



... may explain impaired transcriptional responses in leishmania-infected cells



Unstructured Information Management Architecture

An Apache Incubator Project.



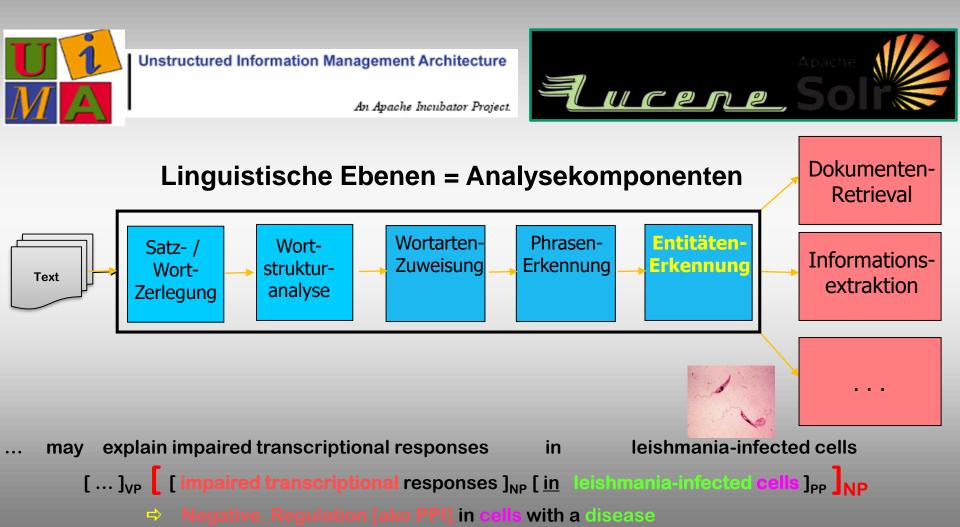


.. may explain impaired transcriptional responses .. MODAL VERB PARTIZIP ADJEKTIV NOMEN

[may explain]_{VP} [[impaired transcriptional responses]_{NP}

in leishmania-infected cells PRÄP PARTIZIP NOMEN

[in leishmania-infected cells]_{PP}]



Fazit: Semedico Suchmaschine

- Ermöglicht einen tieferen semantischontologischen Zugang zu Medline-Texten
- Derzeit sind > 28 Millionen Medline-Dokumente (Abstracts) und 1,8 Mio.
 Volltexte prozessiert und indexiert
- Nutzerstudie an der Medizinischen Hochschule Hannover (Proc. SMC, 2007)
- http://www.semedico.org
- E.Faessler & U. Hahn (2017), ACL Demos.
- Youtube-Demo: http://www.youtube.com/watch?v=f7llPqGh-sc

Natürliche Sprache Ambiguität: lexikalisch-semantisch

- Homografie, Polysemie
 - Konstanz liegt am Bodensee.
 - Bei Konstanz des Luftdrucks ...

- I saw that gasoline can explode
 - [Ich sah diesen Benzinbehälter explodieren]
 - [Ich sah, dass Benzin explodieren kann]

Natürliche Sprache Ambiguität: syntaktisch

Skopus

- die alten Männer und Frauen
 - die alten Männer und [allgemein alle] Frauen
 - die alten Männer und alten Frauen

PP-Anbindung

- Sie sahen den Mann mit dem Fernrohr
 - Sie sahen den Mann mit Hilfe ihres_{INSTRUM} Fernrohrs
 - Sie sahen den Mann, der sein_{POSSESS} Fernrohr trug

Natürliche Sprache Ambiguität: syntaktisch

- PP-Anbindung & Homografie/Polysemie
 - They saw the man with the telescope.

Ambiguität: syntaktisch

- PP-Anbindung & Homografie/Polysemie
 - They <u>saw</u> the man with the telescope.
 - Sie sahen den Mann mit Hilfe ihres_{INSTRUM} Fernrohrs.

Ambiguität: syntaktisch

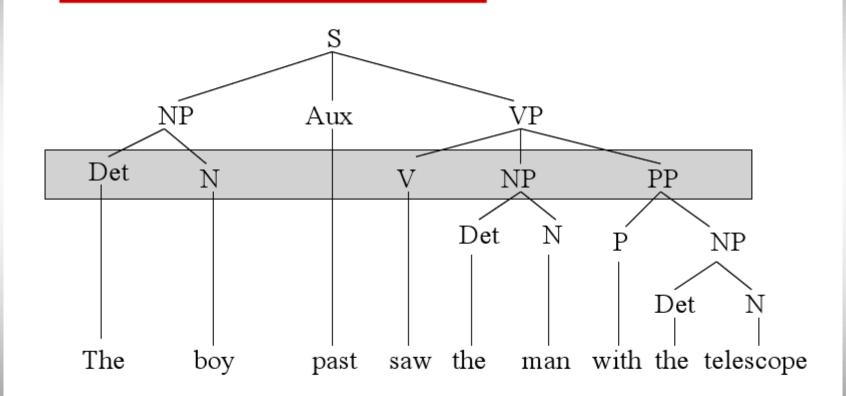
- PP-Anbindung & Homografie/Polysemie
 - They <u>saw</u> the man with the telescope.
 - Sie sahen den Mann mit Hilfe ihres_{INSTRUM} Fernrohrs.
 - Sie sahen den Mann, der sein_{POSSESS} Fernrohr trug.

Ambiguität: syntaktisch

- PP-Anbindung & Homografie/Polysemie
 - They saw the man with the telescope.
 - Sie sahen den Mann mit Hilfe ihres_{INSTRUM} Fernrohrs.
 - Sie sahen den Mann, der sein_{POSSESS} Fernrohr trug.
 - Sie zersägen den Mann mit Hilfe ihres_{INSTRUM} Fernrohrs.
 - Sie zersägen den Mann, der sein_{POSSESS} Fernrohr trug.

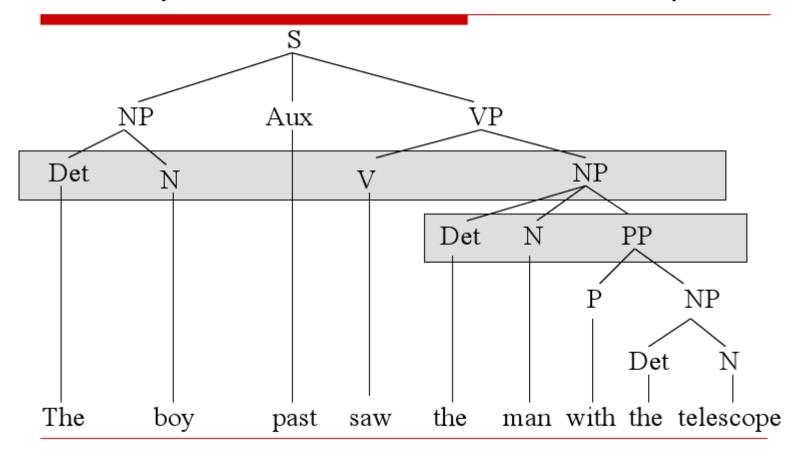
Structural Ambiguity (1)

The boy saw the man with the telescope



Structural Ambiguity (2)

The boy saw the man with the telescope



Natürliche Sprache Ambiguität: syntaktisch

- Wachstumsverhalten v. PP-Anbindungen
 - Die Tarifparteien haben sich [auf der Basis eines Kompromissentwurfs des neutralen Schlichters]₁ [am frühen Mittwoch Morgen]₂ [in Bad Nauheim]₃ [nach einem 16-stündigen Verhandlungsmarathon]₄ [auf einen Tarifabschluss]₅ [in Höhe]₆ [von 1,2 Prozent]₇ [für die 160.000 Beschäftigten]₈ [in der hessischen Bekleidungsindustrie]₉ geeinigt.

Ambiguität: syntaktisch

- Wachstumsverhalten v. PP-Anbindungen
 - Die Tarifparteien haben sich [auf der Basis eines Kompromissentwurfs des neutralen

Schlick [in Bac Verhar Tarifak [für die hessis

$$CAT_{n} = \binom{2n}{n} - \binom{2n}{n-1}$$

$$= \frac{(2n)!}{n! \times (2n-n)!} - \frac{(2n)!}{(n-1)! \times (2n-(n-1))!}$$

Ambiguität: syntaktisch

- Wachstumsverhalten
 - Die Tarifparteien hab CAT_{n=9-1} = 1430 sis eines Kompromisse

Schlick [in Bac Verhar Tarifak [für die hessis

$$CAT_{n} = {2n \choose n} - {2n \choose n-1}$$

$$= \frac{(2n)!}{n! \times (2n-n)!} - \frac{(2n)!}{(n-1)! \times (2n-(n-1))!}$$

Binärbäume !!