

Linguistische Informatik

Einführung

Gerhard Heyer
Universität Leipzig
heyer@informatik.uni-leipzig.de

UNIVERSITÄT LEIPZIG

Institut für Informatik



Informatik und Linguistik

Informatik als Werkzeug oder Theoriegrundlage

Linguistik *mit* dem Computer: Unterstützung der Linguisten bei der praktischen Arbeit

Linguistische Informatik: Anwendung von Konzepten und Verfahren der Informatik *für* die Analyse von Sprache (als System oder Einzelsprache)

- formale Sprachen, Parsing, Semantik und Verifikation
- informationstheoretische Konzepte, graphentheoretische Modellierung

Zwei Traditionen der Linguistischen Informatik

Computerlinguistik

- **Orientierung an kognitiven Modellen des Sprachverstehens**
- **Grundlage: formale Sprachen und regelbasierte Repräsentationen**
- **Ziel ist die Simulation sprachlicher Prozesse auf dem Computer**

Institutionell meist an philologischen Fakultäten angesiedelt,
sehr populär in den 90er Jahren, aber
keine wesentlichen theoretischen oder praktischen Erfolge

Zwei Traditionen der Linguistischen Informatik

Sprachprodukttechnologie

- **Orientierung an empirischen und ingenieurwissenschaftlichen Verfahren der Sprachbeschreibung**
- **Grundlage: Textdatenbanken, IR, informationstheoretische Modelle, graphbasierte Modellierungen**
- **Ziel ist die computerunterstützte Nutzung von Sprache als wesentliches Mittel des Verstehens und der Kommunikation**

**Institutionell Teil der Angewandten Informatik,
starker Auftrieb mit wachsender Bedeutung des Internet**

Anwendungen der Linguistischen Informatik

Informatik *für* Sprachanwendungen

Weder Sprachverstehen noch Sprachproduktion sind bisher gelöst. Jedoch besitzen wir eine Vielzahl von *Verfahren*, die zwar kein Verstehen ermöglichen, aber für viele Anwendungen oft völlig ausreichen.

flache Ansätze (effizient und robust)

statistische Methoden, Mustergrammatiken

tiefe Ansätze (präzise und korrekt)

linguistische Prinzipien, Constraints oder komplexe Regelwerke

Technologische Grundlagen

- ***Mustererkennung*** für die Spracherkennung und -generierung (speech to text / text to speech), Sprechererkennung
- ***Texttechnologien*** für die Textverarbeitung und Übersetzung
- ***Information Retrieval*** für die Suche und Informationsextraktion

Wesentliche Einsatzgebiete

- Sprachprodukte für das **Informations- und Wissensmanagement**
- Sprachprodukte für die **Dokumentenproduktion und -verwaltung**
- Sprachprodukte für die **Mensch-Maschine Kommunikation**
- Sprachprodukte für die **Mensch-Mensch Kommunikation**

Wichtige Trends

- Integration von verschiedenen **Medien und Modi** in der Kommunikation zwischen Menschen bzw. Mensch und Maschine
- Verarbeitung von großen Mengen **multimedialer, multimodaler und multilingualer** Information
- Schrittweise Entwicklung zu **inhaltsbezogenem Umgang** mit Information (*semantische* statt syntaktische Methoden)

Speech Technologies

- **Voice Control Systems**
- **Diktiersysteme**
- **Text-to-Speech Systems (Vorlesesysteme)**
- **Sprecher Identifikations- und Verifikationssysteme**
- **Spoken Information Access**
- **Speech Translation Systems**
- **Dialog Systeme in gesprochener Sprache**
- **... ..**

Texttechnologien

- **Rechtschreibprüfung und Textkorrektur**
- **Grammatikprüfung und Stilometrie**
- **computergestützte Übersetzung (*translation memories*, Terminologiesysteme)**
- **vollautomatische Vorübersetzung**
- **hochwertige vollautomatische Übersetzung**
- **Text Generierung**
- **... ..**

Information Retrieval Technologien

- **Textzusammenfassung**
- **Textklassifikation**
- **Cross-lingual Information Retrieval**
- **Information Extraction**
- **Concept-Based Information Retrieval**
- **Query expansion mit Thesauern, Taxonomien und Ontologien**
-

Beispiel: Ortsnamenserkennung

79312 **Emmendingen**

DV - Expect GmbH LABORSYSTEME

Herbolzheim Belchenstr. 44

HRB 662 K - 06. November 2001

Neueintragungen

HRB 662 K -- 06. November 2001: DV – Expect GmbH LABORSYSTEME in **Herbolzheim (Belchenstr. 44)**. Gesellschaft mit beschränkter Haftung. Gesellschaftsvertrag vom 28. August 2001, geändert in § 1 (Firma) am 09. Oktober 2001. Gegenstand des Unternehmens ist die Montage von medizinischen Geräten und Elektronikteilen und deren Vertrieb, auch im Lohnauftrag für Unternehmen der Gerätemedizin und der Bauindustrie, die Endmontage von Bauelementen, der Vertrieb von Ersatzteilen für Personen- und Güterbeförderungsunternehmen sowie der ambulante Handel und Vertrieb mit Fast-food Produkten. Stammkapital: 25 000,-- EUR. Die Gesellschaft hat einen oder mehrere Geschäftsführer. Ist nur ein Geschäftsführer bestellt, vertritt er die Gesellschaft allein. Sind mehrere Geschäftsführer bestellt, wird die Gesellschaft durch zwei Geschäftsführer gemeinschaftlich oder durch einen Geschäftsführer in Gemeinschaft mit einem Prokuristen vertreten. Einzelvertretungsbefugnis und Befreiung von den Beschränkungen des § 181 BGB kann erteilt werden. Geschäftsführer sind Vidko Umek, geb. am 7. März 1951, **Herbolzheim**, und Dolores Umek, geb. am 8. August 1979, **Herbolzheim**; diese sind einzelvertretungsberechtigt und befugt, die Gesellschaft bei der Vornahme von Rechtsgeschäften mit sich selbst oder als Vertreter eines Dritten uneingeschränkt zu vertreten (§ 181 BGB). Nicht eingetragen: Die Gesellschafterin Dolores Umek hat ihre Einlage in Höhe von 12 500,-- EUR durch Einbringung ihres PKW Audi A3 TDI, amtl. Kennzeichen EM-D 2299, Schätzwert: 14 600,-- EUR, erbracht; im übrigen wird auf die eingereichten Unterlagen Bezug genommen. Bekanntmachungen der Gesellschaft erfolgen im Bundesanzeiger.

Beispiel: Erkennung rechtsrelevanten Vokabulars

79312 Emmendingen

DV - Expect GmbH LABORSYSTEME

Herbolzheim Belchenstr. 44

HRB 662 K - 06. November 2001

Neueintragungen

HRB 662 K -- 06. November 2001: DV – Expect GmbH LABORSYSTEME in Herbolzheim (Belchenstr. 44). **Gesellschaft mit beschränkter Haftung. Gesellschaftsvertrag** vom 28. August 2001, geändert in § 1 (**Firma**) am 09. Oktober 2001. **Gegenstand** des **Unternehmens** ist die Montage von medizinischen Geräten und Elektronikteilen und deren Vertrieb, auch im Lohnauftrag für **Unternehmen** der Gerätemedizin und der Bauindustrie, die Endmontage von Bauelementen, der Vertrieb von Ersatzteilen für Personen- und Güterbeförderungsunternehmen sowie der ambulante Handel und Vertrieb mit Fast-food Produkten. **Stammkapital**: 25 000,-- EUR. Die **Gesellschaft** hat einen oder mehrere **Geschäftsführer**. Ist nur ein **Geschäftsführer bestellt, vertritt** er die Gesellschaft allein. Sind mehrere **Geschäftsführer bestellt**, wird die **Gesellschaft** durch zwei **Geschäftsführer gemeinschaftlich** oder durch einen **Geschäftsführer** in Gemeinschaft mit einem **Prokuristen vertreten. Einzelvertretungsbefugnis** und **Befreiung** von den **Beschränkungen** des § 181 BGB kann erteilt werden. **Geschäftsführer** sind Vidko Umek, geb. am 7. März 1951, Herbolzheim, und Dolores Umek, geb. am 8. August 1979, Herbolzheim; diese sind **einzelvertretungsberechtigt** und **befugt**, die **Gesellschaft** bei der **Vornahme von Rechtsgeschäften** mit sich selbst oder als **Vertreter eines Dritten uneingeschränkt zu vertreten** (§ 181 BGB). Nicht **eingetragen**: Die **Gesellschafterin** Dolores Umek hat ihre **Einlage** in Höhe von 12 500,-- EUR durch **Einbringung** ihres PKW Audi A3 TDI, amtl. Kennzeichen EM-D 2299, Schätzwert: 14 600,-- EUR, erbracht; im übrigen wird auf die eingereichten Unterlagen Bezug genommen. **Bekanntmachungen** der **Gesellschaft** erfolgen im **Bundesanzeiger**.

Beispiel: Erkennung von Firmenzweck

79312 Emmendingen

DV - Expect GmbH LABORSYSTEME

Herbolzheim Belchenstr. 44

HRB 662 K - 06. November 2001

Neueintragungen

HRB 662 K -- 06. November 2001: DV – Expect GmbH **LABORSYSTEME** in Herbolzheim (Belchenstr. 44). Gesellschaft mit beschränkter Haftung. Gesellschaftsvertrag vom 28. August 2001, geändert in § 1 (Firma) am 09. Oktober 2001. Gegenstand des Unternehmens ist die **Montage** von **medizinischen** Geräten und **Elektronikteilen** und deren Vertrieb, auch im Lohnauftrag für Unternehmen der **Gerätemedizin** und der **Bauindustrie**, die **Endmontage** von **Bauelementen**, der Vertrieb von Ersatzteilen für Personen- und **Güterbeförderungsunternehmen** sowie der ambulante Handel und Vertrieb mit **Fast-food Produkten**. Stammkapital: 25 000,-- EUR. Die Gesellschaft hat einen oder mehrere Geschäftsführer. Ist nur ein Geschäftsführer bestellt, vertritt er die Gesellschaft allein. Sind mehrere Geschäftsführer bestellt, wird die Gesellschaft durch zwei Geschäftsführer gemeinschaftlich oder durch einen Geschäftsführer in Gemeinschaft mit einem Prokuristen vertreten. Einzelvertretungsbefugnis und Befreiung von den Beschränkungen des § 181 BGB kann erteilt werden. Geschäftsführer sind Vidko Umek, geb. am 7. März 1951, Herbolzheim, und Dolores Umek, geb. am 8. August 1979, Herbolzheim; diese sind einzelvertretungsberechtigt und befugt, die Gesellschaft bei der Vornahme von Rechtsgeschäften mit sich selbst oder als Vertreter eines Dritten uneingeschränkt zu vertreten (§ 181 BGB). Nicht eingetragen: Die Gesellschafterin Dolores Umek hat ihre Einlage in Höhe von 12 500,-- EUR durch Einbringung ihres PKW Audi A3 TDI, amtl. Kennzeichen EM-D 2299, Schätzwert: 14 600,-- EUR, erbracht; im übrigen wird auf die eingereichten Unterlagen Bezug genommen. Bekanntmachungen der Gesellschaft erfolgen im Bundesanzeiger.

Beispiel: Erkennung numerischer Angaben (Zeit, Geld, §§, HR-Nr.)

79312 Emmendingen

DV - Expect GmbH LABORSYSTEME

Herbolzheim Belchenstr. 44

HRB 662 K - 06. November 2001

Neueintragungen

HRB 662 K -- 06. November 2001: DV – Expect GmbH LABORSYSTEME in Herbolzheim (Belchenstr. 44). Gesellschaft mit beschränkter Haftung. Gesellschaftsvertrag vom **28. August 2001**, geändert in § 1 (Firma) am **09. Oktober 2001**.

Gegenstand des Unternehmens ist die Montage von medizinischen Geräten und Elektronikteilen und deren Vertrieb, auch im Lohnauftrag für Unternehmen der Gerätemedizin und der Bauindustrie, die Endmontage von Bauelementen, der Vertrieb von Ersatzteilen für Personen- und Güterbeförderungsunternehmen sowie der ambulante Handel und Vertrieb mit Fast-food Produkten. Stammkapital: **25 000,-- EUR**. Die Gesellschaft hat einen oder mehrere Geschäftsführer. Ist nur ein

Geschäftsführer bestellt, vertritt er die Gesellschaft allein. Sind mehrere Geschäftsführer bestellt, wird die Gesellschaft durch zwei Geschäftsführer gemeinschaftlich oder durch einen Geschäftsführer in Gemeinschaft mit einem Prokuristen vertreten.

Einzelvertretungsbefugnis und Befreiung von den Beschränkungen des § 181 BGB kann erteilt werden. Geschäftsführer sind Vidko Umek, geb. am **7. März 1951**, Herbolzheim, und Dolores Umek, geb. am **8. August 1979**, Herbolzheim; diese sind einzelvertretungsberechtigt und befugt, die Gesellschaft bei der Vornahme von Rechtsgeschäften mit sich selbst oder als Vertreter eines Dritten uneingeschränkt zu vertreten (**§ 181 BGB**). Nicht eingetragen: Die Gesellschafterin Dolores Umek hat ihre Einlage in Höhe von **12 500,-- EUR** durch Einbringung ihres PKW Audi A3 TDI, amtl. Kennzeichen **EM-D 2299**,

Schätzwert: **14 600,-- EUR**, erbracht; im übrigen wird auf die eingereichten Unterlagen Bezug genommen.



Bekanntmachungen der Gesellschaft erfolgen im Bundesanzeiger.

Beispiel Mensch-Maschine Kommunikation

Google
Jetzt sprechen



Beispiel Mensch-Maschine Kommunikation




Durch die Nutzung unserer Dienste erklären Sie sich damit einverstanden, dass wir Cookies setzen. [Mehr erfahren](#)

WEB-ERGEBNISSE


Was ist Informatik denn nun wirklich? - Joint Bachelor of Arts ...
www.hier-studiert-julia.de/was-ist-informatik.html
Informatiker mit einem Uni-Abschluss sein heißt vor allem, mit anderen Menschen zu kommunizieren, häufig aus anderen Fachdisziplinen. Und Informatik heißt ...

Was ein Informatiker eigentlich so macht | MINT-Blog
mint.studieren-mit-meerwert.de/2015/04/informatik-studium
1. Apr. 2015 ... Ja, was macht so ein Informatiker eigentlich? Oftmals die Computer von Verwandten reparieren ;) Aber natürlich ist das nicht die eigentliche ...

Kapitel 1: Einführung Was ist Informatik
formal.tti.kit.edu/~beckert/teaching/InformatikIM-WS0506/iim01.pdf
B. Beckert: Informatik für IM – Universität Koblenz-Landau – WS 2005/2006. Kapitel 1: Einführung. Was ist Informatik? Begriff und Grundprobleme der. Informatik.

Was ist Informatik? - Die Informatik und ihre Teilgebiete - YouTube
de.ask.com/youtube?q=was+ist+informatik&v=u_9y7LS2mGA
19. Sept. 2016 ... Willst du wissen was Informatik ist, und welche Teilgebiete sie beherbergt? In diesem Video ist genau das einfach und zusammengefasst ...

Was ist Informatik ... für die Gesellschaft? - EDOC HU
edoc.hu-berlin.de/oa/reports/reJfJpzzqDM/PDF/23AJv1brKxPg.pdf
Informatik tritt in der gesellschaftlichen Wahrnehmung in zwei Varianten auf. ... These 1: Informatik ist eine technische Wissenschaft geworden - mit sichtbaren.

Informatik - Wer wie was? - YouTube
de.ask.com/youtube?q=was+ist+informatik&v=M20L26Nh3us
2. Sept. 2008 ... Ein etwas anderer Blick auf die Informatik Image Film zum Thema Informatik und Informatikstudium an der Friedrich-Alexander-Universität ...

Informatik Studium im Überblick - Inhalte, Voraussetzungen ...
www.informatik-studium.org
Informatik-Studium.org ist Ihr unabhängiger Wegweiser für das Studium der Informatik. Alles zu Inhalten, Voraussetzungen, Prüfungen.

Informatik-Ingenieurwesen - TUHH
www.tuhh.de/tuhh/studium/studienangebot/auslaufende-studiengaenge/informatik-ingenieur...
Für alle diese Aufgaben werden einerseits Methoden, Verfahren und Werkzeuge eingesetzt, die den Strukturwissenschaften Mathematik und Informatik ...

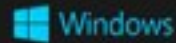
92.12 WasInfDid - Informatik in der Oberstufe
www.oberstufeninformatik.de/WasInfDid.pdf
Informatik Spektrum, 16,1, 3-10, Feb 1993. 92.12.17. Was ist Informatik-Didaktik? Gedanken über die Fachkenntnisse des Informatiklehrers. J. Nievergelt.

Antworten | Video

ENTDECKEN SIE ANTWORTEN ZU

- informatik** ausbildung
- praktikum **informatik**
- wettbewerb **informatik** biber
- informatik** lehrstellen
- schulpraktikum **informatik**
- informatik** schulen
- finanz **informatik**
- informatik** gehalt
- Informatik**
- Was ist **Informatik**
- r **Informatik** e.V
- r **Informatik**

Beispiel Mensch-Mensch Kommunikation



Skype Translator preview



Kostenlos ★★★★★ 11

Herausgegeben von Skype
(C) Microsoft Corp., 2014

Kategorie: Produktivität
Ungefähre Größe: 13,9 MB
Altersfreigabe: 12+

Beziehen von Windows 8.1 zum
Ausführen dieser App [Weitere
Informationen](#)

Weitere Apps nach Skype



Skype WiFi
Kostenlos
★★★★★ 338



Skype
Kostenlos
★★★★★ 61.156



Beschreibung

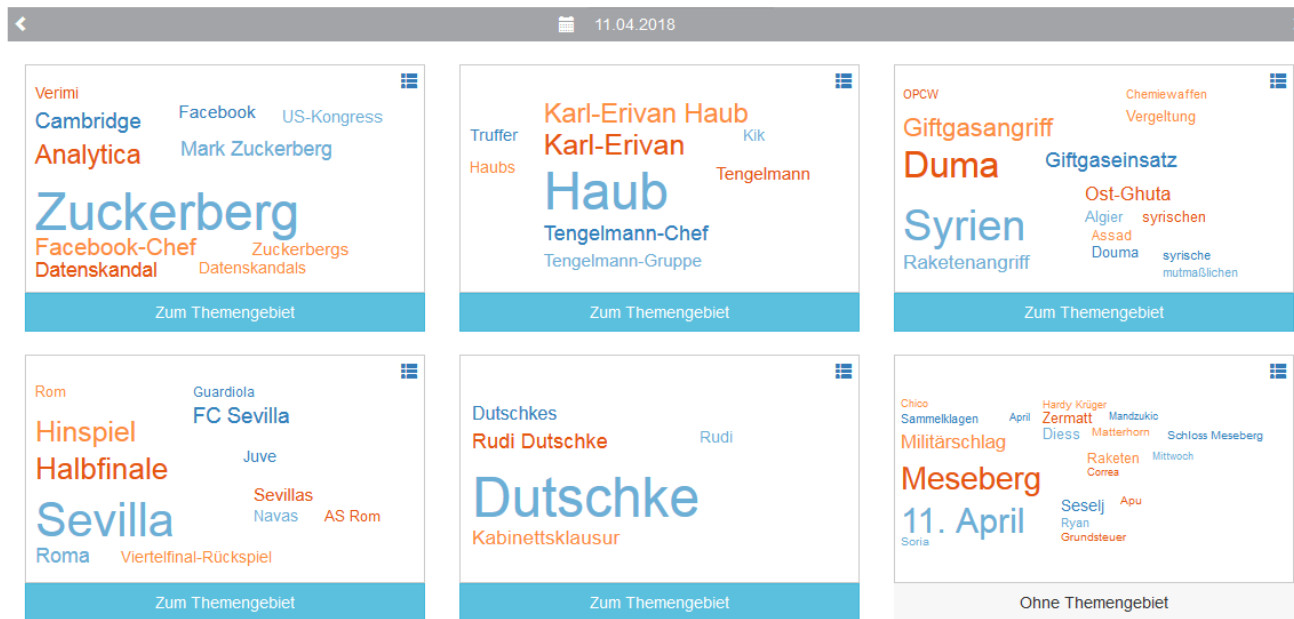
Gestalten Sie die Zukunft von Skype mit, indem Sie Ihre Video- und Sprachanrufe direkt übersetzen lassen – mit dieser Vorshow von Skype Translator.

- Ihre Einzelgespräch- und Videoanrufe werden direkt in beide Richtungen übersetzt.
- Wählen Sie, ob Sie die Übersetzung anhören oder nur eine Mitschrift anzeigen lassen möchten.
- Alle regulären Funktionen von Skype für Modern Windows sind enthalten.
- Stellen Sie sicher, dass alle Beteiligten ein Headset verwenden, um die Anrufqualität zu verbessern.
- Jeder Anruf hilft, die Leistung von Skype Translator zu verbessern.
- Wenn Sie aktuell Skype für Modern Windows verwenden, deinstallieren Sie es zuerst.

Beispiel Informations- und Wissensmanagement

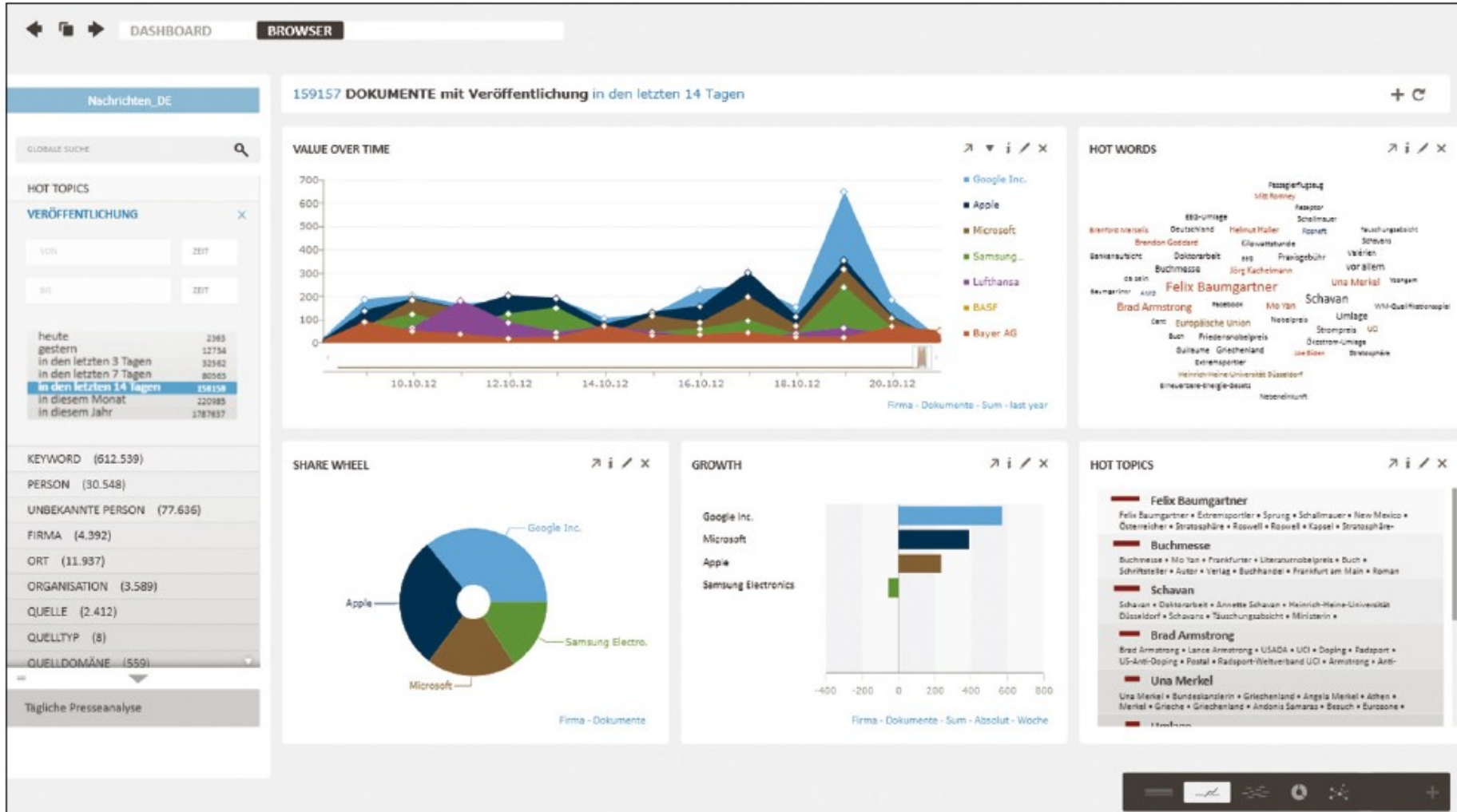


Die »Wörter des Tages« zeigen, welche Begriffe heute besonders aktuell sind. Dazu werden verschiedene Tageszeitungen und Newsdienste täglich ausgewertet. Die »Wörter des Tages« stehen morgens ab etwa 7 Uhr zur Verfügung. Die Aktualität eines Begriffs ergibt sich aus seiner Häufigkeit heute, verglichen mit seiner durchschnittlichen Häufigkeit über längere Zeit hinweg.



Die Daten werden aus sorgfältig ausgewählten öffentlich zugänglichen Quellen automatisch erhoben. Die Beispielsätze werden automatisch ausgewählt und stellen keine Meinungsäußerung des Projektes Deutscher Wortschatz dar. Für die darin enthaltenen Inhalte und Meinungen sind ausschließlich die Autoren verantwortlich.

Beispiel Informations- und Wissensmanagement



Linguistische Ebenen

Linguistische Ebenen bei der Zerlegung eines Satzes

Der Satz „*Das Gewandhaus zu Leipzig befindet sich am Augustusplatz*“ besteht aus:

1) einer Konkatenation von Buchstaben:

D-a-s-G-e-w-a-n-d-h-a-u-s-z-u-L-e-i-p-z-i-g-b-e-f-i-n-d-e-t-s-i-c-h-a-m-A-u-g-u-s-t-u-s-p-l-a-t-z

2) einer Konkatenation von Morphemen:

Das-Gewand-haus-zu-Leipzig-be-find-et-sich-am-Augustus-platz

3) einer Konkatenation von Wortformen:

Das-Gewandhaus-zu-Leipzig-befindet-sich-am-Augustusplatz

4) einer Konkatenation von Phrasen:

Das Gewandhaus-zu Leipzig-befindet sich-am Augustusplatz

Linguistische Ebenen

Explanandum

Laute (tokens)

Lautgruppen

*Phonem: kleinste bedeutungs-
unterscheidende Einheit*

Gruppen von Phonemen

*Morphem: kleinste bedeutungstragende
Einheit*

*Allomorphe: bedeutungsäquivalente
Morpheme*

Bsp.: sprech={sprech, sprich, sprach, ...}

Explanans

Phonetik

Phonologie

Morphologie

Linguistische Ebenen

Gruppen von Morphemen

Wortformen: flektierte Formen eines Wortes

Wort: Äquivalenzklasse v. Wortformen

Gruppen von Wörtern

Phrasen: zulässige Kombination von Wortformen

Sätze: grammatisch vollständige Sequenz von Phrasen

Aussagen: wahrheitsfähige Sätze

Sprechakte: zustandsverändernd

Lexikon

Syntax

Semantik /
Pragmatik

Phonetische Analyse

- **MAUS service (BAS)**

<https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic>

Online Ressourcen Daten und Verfahren

- <http://wortschatz.uni-leipzig.de/en>
- <http://corpora.informatik.uni-leipzig.de/>
- <http://www.nltk.org/data>
- <http://alias-i.com/lingpipe/index.html>
- <http://asv.informatik.uni-leipzig.de/asv/methoden>
- <http://nltk.googlecode.com/svn/trunk/doc/api/index.html>

Literaturempfehlung

Grewendorf/Hamm/Sternefeld, Linguistisches Wissen,
Suhrkamp (stw 695), Frankfurt (13) 2004

C.Manning/H.Schütze, Foundations of Statistical
Natural Language Processing, MIT Press 2000

D.Jurafsky/J.Martin, Speech and Language
Processing, Prentice Hall, NY 2000

Computerlinguistik im Internet, u.a.
<http://www.sil.org>
<http://www.aclweb.org>