Adaptive Informationsaufbereitung aus heterogenen Quellen (AIPHES)

– Herausforderungen und Werkzeuge –

Iryna Gurevych und **Christian M. Meyer**

RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG

Heidelberger Institut für
Theoretische Studien

TECHNISCHE
UNIVERSITÄT
DARMSTADT

DFG-Graduiertenkolleg № 1994 AIPHES

http://www.aiphes.tu-darmstadt.de

Informationsaufbereitung

TECHNISCHE UNIVERSITÄT DARMSTADT

Intensive Recherche ist zentral für sämtliche journalistische Tätigkeiten

- Sehr enge Zeitvorgaben
- Hochgradig heterogene Quellen
- Unterschiedliche Informationsqualität
- Explodierende Informationsmenge

Manuelle Auswertung kaum mehr praktikabel

Aber: Effektive Informationsaufbereitung ist erfolgsentscheidend

Datenjournalismus

TECHNISCHE UNIVERSITÄT DARMSTADT

Sammlung, Aufbereitung und Auswertung von

- **strukturierten Daten** (Tabellen, Datenbanken, Open Data)
  - Automatische Methoden der Datenanalyse
  - *Beispiel: Mindestlohn im weltweiten Vergleich (The Guardian Datablog)*

New York's $15 minimum wage would be the highest in the world

Fast-food workers in New York are now set for a minimum wage of $15 an hour. Here is how that compares with other states, countries and major jurisdictions

http://www.theguardian.com/news/datablog/2015/jul/24/new-york-15-dollar-minimum-wage-highest-in-world

Analyse strukturierter Daten

- Eingabe: Daten zum Mindestlohn weltweit
- Vereinheitlichung z.B. Währung
- Visualisierung

Adaptiert von : http://www.theguardian.com/news/datablog/2015/jul/24/new-york-15-dollar-minimum-wage-highest-in-world

Datenjournalismus

TECHNISCHE UNIVERSITÄT DARMSTADT

Sammlung, Aufbereitung und Auswertung von

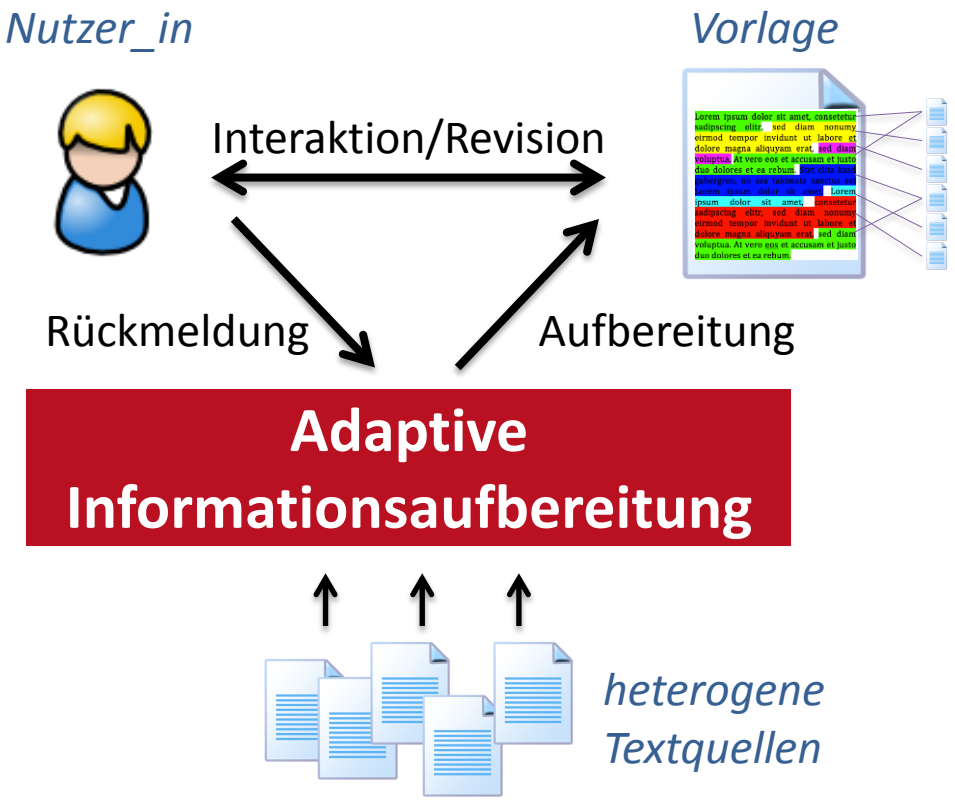
- **strukturierten Daten** (Tabellen, Datenbanken, Open Data)
  - Automatische Methoden der Datenanalyse
  - *Beispiel: Mindestlohn im weltweiten Vergleich (The Guardian Datablog)*
- **unstrukturierten Daten** (Interviews, Pressemitteilungen, Protokolle, Blogs,…)
  - Häufig: Erschließung großer Textsammlungen durch Suche
  - Für weiterführende Auswertungen: Automatische Methoden der Textanalyse
  - **Hohes Potenzial und Forschungsbedarf**

Graduiertenkolleg AIPHES

TECHNISCHE UNIVERSITÄT DARMSTADT

Vision:

- Wissen aus heterogenen Textquellen automatisiert extrahieren
- und zu einem informativen und stilistisch homogenen Dossier aufbereiten
- Anpassung an unterschiedliche Textsorten, Sachgebiete, Nutzergruppen und Sprachen
- Enge Kooperation mit Online-Redaktionen und internationalen Partnern

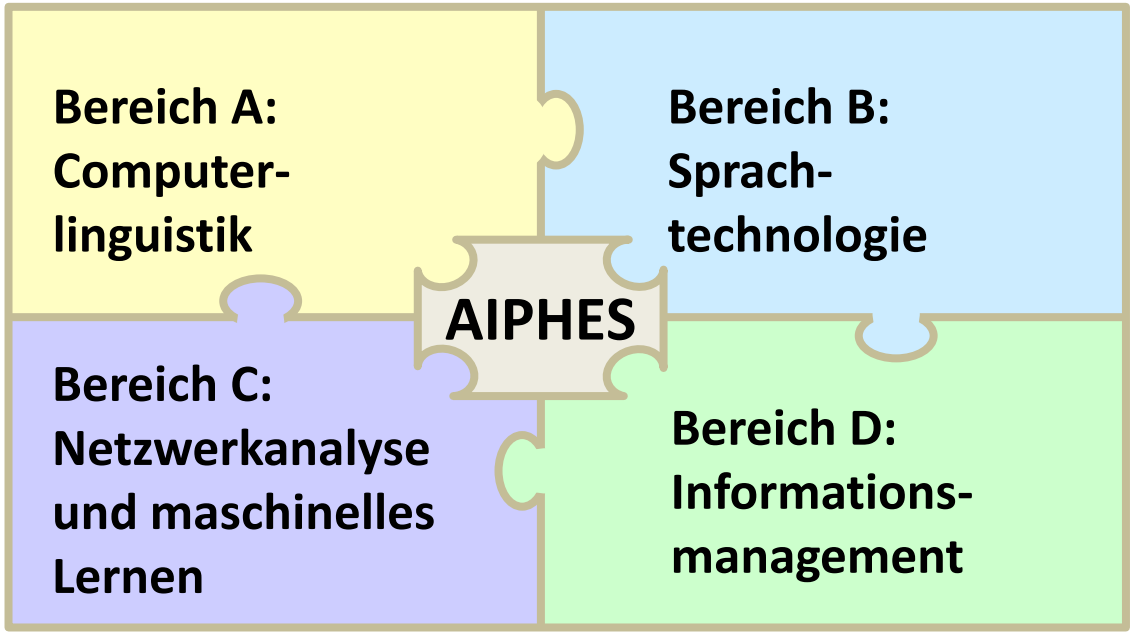


UKP

Graduiertenkolleg AIPHES

TECHNISCHE UNIVERSITÄT DARMSTADT

- Kooperation der Universitäten Darmstadt und Heidelberg sowie des Heidelberger Instituts für Theoretische Studien (HITS)
- 11 Promovierende aus 4 thematischen Bereichen
- Netzwerk von Assoziierten und Kooperationspartnern
- Auftakt: 1. April 2015



https://www.aiphes.tu-darmstadt.de/

UKP

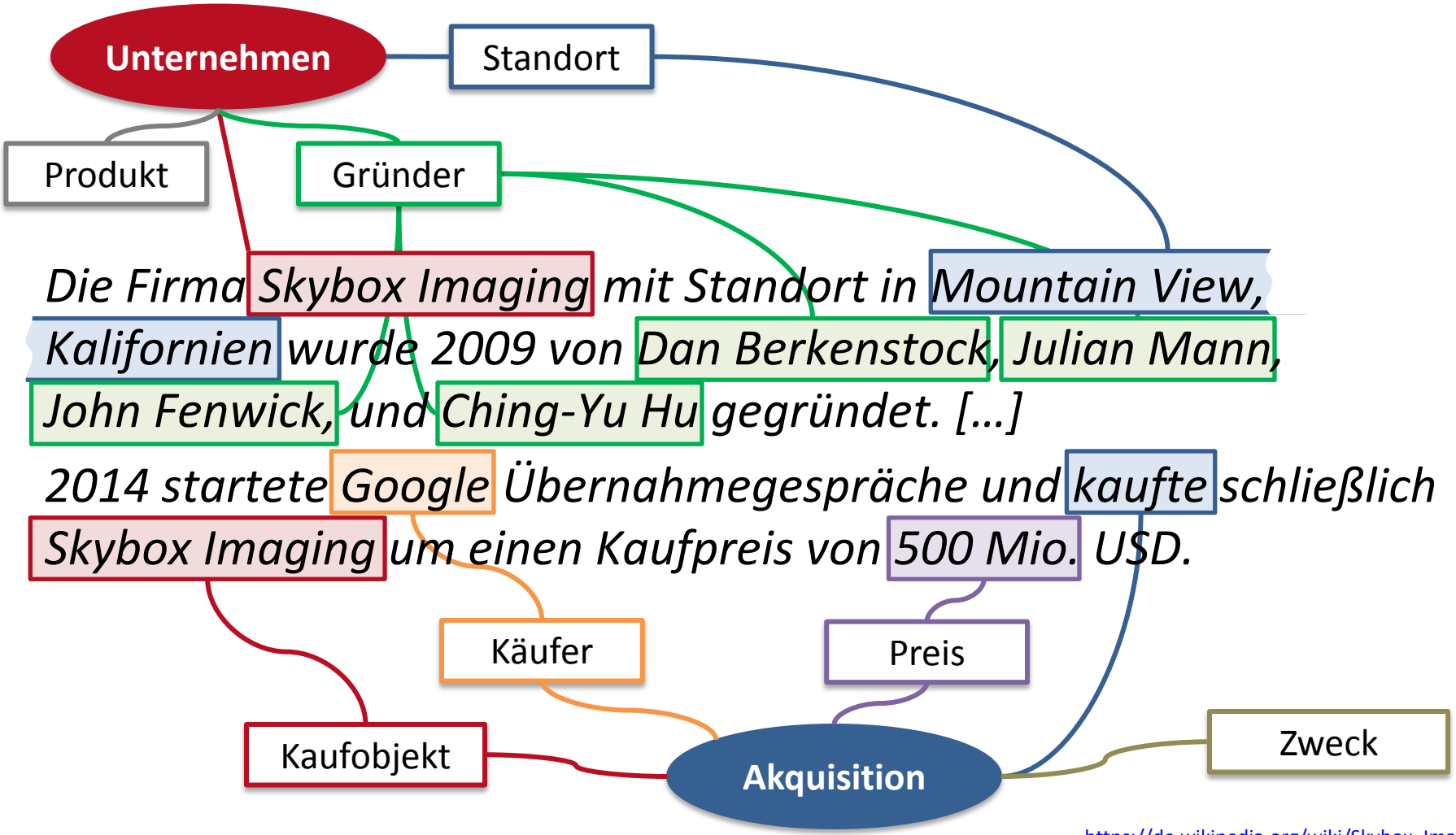
Automatische Zusammenfassung

TECHNISCHE UNIVERSITÄT DARMSTADT

- Prototypische Anwendung in AIPHES
- Automatische Identifikation der zentralen Aussagen einer Dokumentensammlung
- Stilistische Vereinheitlichung von heterogenen Texttypen
- Erprobung neuartiger Verfahren der Diskursanalyse und des maschinellen Lernens

UKP

Diskursanalyse

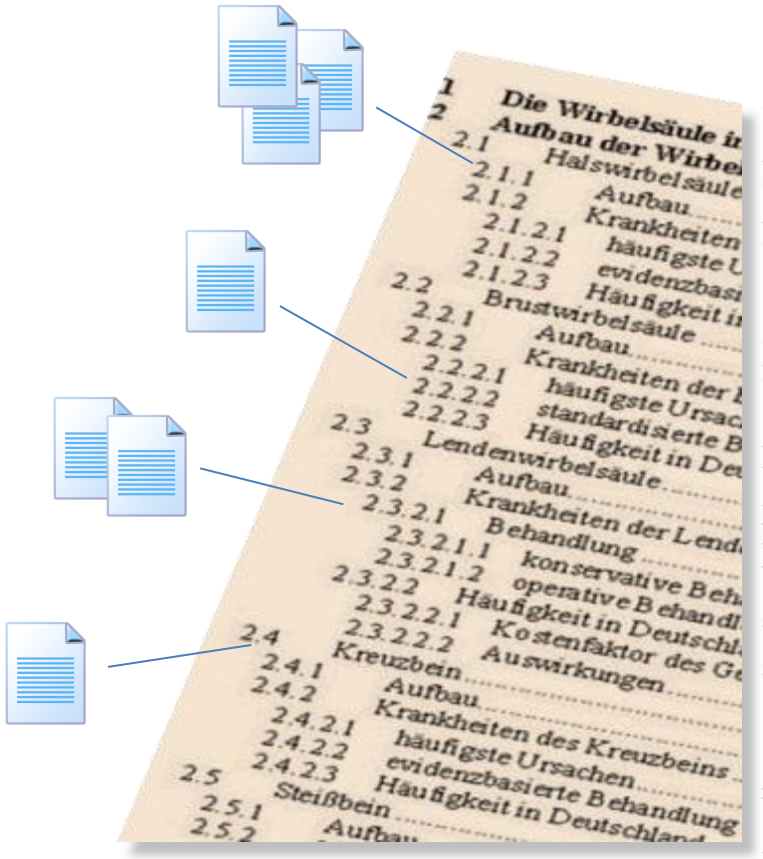


https://de.wikipedia.org/wiki/Skybox_Imaging

Interaktive Navigationsverzeichnisse

TECHNISCHE UNIVERSITÄT DARMSTADT

- Neue Zugriffspfade zu heterogenen Dokumentsammlungen nach…
  - Zeit
  - Personen
  - Kernaussagen
  - Ereignisstrukturen und komplexen Zusammenhängen
  - Argumenten und Standpunkten
  - Informationsqualität



Fact-Checking

TECHNISCHE UNIVERSITÄT DARMSTADT

POLITIFACT
WINNER OF THE PULITZER PRIZE

EDITIONS | TRUTH-O-METER™ | 2016 | PEOPLE | PROMISES | PANTS-ON-FIRE

Our latest fact-checks

HILLARY CLINTON

"There have been seven investigations (of Benghazi) led mostly by Republicans in the Congress" that concluded "nobody did anything wrong, but there were changes we could make."

And the Benghazi committee makes 8

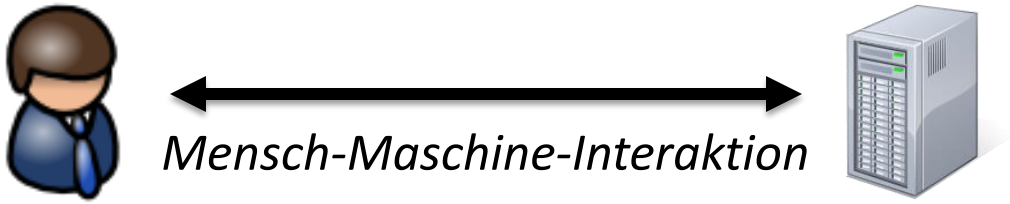
DAVE BRAT

Says Republican Rep. Charlie Dent wants to kick the Freedom Caucus out of the Republican conference "for voting our conscience."

UKP

Mensch-Maschine-Interaktion

Automatische Ansätze sind zwangsläufig fehlerbehaftet, daher:



Interaktion zwischen Mensch und Maschine auswerten

- Verbesserte automatische Verfahren
  „Maschine lernt von menschlichen Expert_innen“
- Schnellere, effektivere Bearbeitung
  „Computer unterstützt Menschen“

Methodische Grundlagen: Verfahren der künstlichen Intelligenz

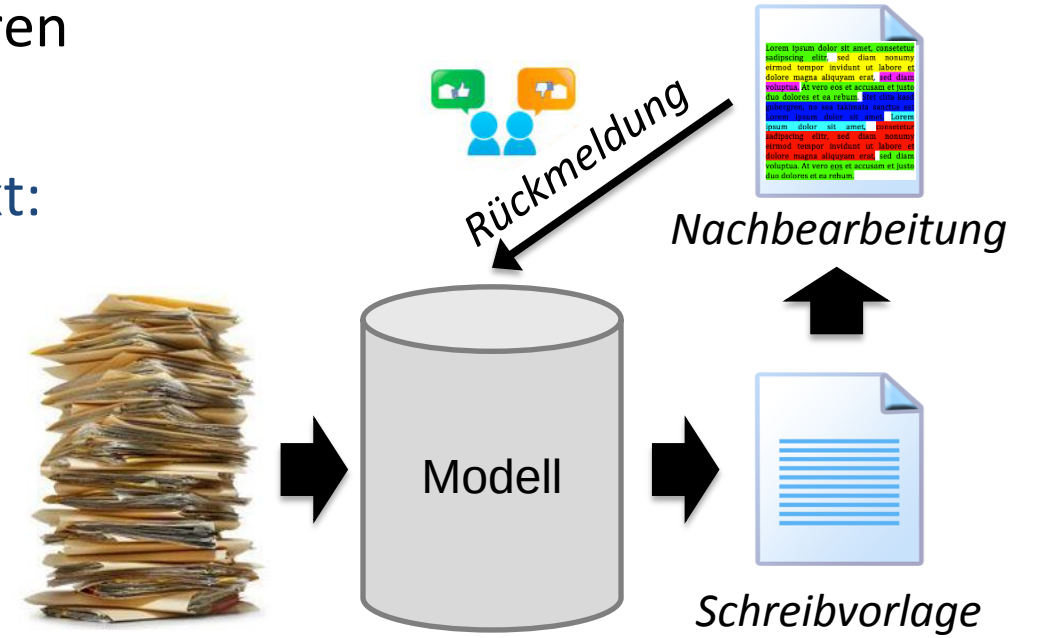
Mensch-Maschine-Interaktion

Prototypische Umsetzung: **Intelligente Schreibassistenzsysteme**

Beispiele:

- System unterstützt die Identifikation wichtiger Aussagen
- Schreiben eines kohärenten Textes mit Unterstützung durch automatische Verfahren

Im Zusammenfassungskontext:
Lassen sich die menschlichen Nachbearbeitungsschritte minimieren?



Kooperationsmöglichkeiten

TECHNISCHE UNIVERSITÄT DARMSTADT

Woran wir interessiert sind:

- Wissenstransfer
- Definition von Anforderungen
- Erprobung von Prototypen
- Evaluation und Referenzdaten

Kooperationsmöglichkeiten

TECHNISCHE UNIVERSITÄT DARMSTADT

Woran wir interessiert sind:

- Wissenstransfer
- Definition von Anforderungen
- Erprobung von Prototypen
- Evaluation und Referenzdaten

Was wir nicht haben:

- Fertige Softwarelösungen für den Produktivbetrieb

Vielen Dank für die Aufmerksamkeit!

TECHNISCHE UNIVERSITÄT DARMSTADT

Kontakt / Contact

Iryna Gurevych und **Christian M. Meyer**
Technische Universität Darmstadt
Ubiquitous Knowledge Processing Lab

Hochschulstr. 10, 64289 Darmstadt, Germany
+49 (0)6151 16–5430
+49 (0)6151 16–5455
{gurevych, meyer} (at) ukp.informatik.tu-darmstadt.de