

Adaptive Informationsaufbereitung aus heterogenen Quellen (AIPHES)

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>



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

Sammlung, Aufbereitung und Auswertung von

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

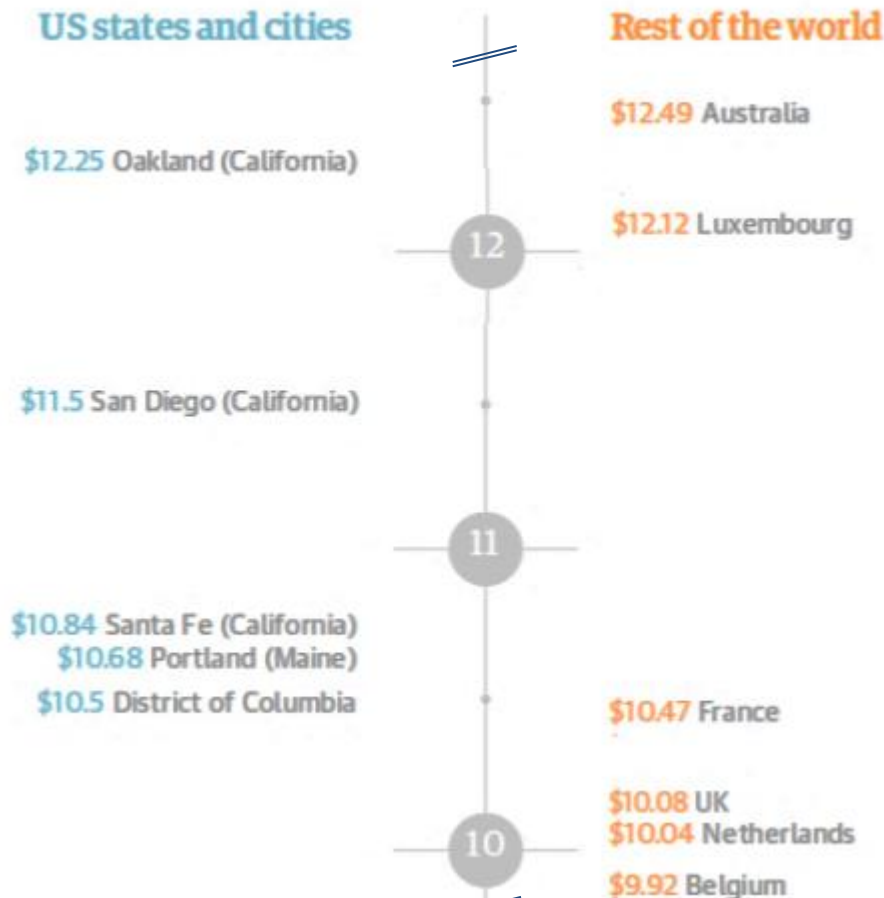
New York's \$15 minimum wage would be the highest in the world

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

Minimum wage compared



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

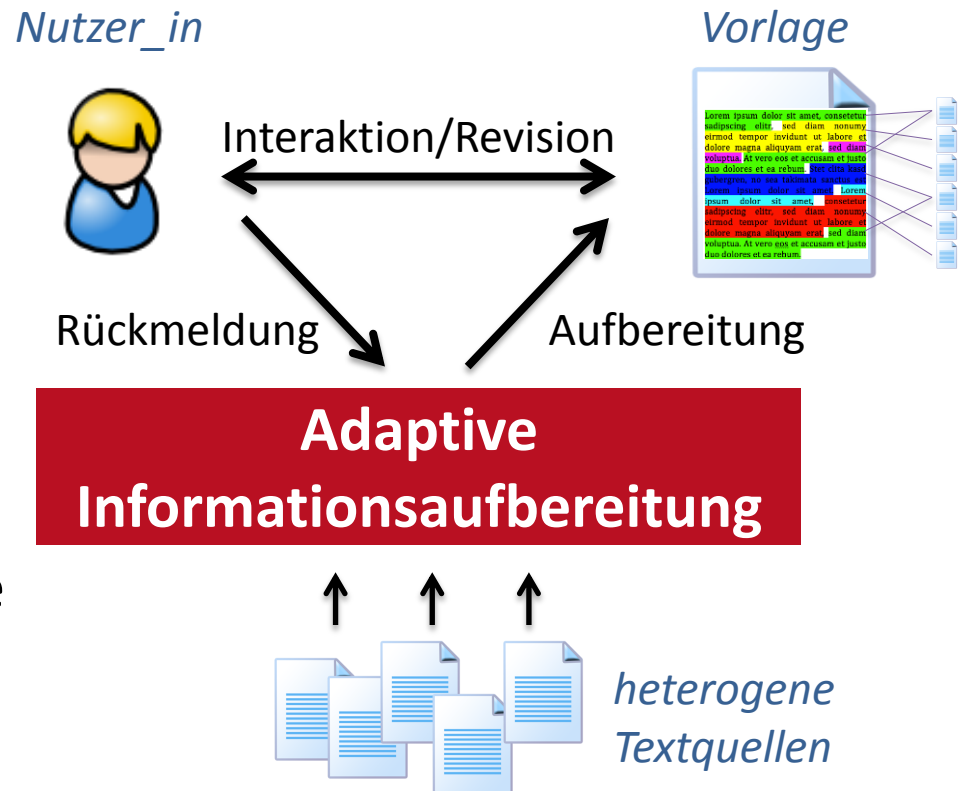
Adaptiert von : <http://www.theguardian.com/news/datablog/2015/jul/24/new-york-15-dollar-minimum-wage-highest-in-world>

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**

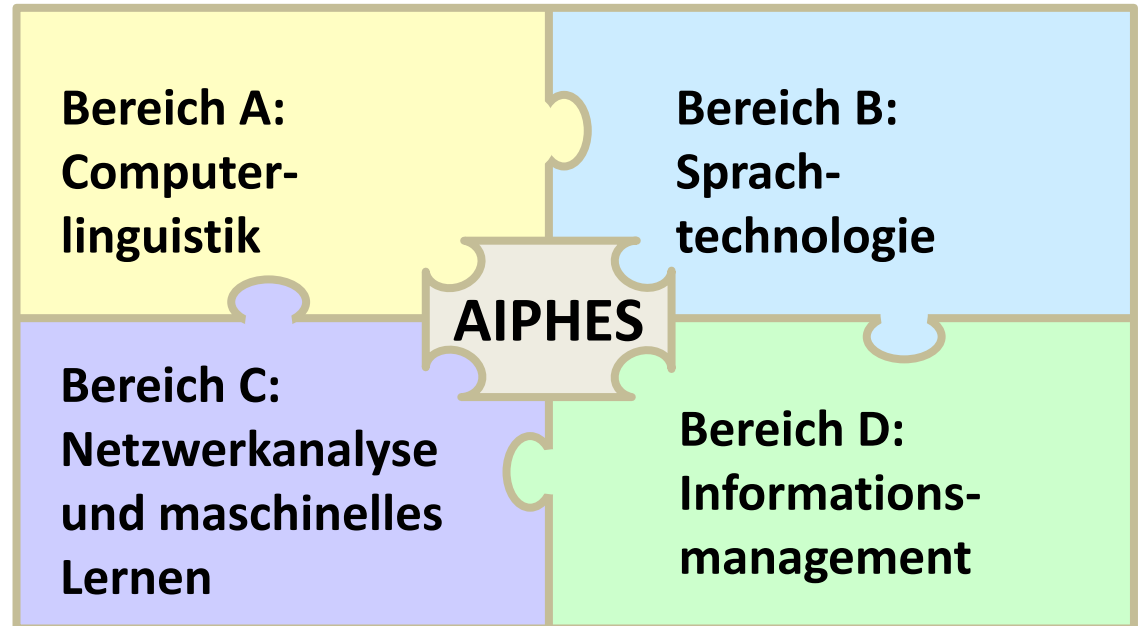
Vision:

- 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



Graduiertenkolleg AIPHES

- 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



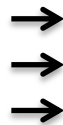
<https://www.aiphes.tu-darmstadt.de/>

Automatische Zusammenfassung

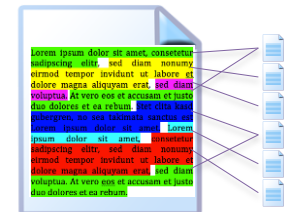
- Prototypische Anwendung in AIPHES mit möglichen Anknüpfungspunkten v.a. im Investigativ- und Datenjournalismus
- Automatische Identifikation der zentralen Aussagen einer Dokumentensammlung
- Stilistische Vereinheitlichung von heterogenen Texttypen
- Erprobung neuartiger Verfahren der Diskursanalyse und des maschinellen Lernens



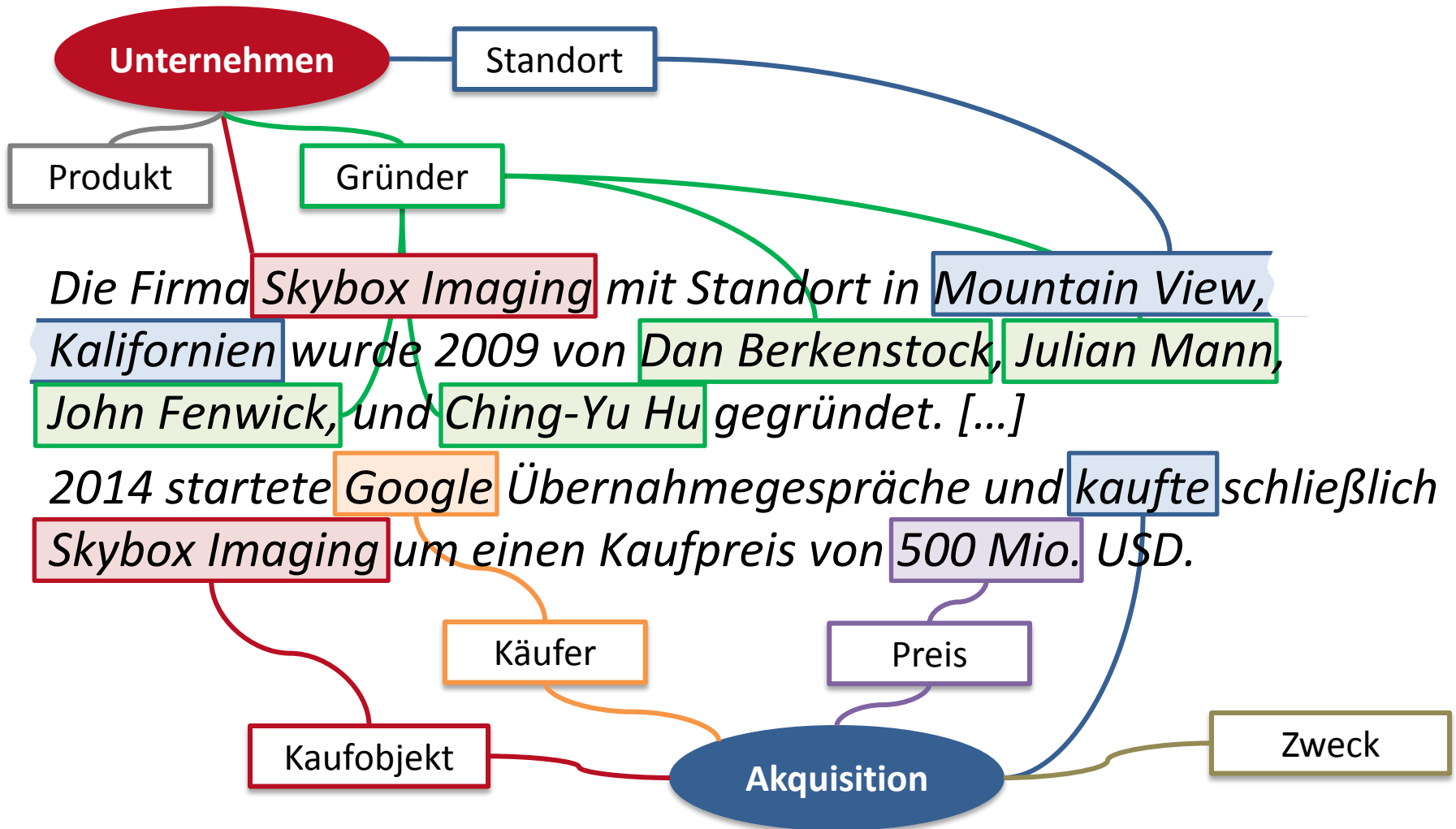
*heterogene
Textquellen*



**Adaptive
Informationsaufbereitung**



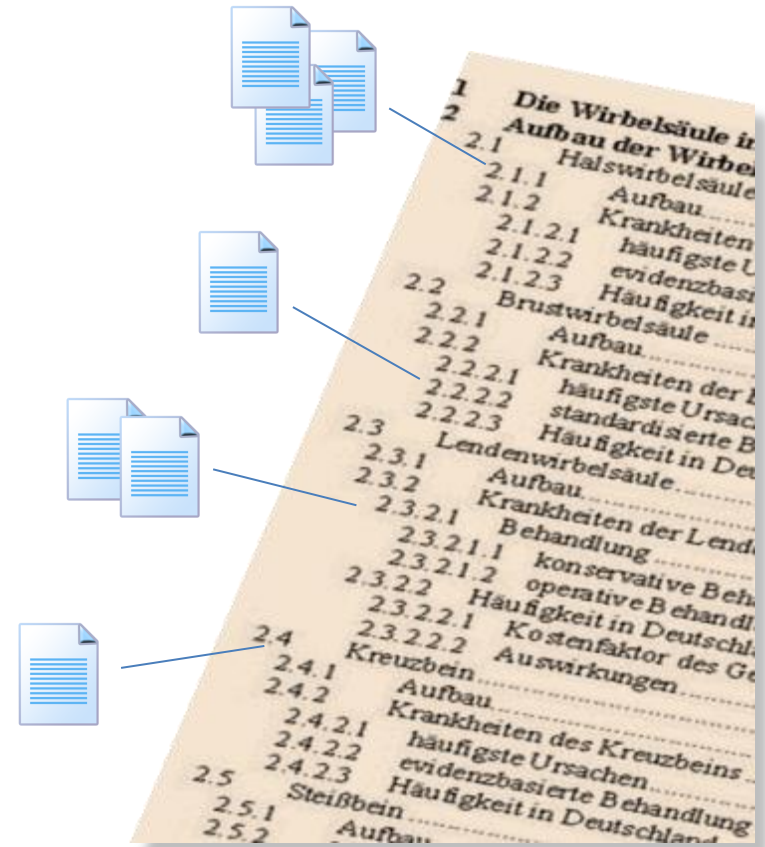
Diskursanalyse



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

Interaktive Navigationsverzeichnisse

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



Fact-Checking



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."



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“

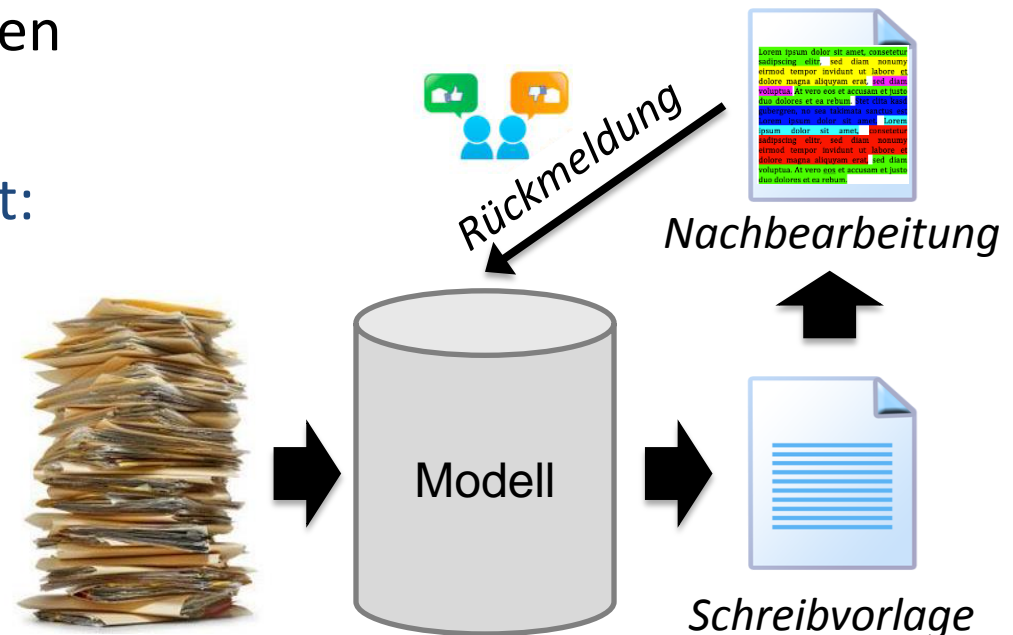
Methodische Grundlagen: Verfahren der künstlichen Intelligenz

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

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!


Kontakt / Contact


Iryna Gurevych und Christian M. Meyer

Technische Universität Darmstadt

Ubiquitous Knowledge Processing Lab

 Hochschulstr. 10, 64289 Darmstadt, Germany

 +49 (0)6151 16–25293

 +49 (0)6151 16–25295

 {gurevych, meyer} (at) ukp.informatik.tu-darmstadt.de

Rechtliche Hinweise

Die Folien sind für den persönlichen Gebrauch der Vortragsteilnehmer gedacht. Im Vortrag verwendete Photographien, Illustrationen, Wort- und Bildmarken sind Eigentum der jeweiligen Rechteinhaber oder Lizenzgeber. Um Missverständnisse zu vermeiden, wäre eine kurze Kontaktaufnahme vor Weitergabe oder -nutzung der Vortragsmaterialien empfehlenswert. Sofern Sie Ihre Rechte verletzt sehen, bitte ich ebenfalls um Kontaktaufnahme zur Klärung der Sachlage.

Legal Issues

The slides are intended for personal use by the audience of the talk. Photographies, illustrations, trademarks, or logos are property of the holder of rights. To avoid any misconceptions, I would strongly recommend to get in touch before reusing or redistributing the slides or any additional material of the talk. The same applies if you consider your rights infringed – please let me know to initiate further clarification.