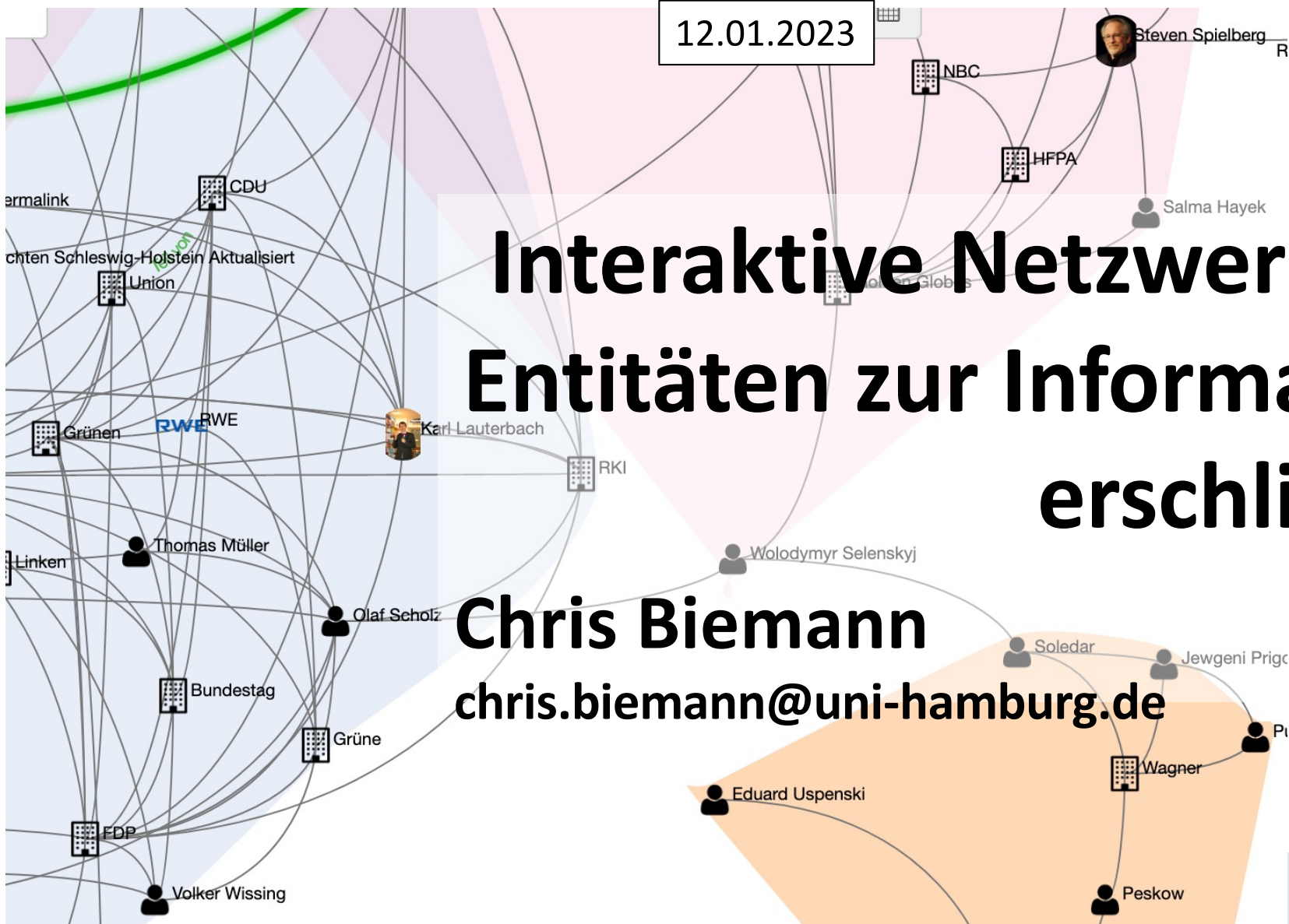


12.01.2023

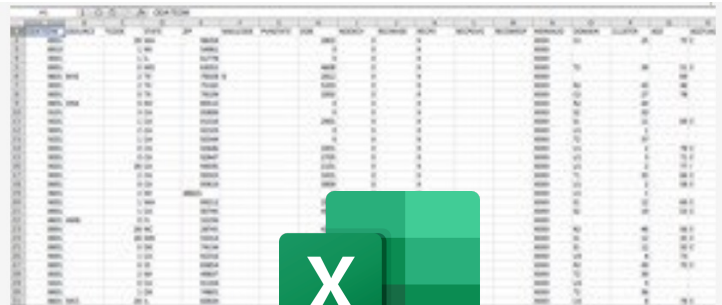

Interaktive Netzwerke von Entitäten zur Informations- erschließung

Chris Biemann
chris.biemann@uni-hamburg.de

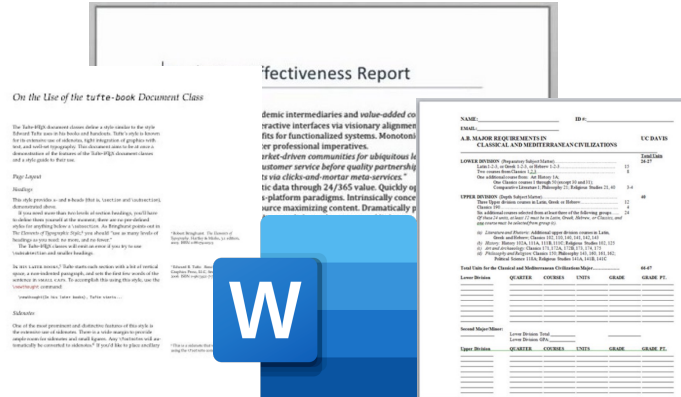



Warum reden wir über Text?

- Wie viel Daten gibt es auf der Welt?
Schätzung: ~2 Zettabytes (Billionen GBs)
- Der Hauptteil der Daten ist unstrukturiert (geschätzt 70%-85%)
- Auch in strukturierten Daten fehlt häufig der Zugang zu unstrukturierten Teilen
 - Unstranskrierte Felder, Notizen, Kommentare
 - Beispiel: Notizen von Untersuchungen, Prüfungsprotokolle, Surveys mit Freitextfeldern ..



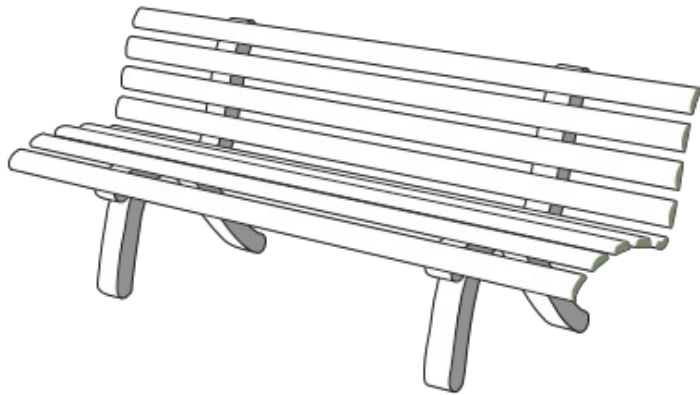
VS.



Warum Sprache schwer ist



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG



polysem

synonym

Er saß auf der **Bank** und zählte seine **Kohle**.

Sie ging zur **Bank** und hob **Geld** ab.

Konzept-
ebene

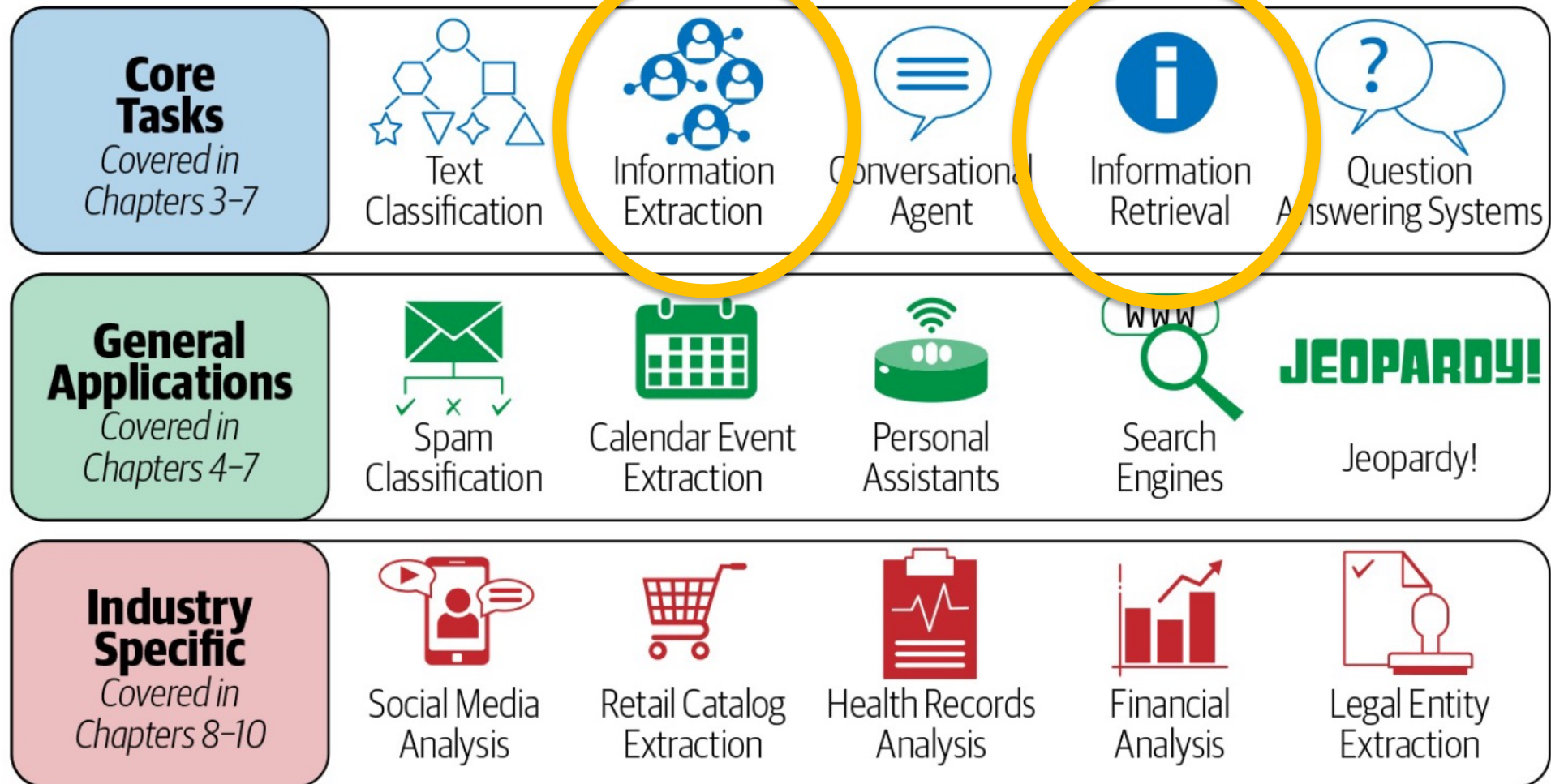
lexikalische
Ebene

Natural Language Processing Tasks



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Taken From: Practical Natural Language Processing : A Comprehensive Guide to Building Real-World NLP Systems
by Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana



Entitätenerkennung im Text

- “Named Entity Recognition” der Klassen Person, Organisation, Ort und Sonstige ist eine sprachtechnologische Standardaufgabe
- Sequenztaggingverfahren werden mit annotiertem Text trainiert.

*Nehmen Sie die berühmte Rede von **Richard Feynman**, **There 's Plenty of Room at the Bottom**, von 1959, die ist damals in der Zeitschrift des **California Institute of Technology** abgedruckt worden.*

- Genauigkeit abhängig von Sprache, Domäne, Anzahl Klassen, Volumen Training

- Englisch, Zeitungstext, 4 Klassen, 1M Tokens: 94% ; bei feineren Klassen 55-95%.

- Angriffe, welches sich die „Wäschfra“

English: "[the] attacks that the Wäschfra [has taken the liberty to conduct] against [...]"

NE GT - Angriffe, welches sich Die **PRODUCT** „3Sd)cb.fra" gegen

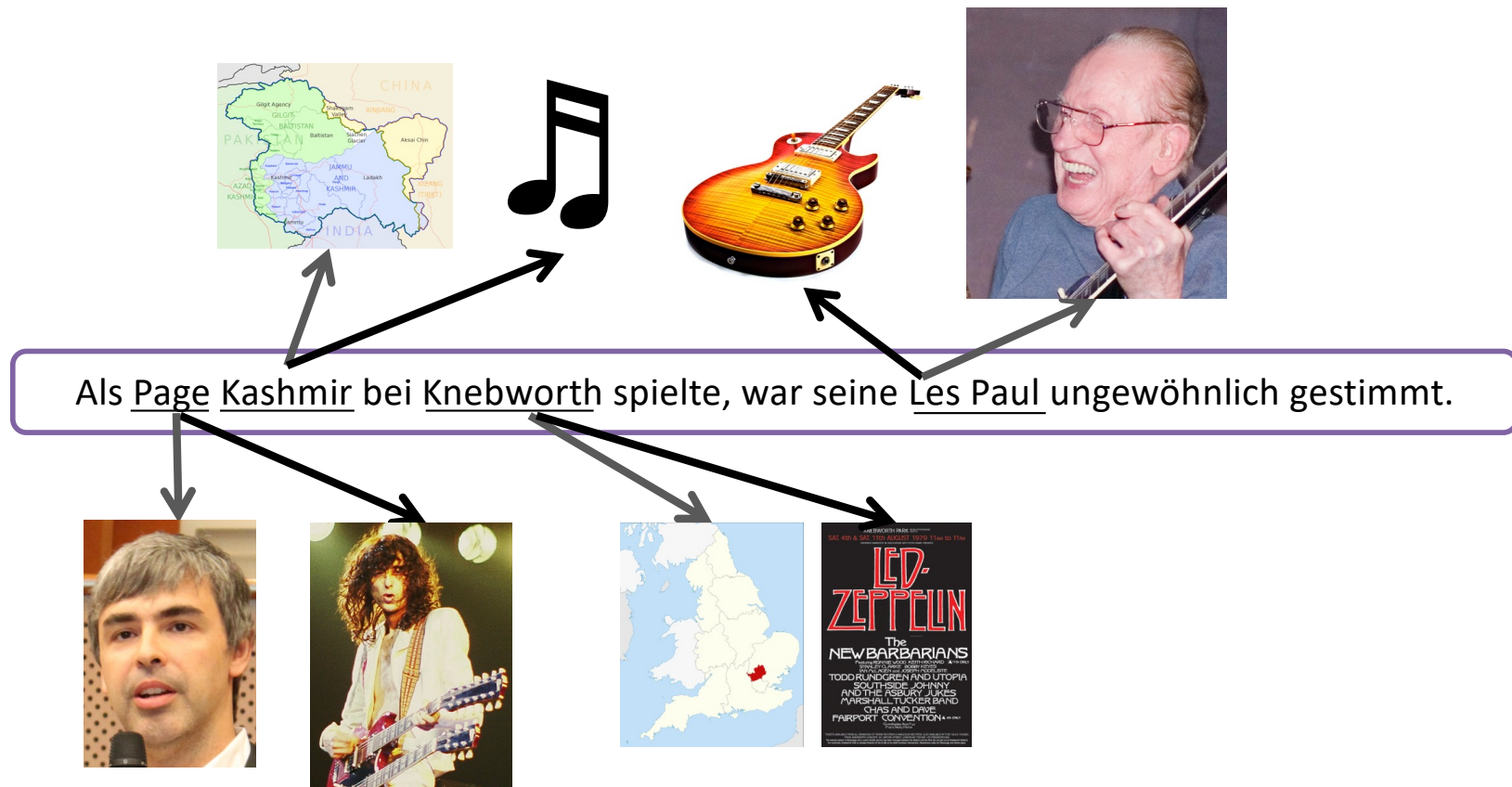
Khai Mai, Thai-Hoang Pham, Minh Trung Nguyen, Tuan Duc Nguyen, Danushka Bollegala, Ryohei Sasano, and Satoshi Sekine. 2018. An Empirical Study on Fine-Grained Named Entity Recognition. In Proceedings of the 27th International Conference on Computational Linguistics, pages 711–722, Santa Fe, New Mexico, USA.

Maud Ehrmann, Ahmed Hamdi, Elvys Linhares Pontes, Matteo Romanello, Antoine Doucet: Named Entity Recognition and Classification on Historical Documents: A Survey. CoRR abs/2109.11406 (2021)

Entity Linking: Identifikation der referierten Entität




Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG



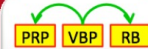
Setzt voraus, dass die Entität auch in der dahinterliegenden Ontologie bekannt ist.

Schnelles Erzeugen von Trainingstext für eigene Kategorien


 **Automation**


WebAnno | [Home](#)

Help | User: chris | [Log out](#)

 **WebAnno**

Document

 [Open](#)


 [Prev.](#)


 [Next](#)

 [Export](#)

 [Settings](#)

Page


 [First](#)

 [Prev.](#)


5

[Go to](#)


 [Next](#)

 [Last](#)

Help

 [Guidelines](#)

Workflow

 [Done](#)

ljl-auto-Init-Sugg/abstract8.tsv

showing 5-6 of 10 sentences

Annotation

- 5 Over the past decade , chronic inflammation in visceral adipose tissue (VAT) has gained acceptance as a lead promoter of insulin resistance in obesity .
- 6 A great deal of evidence has pointed to the role of adipokines and innate immune cells , in particular , adipose tissue macrophages , in the regulation of fat inflammation and glucose homeostasis .

User: Suggestion

- 5 Over the past decade , chronic inflammation in visceral adipose tissue (VAT) has gained acceptance as a lead promoter of insulin resistance in obesity .
- 6 A great deal of evidence has pointed to the role of adipokines and innate immune cells , in particular , adipose tissue macrophages , in the regulation of fat inflammation and glucose homeostasis .

Yimam, S.M., Biemann, C., Majnaric, L., Šabanović, Š., Holzinger, A. (2015): Interactive and Iterative Annotation for Biomedical Entity Recognition, International Conference on Brain Informatics and Health (BIH'15), London, UK

Nach Annotation von 5 Abstracts



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

5 Over the past decade , DISORDER chronic inflammation in CONDITION visceral adipose tissue (VAT) has gained acceptance as a
lead promoter of DISORDER insulin resistance in CONDITION obesity .

6 A great deal of evidence has pointed to the role of MOLECULE adipokines and CELL innate immune cells , in particular ,
CELL adipose tissue macrophages , in the regulation of DISORDER fat inflammation and CONDITION glucose homeostasis .

5 Over the past decade , CONDITION chronic inflammation in CONDITION visceral adipose tissue (VAT) has gained acceptance as a lead promoter of insulin resistance in obesity .

6 A great deal of evidence has pointed to the role of adipokines and innate CELL immune cells , in particular ,
CONDITION adipose tissue macrophages , in the regulation of fat inflammation and glucose homeostasis .

Nach Annotation von 9 weiteren Abstracts



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

5 Over the past decade , DISORDER chronic inflammation in CONDITION visceral adipose tissue (VAT) has gained acceptance as a lead promoter of insulin resistance in DISORDER obesity .

6 A great deal of evidence has pointed to the role of MOLECULE adipokines and CELL innate immune cells , in particular , CELL adipose tissue macrophages , in the regulation of fat DISORDER inflammation and CONDITION glucose homeostasis .

5 Over the past decade , DISORDER chronic inflammation in CONDITION visceral adipose tissue (VAT) has gained acceptance as a lead promoter of insulin resistance in DISORDER obesity .

6 A great deal of evidence has pointed to the role of CELL adipokines and innate immune cells , in particular , CONDITION adipose tissue macrophages , in the regulation of fat DISORDER inflammation and DISORDER glucose homeostasis .

Entitätennetzwerke zur Informationsserschließung

- Von content-basierter zur entitätenzentrierten Informationsrepräsentation
 - Z.B. LDA Topic modelling: content-basierte Sicht auf Dokumente, um die Frage zu beantworten:
was enthält meine Kollektion thematisch? (und wie verändert sich das über die Zeit?)
 - Nun: Ordnen der Information um Entitäten (wie Personen, Organisationen, eigene Klassen), um die Frage zu beantworten:
wer macht was mit wem wann wo?
- Interaktive Kontrolle und Provenienz
 - Kontrolle geht vom Nutzer aus, nicht vom Algorithmus
 - Verlinkung auf die Originalquellen ist IMPERATIV!

Beispiele für entitätenzentrierte Tools

- Tagesnetzwerk: www.tagesnetzwerk.de

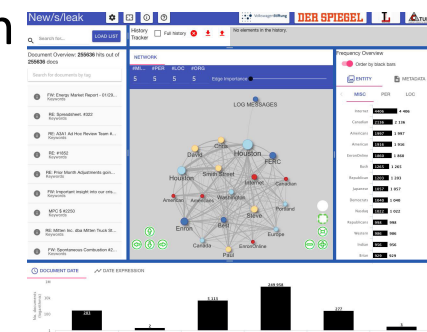
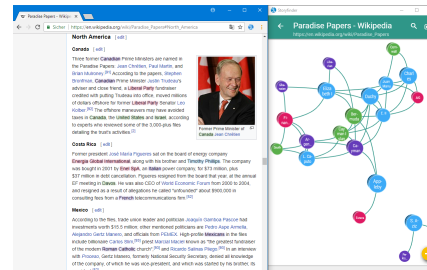
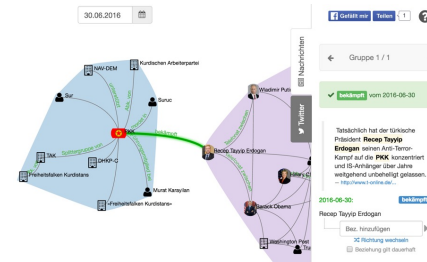
Anhand von Entitäten aggregierender News-Reader

- Storyfinder <https://uhh-lt.github.io/storyfinder/>

Individueller entitätenzentrierter Nachrichten-Assistent als Browser-Plugin

- New/s/leak: www.newsleak.io

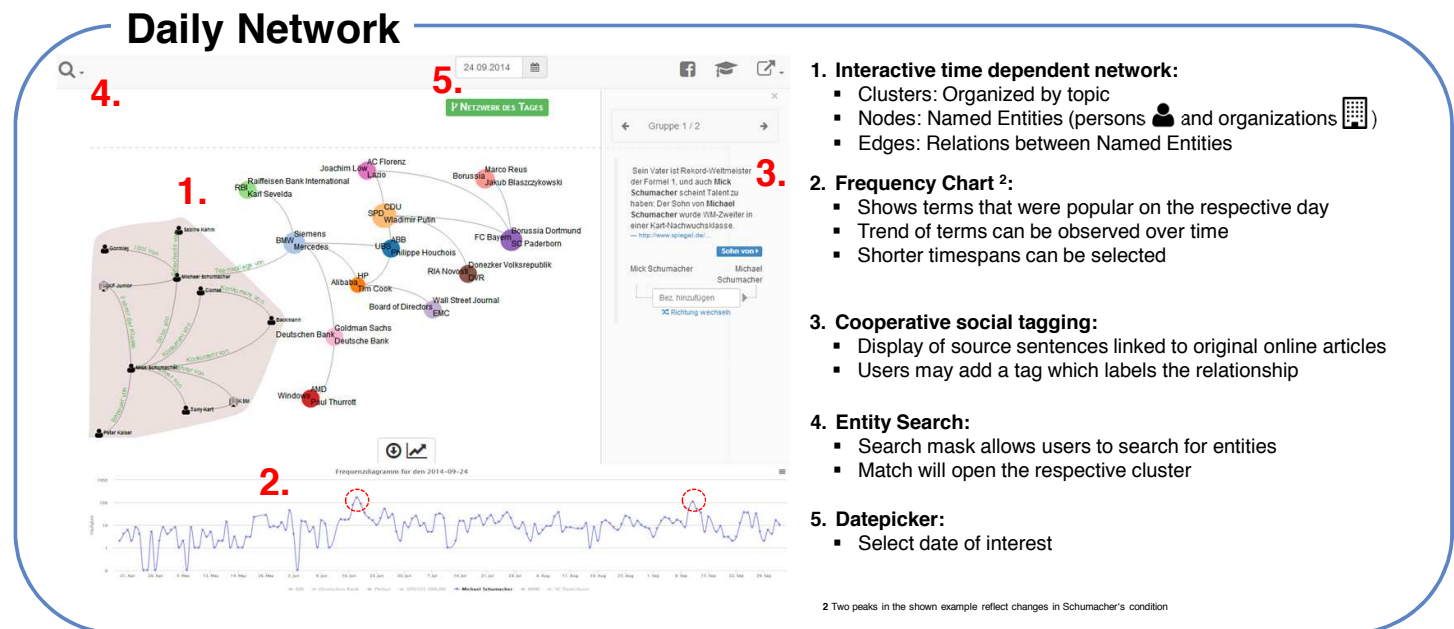
Interaktive facettierte Suchmaske für investigativen Journalismus



Biemann, C., Heyer, G., Quasthoff, U.
(2022): Wissensrohstoff Text. Eine
Einführung in das Text Mining. Springer
Vieweg

- Tägliche Sammlung von deutschen Online-Nachrichten über RSS-Feeds und Webscraping
- Namenserkennung; Indizierung von gemeinsamen Auftreten von Entitäten (Personen und Organisationen)

- Tägliches Netzwerk für Browsing und Annotation von Relationen



What's in the News? (13.1.2023)



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Sophia Schaefer

Filter

13.01.2023



13.01.2023



← Gruppe 1 / 3 →

Der Handel zwischen **RWE**, der Landesregierung und Bundesklima- und Wirtschaftsminister **Robert Habeck** geht sogar noch weiter über diese Bedarfsrechnungen hinaus.

— <https://www.nd-aktuell.de/...>

← Gruppe 2 / 3 →

Sie haben geglaubt, der Kompromiss von **Robert Habeck** und Mona Neubaur mit **RWE** stehe für sich.

— <https://www1.wdr.de/...>

← Gruppe 3 / 3 →

Der Kohle-Abbau in Lützerath geht nämlich auf einen Kompromiss zurück, den der grüne Bundeswirtschaftsminister **Robert Habeck** (53) und die grüne Vize-Ministerpräsidentin Mona Neubaur (45) im Oktober 2022 mit **RWE** ausgehandelt hatten.

<https://www.bild.de/>

WDR



Wetter



Verkehr

im WDR

Nachrichten

Sport

Wissen

Verbraucher

Kultur

Unterhaltung



Kommentar: Der Eiertanz der Grünen um Lützerath

Stand: 11.01.2023, 13:30 Uhr

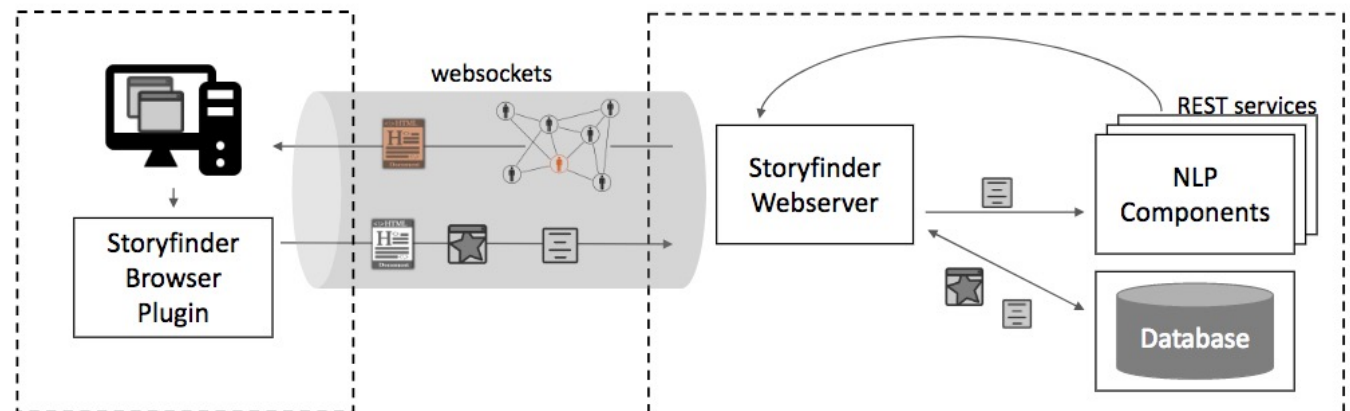
Statt den früheren Kohleausstieg und den Erhalt vieler Dörfer als Erfolg zu feiern, statt Freudentänze am Rande der Kohlegruben aufzuführen, geben sich die regierenden Grünen staatstragend - und zerknirscht.

Storyfinder: Persönlicher Newsreader

<https://uhh-lt.github.io/storyfinder/>

- Ambientes Browser-Plugin
- Erinnerung von Nachrichtenartikeln und inkrementeller Aufbau eines Entitätennetzwerkes
- Momentaner Artikel wird zum globalen personalisierten Entitätennetzwerk in Beziehung gesetzt
- web-basiertes Tool, kann auf Server oder wegen Datenschutz lokal betrieben werden.

Remus, S., Kaufmann, M., Ballweg, K., von Landesberger, T., and Biemann, C. (2017): Storyfinder: Personalized Knowledge Base Construction and Management by Browsing the Web. In Proceedings of the 26th ACM International Conference on Information and Knowledge Management. Singapore, Singapore. pp. 2519-252



Storyfinder

Online-Artikel

× New element SAVE

Name:

Corker

Please select a type:

☐ Person

☐ Organisation

☐ Location

☒ Other

☒ Element is always visible

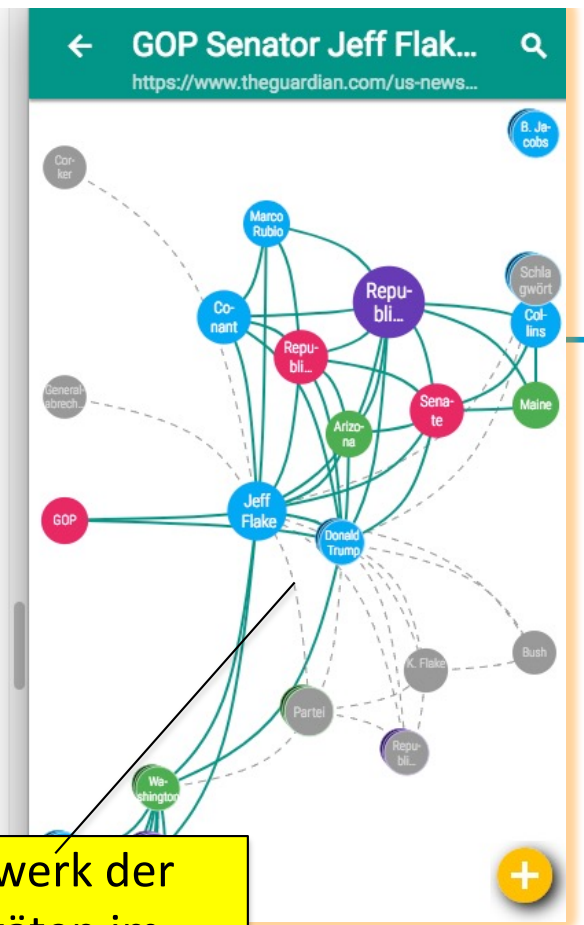
☒ Autodetect relations

Senior Republican Bob Corker announced plans to retire, including Matt of Michigan, Schert of Washington, Lehtinen of Florida, leaving districts. Collins said earlier this month leave to run for governor of "The Republican nominee for this it was a "troubling trend" for reliable conservatives like Flake to feel as though they no longer had a place in the Republican party.

"It's no secret that there's a lot of divisions within the Republican party right now. A lot of Republican leaders are uncomfortable with the direction that Trump is leading us," Conant said.

Additional reporting by Ben Jacobs and David Smith

This article was amended on 25 October 2017 to clarify Susan Collins is not up for re-election in 2018.



Netzwerk der Entitäten im Artikel

- Entitäten können editiert/manuell hinzugefügt werden
- Netzwerk zeigt auch relevante Entitäten aus vorher gelesenen Webseiten

Globale Sicht, Relationen, Provenienz

Sites

theguardian.com
 Senior Republican Bob Corker feuds with Trump: 'He's 'debasing our nation' | US news | The Guardian

tagesschau.de
 Abrechnung mit Trump: 'Nicht länger mitschuldig sein' | tagesschau.de

en.wikipedia.org
 Singapore - Wikipedia

theguardian.com
 Toon and Tyne: Newcastle united ... on a city walking tour | Travel | The Guardian

tagesschau.de
 Transparenz im Netz: Twitter will Wahlwerbung

← Group of 2 sites

GOP Senator Jeff Flake... Senior Republican Bob...



Sites

variety.com
 Jeff Dunham's Growing Empire Looks Forward - Variety

chicago.suntimes.com
 Jeff Bezos, Ken Griffin among those Rahm Emanuel emailed this year | Chicago Sun-Times

theguardian.com
 Senior Republican Bob Corker feuds with Trump: 'He's 'debasing our nation' | US news | The Guardian

tagesschau.de
 Abrechnung mit Trump: 'Nicht länger mitschuldig sein' | tagesschau.de

en.wikipedia.org
 Singapore - Wikipedia

theguardian.com
 Toon and Tyne: Newcastle united ... on a city walking tour | Travel | The Guardian

Global network

Entities

Jeff Flake

Jeff Bezos

Jeff Dunham

Jeff Glor An...

Sites

← Jeff Flake

Person

SOURCES

RELATIONS

Label	Neigh
—	Trump
—	Donald Trump
—	Flake
—	Republican
—	Republican party

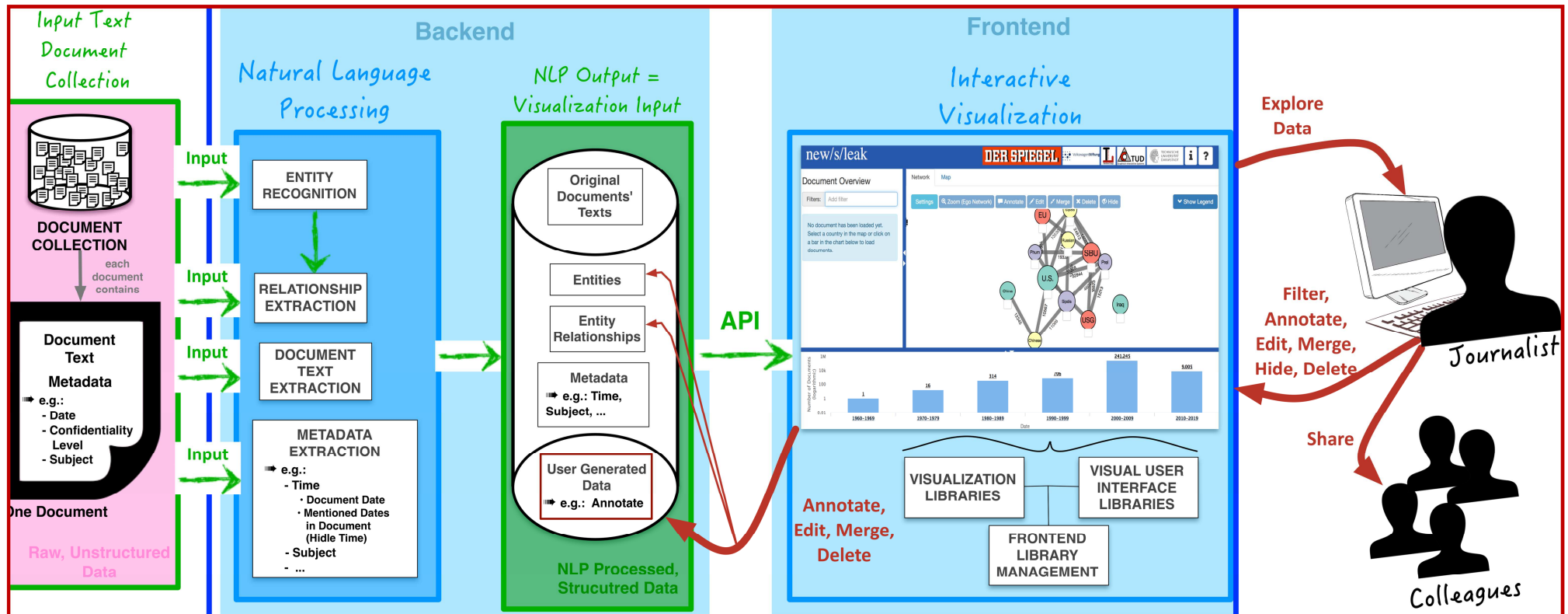
»GOP Senator **Jeff Flake** attacks 'reckless, outrageous and undignified' **Trump**«
 - Source GOP Senator Jeff Flake attacks 'reckless, outrageous and undignified' Trump | US news | The Guardian visited on 2017-10-25T13:05:12.000Z -

Alle Webseiten werden
local gecached

- Nutzer kann eigenen Wissensgraph aufbauen

New/s/leak

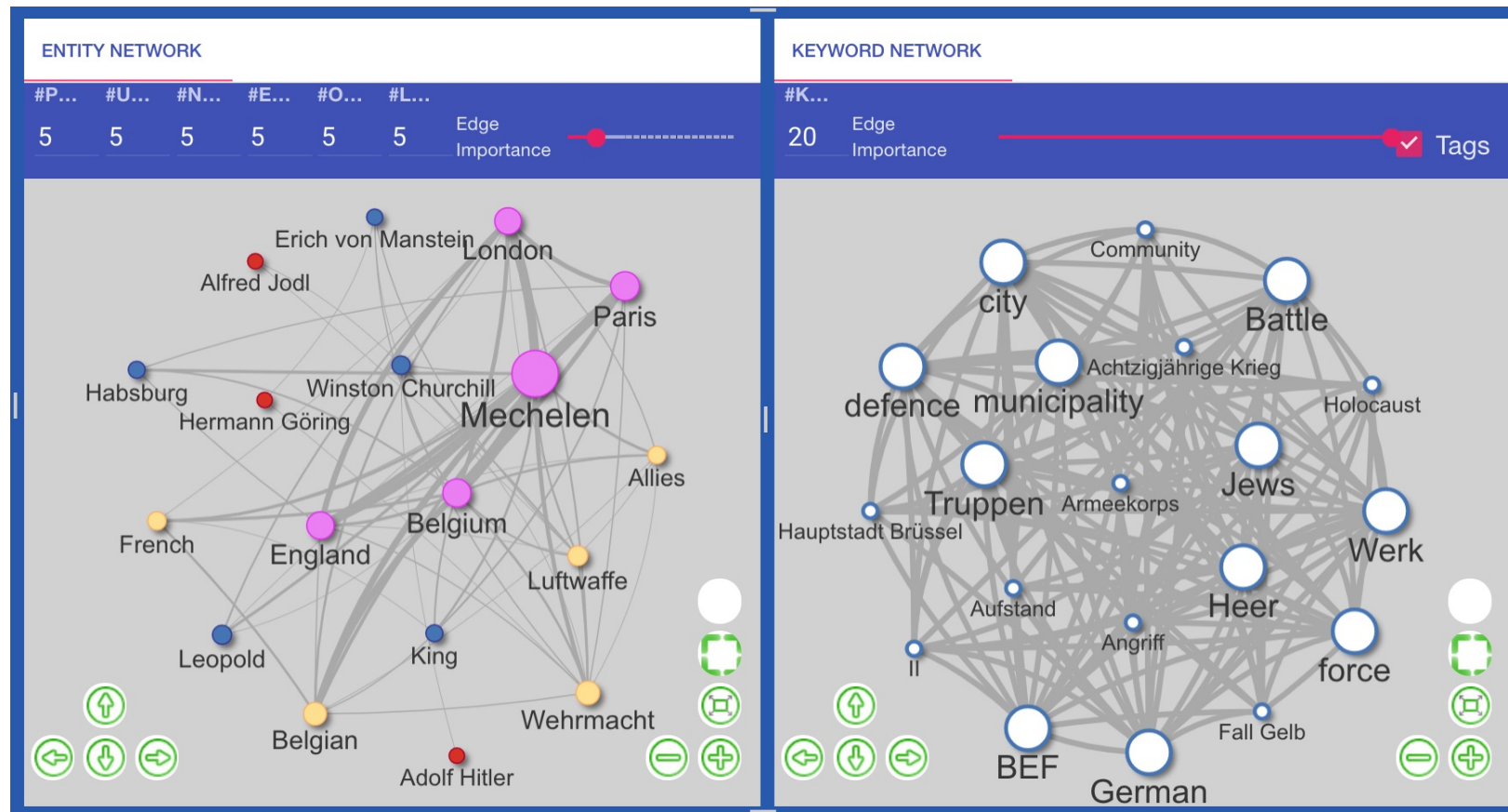
www.newsleak.io



New/s/leak: entitäten- und schlüsselwortzentrierte Ansicht



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

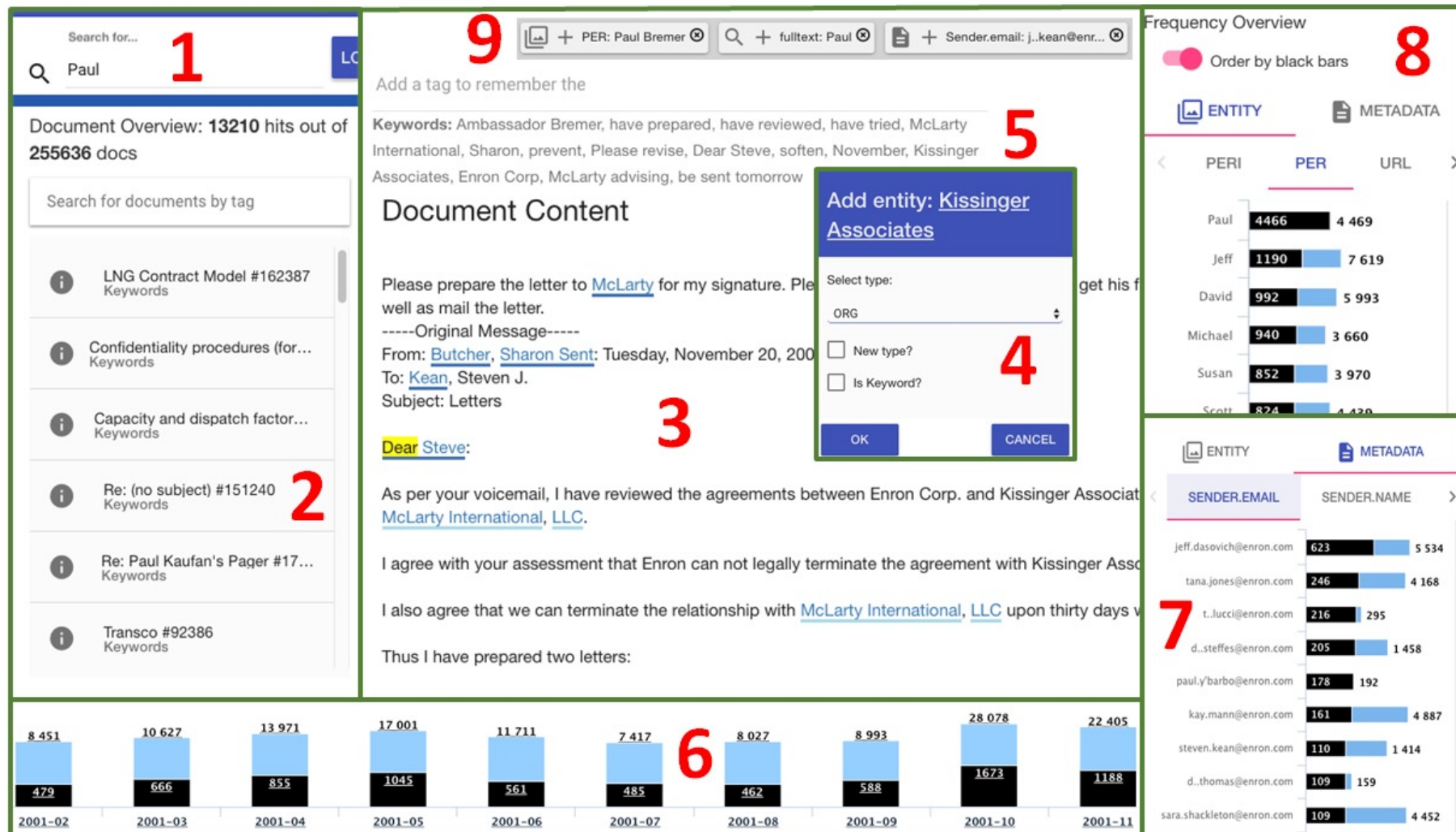


- Alles kann als Filter definiert werden

New/s/leak: Facettierte Suche und Volltextansicht



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG



- Motivation: Text Mining im unstrukturierten Text
- Erkennung von Entitäten im Text
 - Von der Stange
 - Und nach Maß
- Entitätenzentrierte Netzwerke und Facettierte Suche
 - Tagesnetzwerk
 - Storyfinder
 - New/s/leak

**HERZLICHEN DANK FÜR
DIE AUFMERKSAMKEIT**

