

INTERACTIVE EN TITY NETWORKS FOR INFORMATION ACCESS TO TEXT

CHRIS BIEMANN

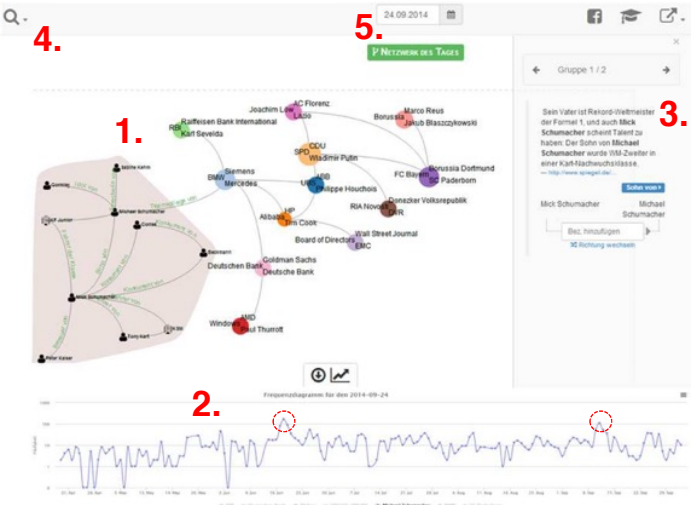
[HTTP://LT.INFORMATIK.UNI-HAMBURG.DE](http://lt.informatik.uni-hamburg.de)

[BIEMANN@INFORMATIK.UNI-HAMBURG.DE](mailto:biemann@informatik.uni-hamburg.de)



- FROM CONTENT-BASED TO ENTITY-CENTRIC VIEW ON DOCUMENT COLLECTIONS
 - FUZZY DOCUMENT CLUSTERING ALGORITHMS SUCH AS LDA TOPIC MODELLING PROVIDE A CONTENT-BASED VIEW ON DOCUMENTS TO ANSWER THE QUESTION: **WHAT IS IN THERE AND HOW DOES IT CHANGE?**
 - NOW: CENTER INFORMATION ON ENTITIES, SUCH AS PERSONS, ORGANIZATIONS AND CUSTOM-DEFINED ENTITIES, TO ANSWER THE QUESTION:
WHO DOES WHAT WITH WHOM WHEN?
- INTERACTIVE CONTROL AND PROVENANCE
 - CONTROL SHOULD BE FULLY SITUATED WITH THE USER, NOT WITH THE ALGORITHM
 - LINKS TO THE ORIGINAL SOURCES FOR INFORMATION ARE CRUCIAL

- COLLECTING GERMAN ONLINE NEWS DAILY FROM RSS FEEDS AND WEBPAGES
- PERFORM NAMED ENTITY RECOGNITION AND INDEX DOCUMENTS BY ENTITY CO-OCCURRENCE
- OFFER A DAILY NEWS ENTITY NETWORK FOR BROWSING AND ANNOTATION

Daily Network



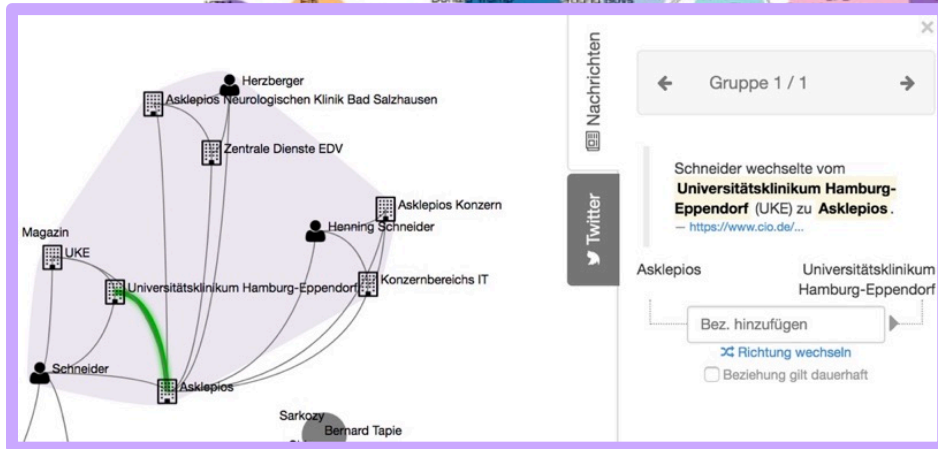
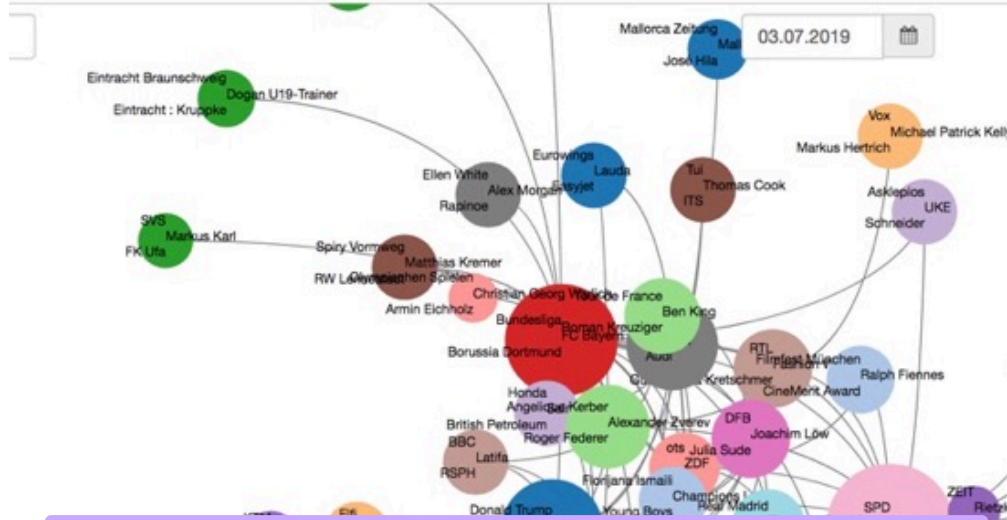
The screenshot shows the 'Daily Network' interface for the date 24.09.2014. It features a network graph (1) with nodes representing named entities and edges representing relationships. A frequency chart (2) at the bottom shows the trend of terms over time. A text snippet (3) on the right is annotated with tags (4) and a date picker (5). The interface includes a search bar, a date picker, and a search mask.



1. **Interactive time dependent network:**
 - Clusters: Organized by topic
 - Nodes: Named Entities (persons  and organizations )
 - Edges: Relations between Named Entities
2. **Frequency Chart ²:**
 - Shows terms that were popular on the respective day
 - Trend of terms can be observed over time
 - Shorter timespans can be selected
3. **Cooperative social tagging:**
 - Display of source sentences linked to original online articles
 - Users may add a tag which labels the relationship
4. **Entity Search:**
 - Search mask allows users to search for entities
 - Match will open the respective cluster
5. **Datepicker:**
 - Select date of interest

2 Two peaks in the shown example reflect changes in Schumacher's condition

WHAT'S IN THE NEWS? (3.7.2019)

Not Secure | tagesnetzwerk.de



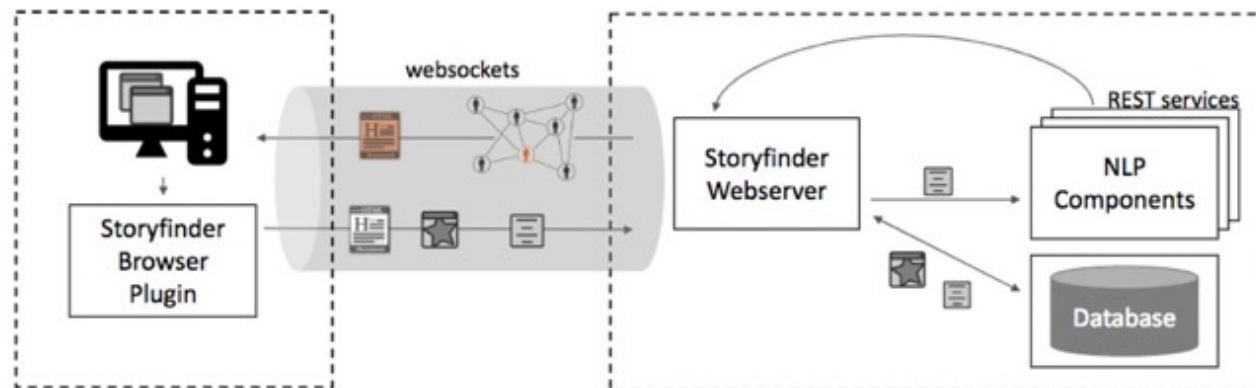
https://www.cio.de/a/berliner-charite-nutzt-spracherkennung-fuer-mehr-effizienz,3601478?utm_sour...
CIO Newsletter Whitepaper Jobs Shop
Healthcare IT
 Digital Healthcare
Berliner Charité nutzt Spracherkennung für mehr Effizienz
 02.07.2019 Von Julia Lammi (Redakteurin)
 Die Berliner Charité setzt im Krankenhausbetrieb auf Datenanalyse in Echtzeit. Der erste Schritt ist der Aufbau einer Infrastruktur mit Spracherkennung.

 Die Software zur Spracherkennung wird an allen vier Standorten der Charité eingesetzt.
 Seit rund einem Jahr nutzt die Charité  Universitätsmedizin Berlin an allen Standorten in der Hauptstadt das Spracherkennungssystem Dragon Medical der Firma Nuance. Mittlerweile sind 1.200 User dafür registriert. Neben Ärzten setzen auch Mitarbeiter der Verwaltung und anderer Geschäftsbereiche die Software ein. Als Stand-Alone-Anwendung ist sie

STORYFINDER: PERSONALIZED NEWSREADER

[HTTPS://UHH-LT.GITHUB.IO/STORYFINDER/](https://UHH-LT.GITHUB.IO/STORYFINDER/)

- AMBIENT BROWSER PLUGIN
- REMEMBERING NEWS STORIES AND INCREMENTALLY BUILDING A NETWORK OF ENTITIES
- RELATING CURRENT ARTICLE TO GLOBAL PERSONALIZED ENTITY NETWORK
- WEB-BASED TOOL, DEPLOYED ON A SERVER OR IN A LOCAL DOCKER CONTAINER FOR PRIVATE REASONS

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

news article

X New element **SAVE**

Name:

Please select a type:

- Person
- Organisation
- Location
- Other

Element is always visible

Autodetect relations

Look Up "Corker"

- Copy
- Search Google for "Corker"
- Print...
- AdBlock
- Add to Storyfinder**
- Inspect
- Speech
- Services



Senior **Republican** Bob Corker feuds with Trump: 'He's 'debasing our nation'

[Read more](#)

announcement, which followed a slew of similar decisions from **Republicans** in swing states.

Several **Republicans** serving in the **House of Representatives** have announced plans to retire, including David Trott of Michigan, David Reichert of Washington, Charlie Dent of Pennsylvania and Ileana Ros-Lehtinen of Florida, leaving

an opening for **Democrats** in competitive districts.

By contrast, moderate **Republican** Susan Collins said earlier this month she would stay in the **Senate** rather than leave to run for governor of Maine.

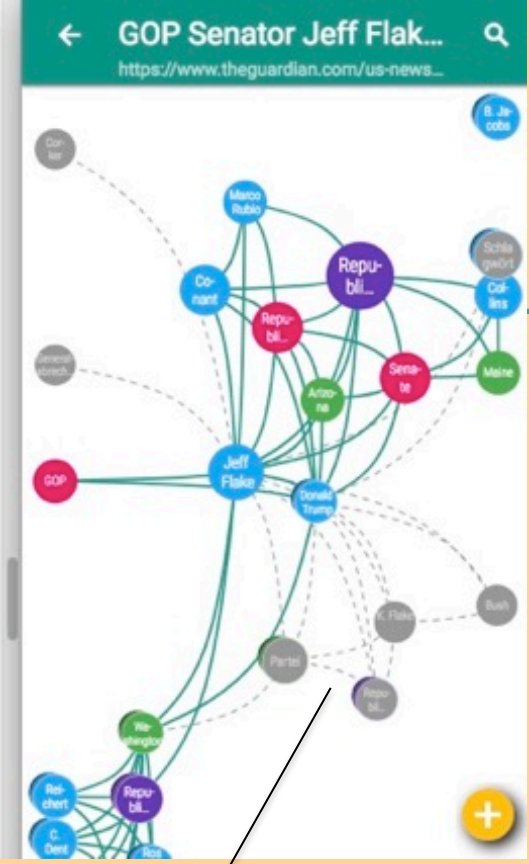
Alex Conant, a partner at Firehouse Strategies who worked on **Marco Rubio's** presidential campaign, said it was a "troubling trend" for reliable conservatives like Flake to feel as though they no longer

party.

lt of divisions within the **Republican party** leaders are uncomfortable with the g us," Conant said.

ncobs and David Smith

in 25 October 2017 to clarify that election in 2018.

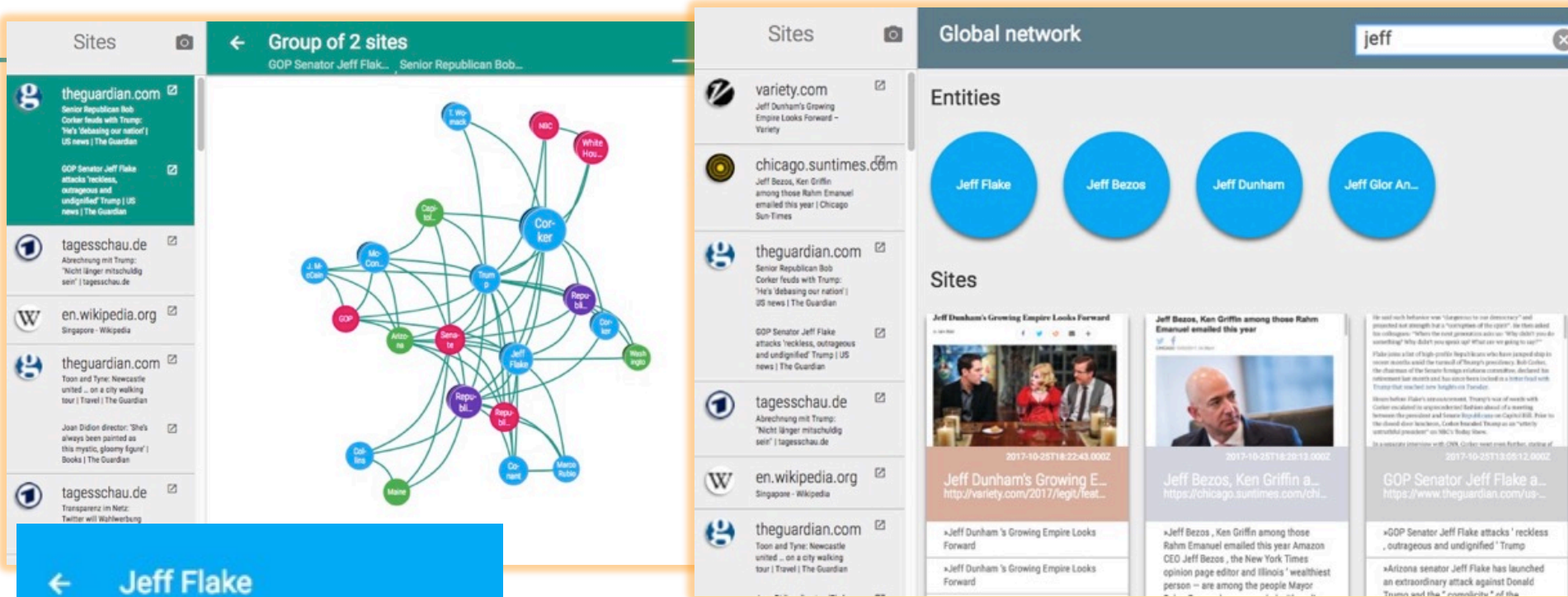


Graph of document entities

- ENTITIES CAN BE ADDED/JOINED/REMOVED
- GRAPH SHOWS RELEVANT ENTITIES FROM PREVIOUSLY READ ARTICLES
- USERS CAN RETRIEVE PREVIOUS ARTICLES



GLOBAL VIEW, RELATIONS AND PROVENANCE



The screenshot displays a web application interface with several panels:

- Left Panel (Sites):** Lists various news sources such as theguardian.com, tagesschau.de, and en.wikipedia.org.
- Center Panel (Group of 2 sites):** Shows a network graph with nodes representing entities like 'Cor-kor', 'Trump', 'Jeff Flake', 'Rep-ubl', and 'Senat'. Edges connect these entities, representing relationships.
- Right Panel (Global network):** Shows search results for 'jeff', including entities like 'Jeff Flake', 'Jeff Bezos', 'Jeff Dunham', and 'Jeff Glor An...', along with a list of sites.
- Bottom Panel (Jeff Flake):** A detailed view of the 'Jeff Flake' entity, showing a list of sources and a table of relations.

← Jeff Flake

Person

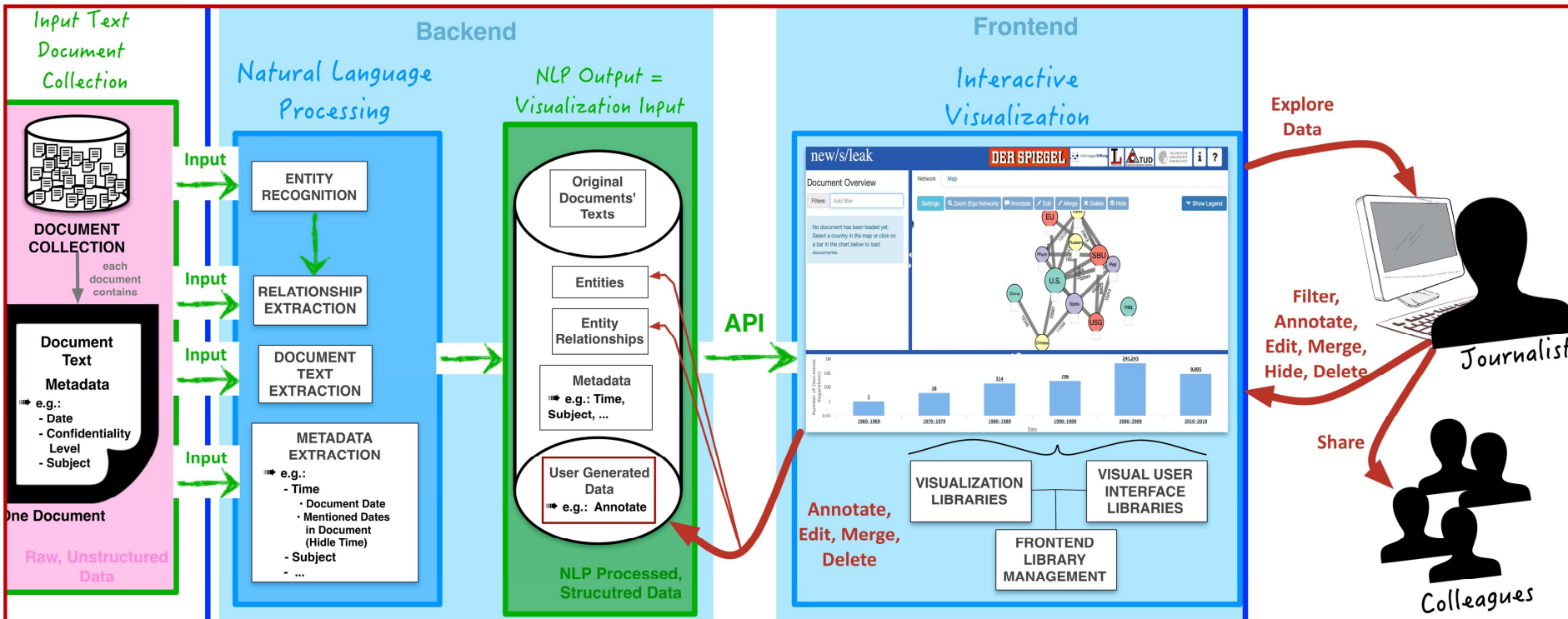
SOURCES		RELATIONS
Label	Neigh	
—	Trump	
—	Donald Trump	
—	Flake	
—	Republican	
—	Republican party	

»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 —

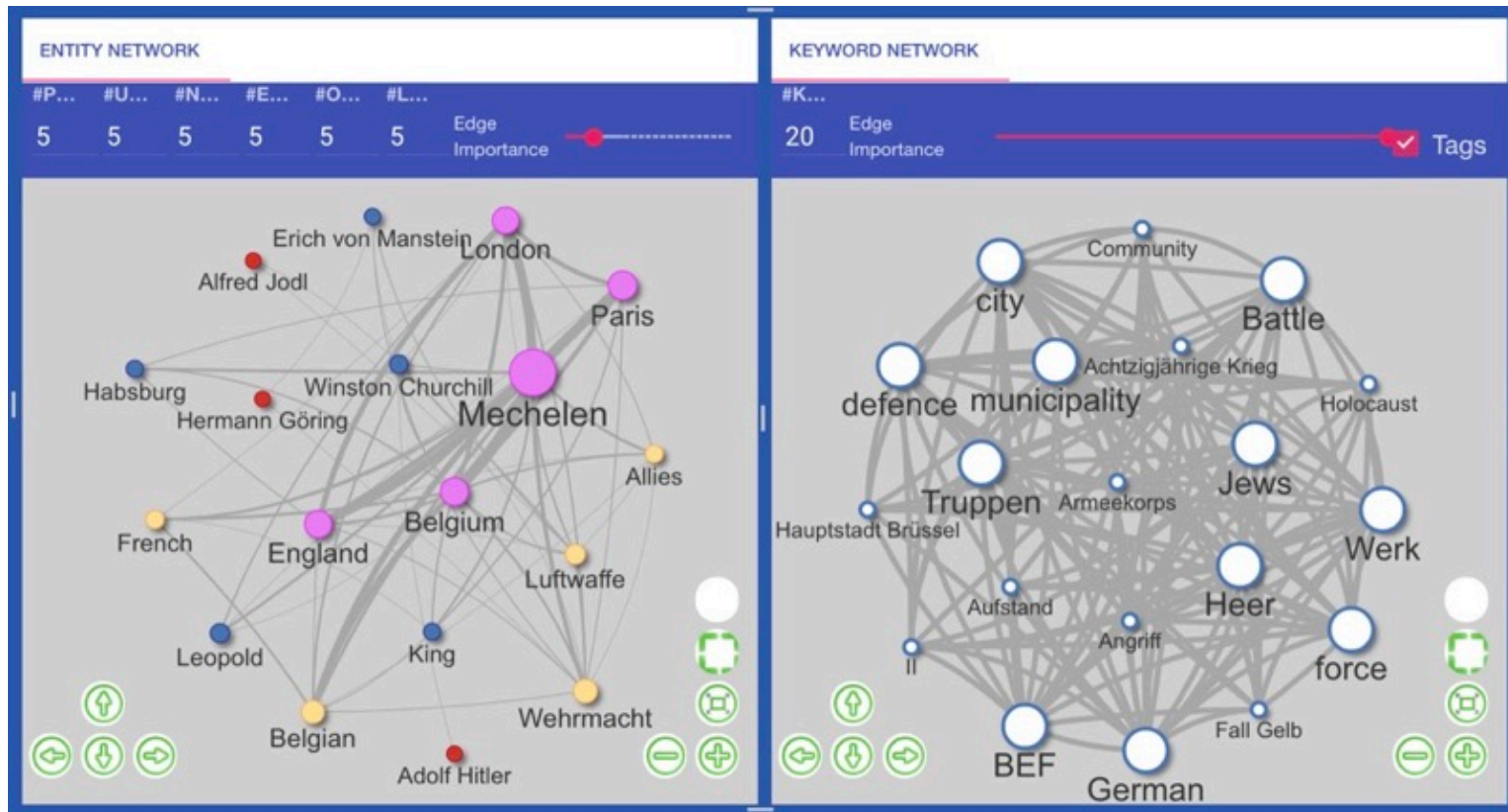
Please enter a label for the relation...

- WEBPAGES ARE CACHED LOCALLY
- USER BUILDS ITS OWN KNOWLEDGE GRAPH



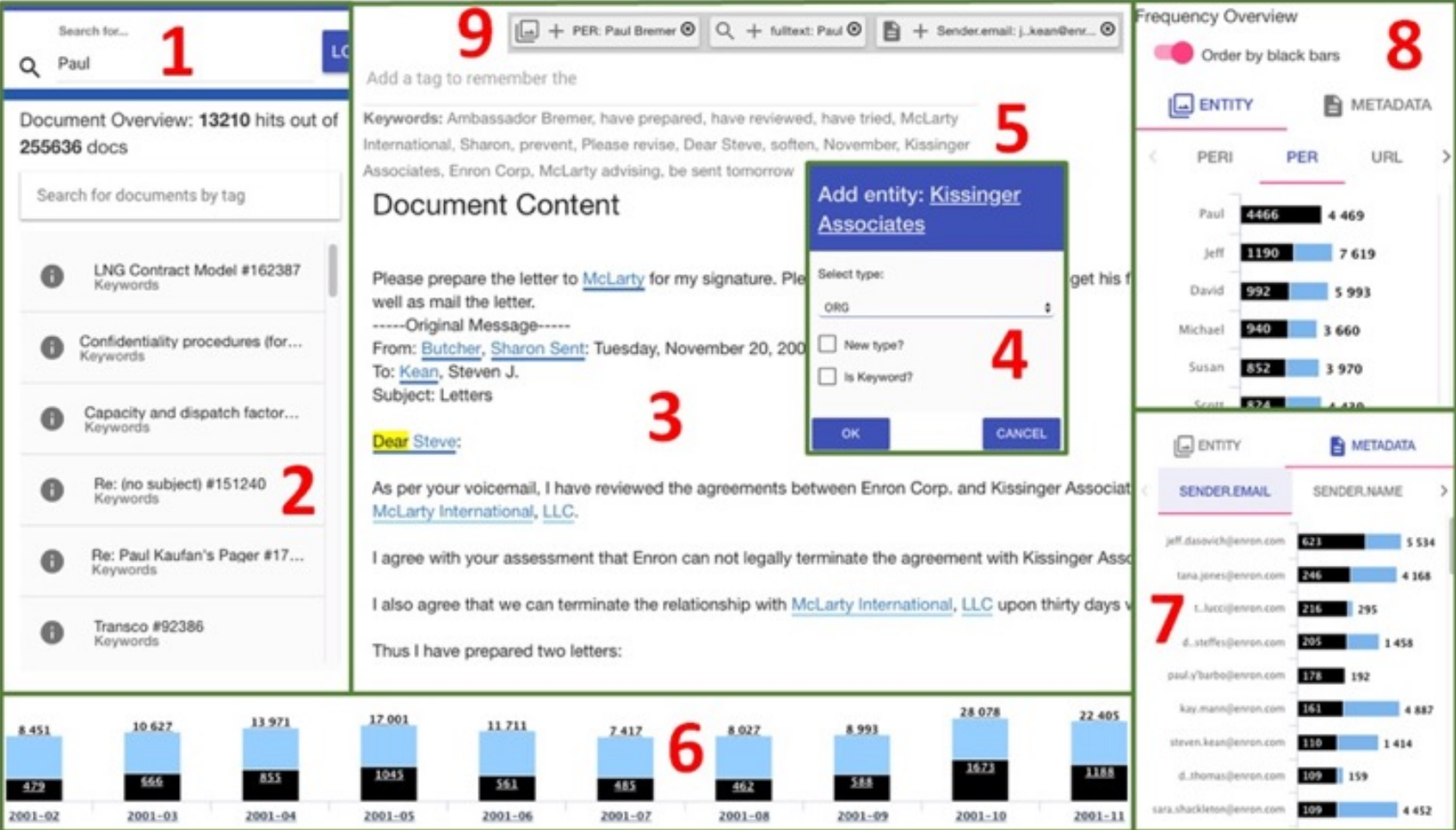
NEW/S/LEAK

ENTITY- AND KEYWORD-CENTRIC VIEW



■ ... AND EVERYTHING TRIGGERS A FILTER

NEW/S/LEAK: FACETED SEARCH AND FULL-TEXT VIEW



1 Search for... Paul

2 Document Overview: 13210 hits out of 255636 docs

3 Document Content

4 Add entity: Kissinger Associates

5 Keywords: Ambassador Bremer, have prepared, have reviewed, have tried, McLarty International, Sharon, prevent, Please revise, Dear Steve, soften, November, Kissinger Associates, Enron Corp, McLarty advising, be sent tomorrow

6 Frequency Overview (Bottom)

7 Frequency Overview (Right)

8 Frequency Overview (Top Right)

9 Add a tag to remember the

Document Content Text:

Please prepare the letter to [McLarty](#) for my signature. Please well as mail the letter.

-----Original Message-----

From: [Butcher, Sharon Sent](#): Tuesday, November 20, 2001

To: [Kean, Steven J.](#)

Subject: Letters

Dear Steve:

As per your voicemail, I have reviewed the agreements between Enron Corp. and Kissinger Associates [McLarty International, LLC](#).

I agree with your assessment that Enron can not legally terminate the agreement with Kissinger Associates.

I also agree that we can terminate the relationship with [McLarty International, LLC](#) upon thirty days written notice.

Thus I have prepared two letters:

Frequency Overview (Bottom) Data:

Year	Count
2001-02	8 451
2001-03	10 627
2001-04	13 971
2001-05	17 001
2001-06	11 711
2001-07	7 417
2001-08	8 027
2001-09	8 993
2001-10	28 078
2001-11	22 405

Frequency Overview (Right) Data:

Entity	Count
Paul	4 466
Jeff	1 190
David	992
Michael	940
Susan	852
Scott	874

Frequency Overview (Bottom Right) Data:

Sender Email	Count
jeff.dasovich@enron.com	621
tana.jones@enron.com	246
t.lucci@enron.com	216
d.steffes@enron.com	205
paul.ybarbo@enron.com	178
kay.mann@enron.com	161
steven.kean@enron.com	110
d.thomas@enron.com	109
tara.shackleton@enron.com	109

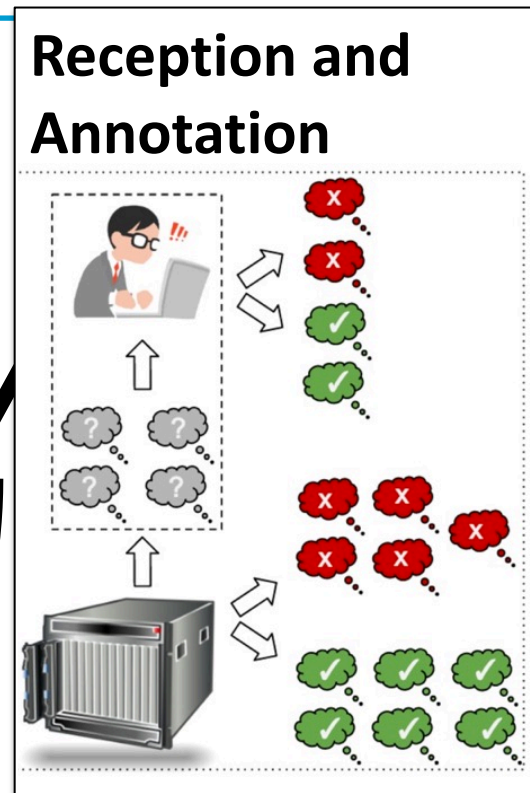
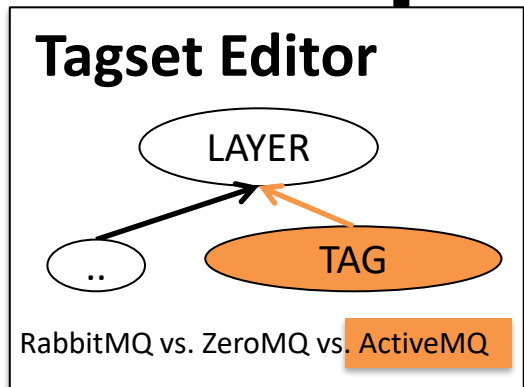
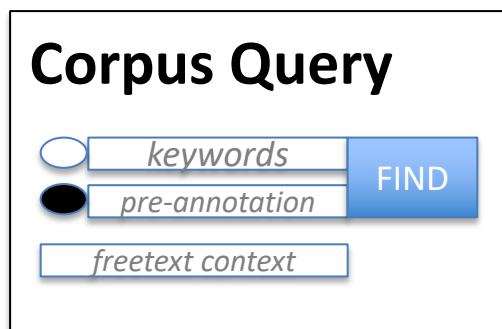
COMPSCI CHALLENGES WITH INTERACTIVE LEARNING

- User Motivation
 - Tooling and GUI is user-facing, must be intuitive and robust to support users with their task
 - For explicit user signals: user must see the immediate effect on improvement
 - user shall have a reasonably accurate mental model of the process
- Formal Evaluation and Reproducibility
 - Need to aggregate over a range of users
 - Comparison is difficult even if user judgments are made public, since changes in the automatic part might lead to a different behavior, cf. IR Evaluation Campaigns

COMPSCI PREMISES OF INCREMENTAL LEARNING

- Handling concept drift: Changes in the stream are quickly reflected in system behavior.
Research question: Recency and forgetting?
- Personalization/model adaptation: True personal assistants that improve in aspects that are important to individual users.
Research question: Mixing global and private signals under privacy constraints?
- User incentives and transparency: System adopts to users in a predictable fashion
Research question: How to make system decisions predictable and transparent? How to perform learning in absence of huge training data?

HUMANIST IN THE LOOP (HERMENEUTIC SPIRAL)



ABSTRACT

Title:

Interactive Entity Networks for Information Access to Text

Abstract:

In this talk, I will show some examples from language technology on how to extract and visualize entity networks from text in order to represent their interaction. A special emphasis will be set on interactivity and provenance to enable users to adapt the network towards their (research) interests in an informed fashion that is firmly rooted in the data. Partially rooted in machine learning techniques already, I will discuss their application in such rather personalized settings for accelerating the process of annotated entity network creation.

Bio:

Chris Biemann obtained his doctorate in Computer Science / Natural Language Processing in 2007 from the University of Leipzig, before joining the San-Francisco-based semantic search startup Powerset, which was acquired by Microsoft to form the Bing.com search engine. In 2011, he got appointed as assistant professor for language technology in the computer science department at TU Darmstadt; since October 2016, Chris is professor for language technology at the University of Hamburg in Germany. His current research is focused on adaptive natural language processing, web-scale statistical semantic methods, machine learning from crowdsourcing signals and on applications in the humanities and social sciences. He has co-authored about 200 publications in the field of natural language processing. His research group regularly releases open datasets and open source software to the community.