







STEFFEN REMUS, MANUEL KAUFMANN, KATHRIN BALLWEG, TATIANA VON LANDESBERGER, CHRIS BIEMANN

# STORYFINDER: PERSONALIZED KNOWLEDGE BASE CONSTRUCTION AND MANAGEMENT BY BROWSING THE WEB

HTTPS://UHH-LT.GITHUB.IO/STORYFINDER



Several Republicans serving in the Senior Republican Bob House of Representatives have Corker feuds with Trump: 'He's 'debasing David Trott of Michigan, our nation David Reichert of Washington

......

Read more

Ileana Ros-Lehtinen of Florida, leaving an opening for Democrats in competitive districts.

By contrast, moderate Republican Susan Collins said earlier this month

in swing states.



# **ABSTRACT**

• We present **STORYFINDER**, a user application which consists of **a browser plugin** and a web server backend with the goal to **highlight** and **manage** the





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 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 that Susan Collins is not up for re-election in 2018

Figure: Highlight named entities and show their relations in a graph

## INTRODUCTION

**INFORMATION EXTRACTION** 

**KNOWLEDGE** & MANAGEMENT

#### Information is vast and broadcast is fast:

 The web with its sheer unmeasurable speed of broadcasting new data and its vast quantity of available knowledge is the first choice for information seekers. It is a user's privilege to read or skim a webpage or bookmark it for later reference, but considering that the human memory can be deceptive, it also is a **user's obligation** to keep information ordered and easily accessible for later reference.

#### **Manual instrumentation methods exist:**

- Concept maps, or mind maps provide the necessary methodology and have been implemented in a multitude of prolific, computerized toolkits, which go beyond simple bookmarking.
  - $\rightarrow$  Manual compilation is not maintainable

**information** contained in web pages by combining techniques from natural language processing and visual analytics. Named entities and keywords are extracted and stored per web page and used in a graph of entities and their relations. The scope can be set to a single web page, multiple web pages, or the complete set of analyzed web pages in a user's history.

## **SCHEMA**

#### Web browser plugin:

- Listens and reacts to user events
- Initiates analysis on the server
- Provides side pane view of interactive webpage

#### Server backend:

- Analyzes webpage
- Extracts metadata: Named entities, keywords

#### Interactive webpage:

- Provides graph view of articles
- Adheres to standards of Visual Analytics



Figure: Search for named entities

## **FRONT- & BACKEND INTERACTION**



#### Information is fully editable:

- Add or remove named entities
- Add or remove relations
- Label primary relations, i.e. generally valid
- Label secondary relations, i.e. valid for a particular

#### sentence

Corke run n

Trum

Senat

Repu

docker

ook Up "Corker"	× New element SAVE
	K Name: Corker
AdBlock	Please select a type: Person



### Managing knowledge:

- Induce taxonomic or ontological knowledge, i.e. general relations between concepts, purely from text
- Identify named entities and relations between them and show them in a so-called network of named entities
- Magpie or ESpotter highlight named entities in websites
- With **STORYFINDER**, we aim to support the user to quickly grasp the key concepts of a webpage, make it easily accessible for later usage, and put the new information into relation with previously visited web pages.
  - a) highlighting within webpage
  - separate graph-based view with entities and **b**) their relations
  - $\rightarrow$  grasp the "bigger picture"

Figure: Show graph for a group of pages

## **TECHNOLOGY STACK**

 Docker containers are used for NLP components and **Database integration**  $\rightarrow$  ensures scalability



- Websockets technology is used for communication from server to client, i.e. the interactive webpage
- $\rightarrow$  enables asynchronous events

÷	Jeff Fla	ake		
	SOURCES	RELATIONS		
	Label	Neighbor		
	-	Trump		
	_	Donald Tru	mp	

LANGUAGE TECHNOLOGY GROUP **UNIVERSITY OF HAMBURG** 

uhh-lt.github.io/storyfinder/

