

„Ja, jetzt ist das langweilig. Aber in zwanzig Jahren!“ –

Bereitstellung, Zugang und Analyse literarischer Blogs am Beispiel des Techniktagebuchs

André Blessing¹, Jan Hess², Kerstin Jung¹, Nicolas Schenk²

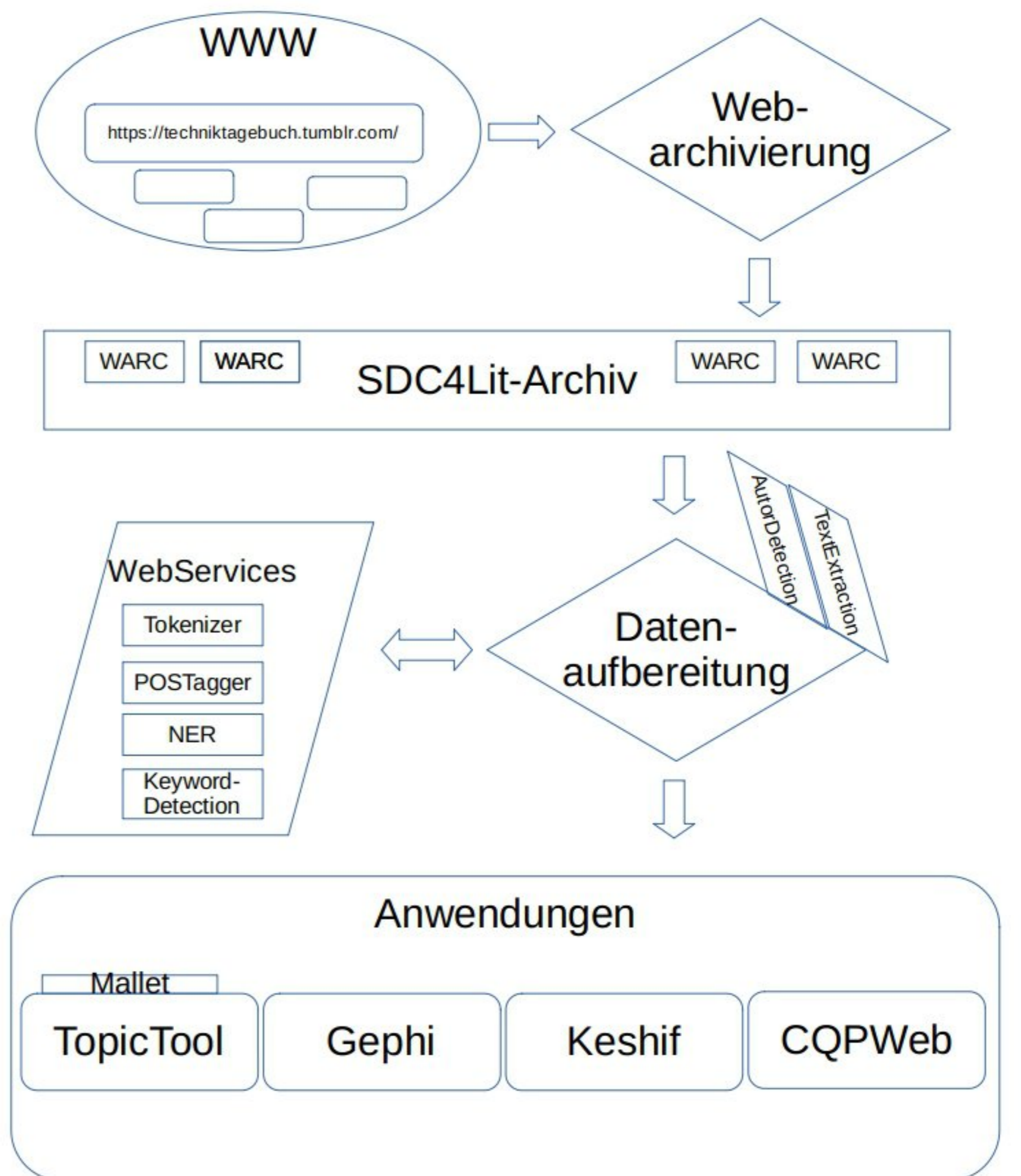
¹ Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

² Deutsches Literaturarchiv Marbach

<https://www.sdc4lit.de>

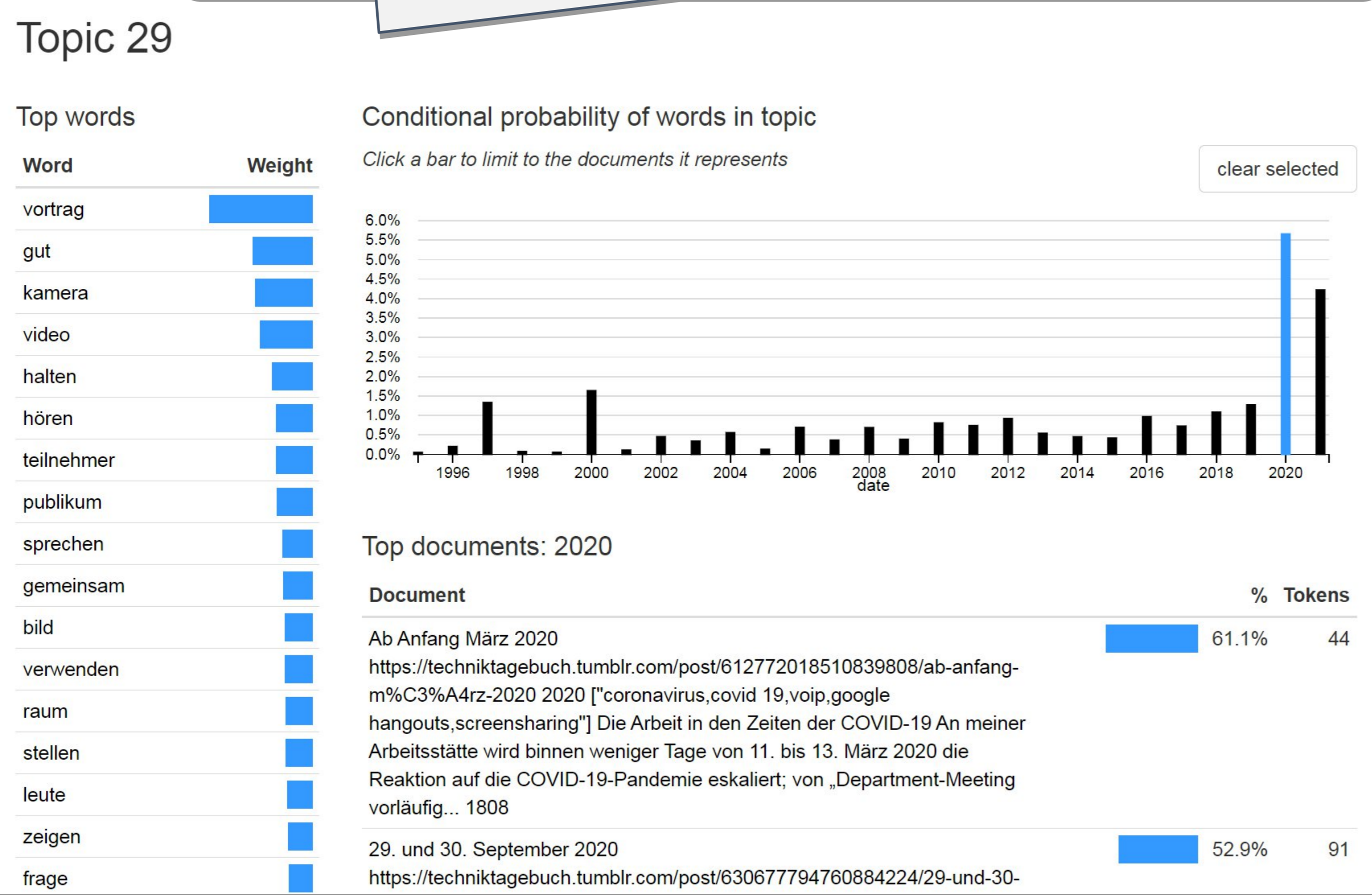
Techniktagebuch

- 2014 von der Schriftstellerin Kathrin Passig ins Leben gerufenes literarisches Blog
- <https://techniktagebuch.tumblr.com>
- Über 60 regelmäßige Autor:innen bloggen über ihre Erfahrungen und (Miss-)Erfolge mit technischen Themen und Gegenständen.
- Besonderheit: retrospektive Einträge sind mittels Rückdatierung nach dem Zeitpunkt der erzählten Erfahrung sortiert
- Blog-Crawl (ca. 9 GB) aus Sammlung des DLA Marbach
- im WARC-Format (Standard zur Archivierung von Webinhalten)
- umfasst mehr als 7000 Blogbeiträge (Stand: 18.5.2021)



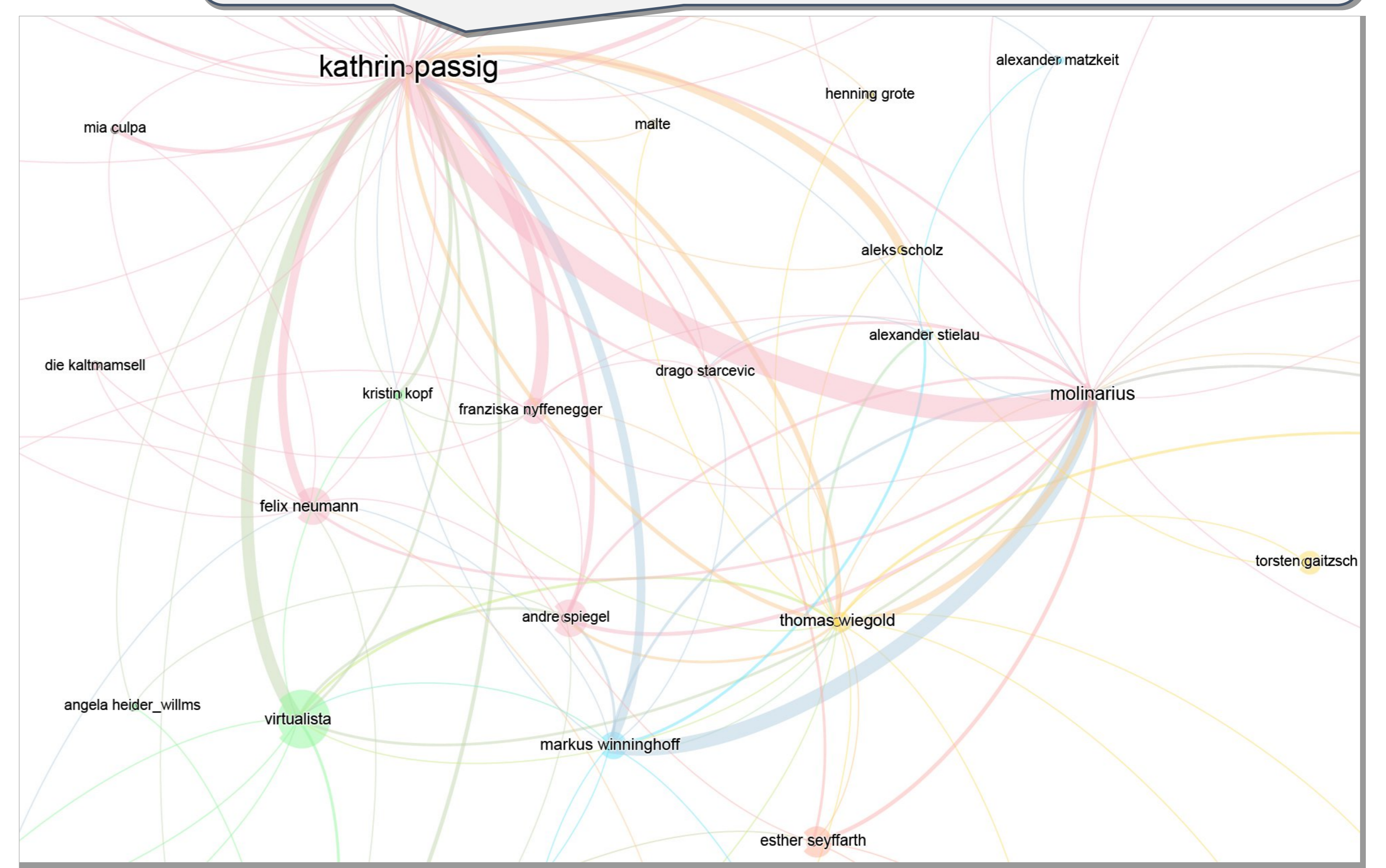
Verarbeitung durch die SDC4Lit-Infrastruktur

Lassen sich thematische 'Trends' erkennen?



Topic-Analyse mit dem Topic Model Browser.

Welche Autor:innen nehmen innerhalb des Blogs aufeinander Bezug?



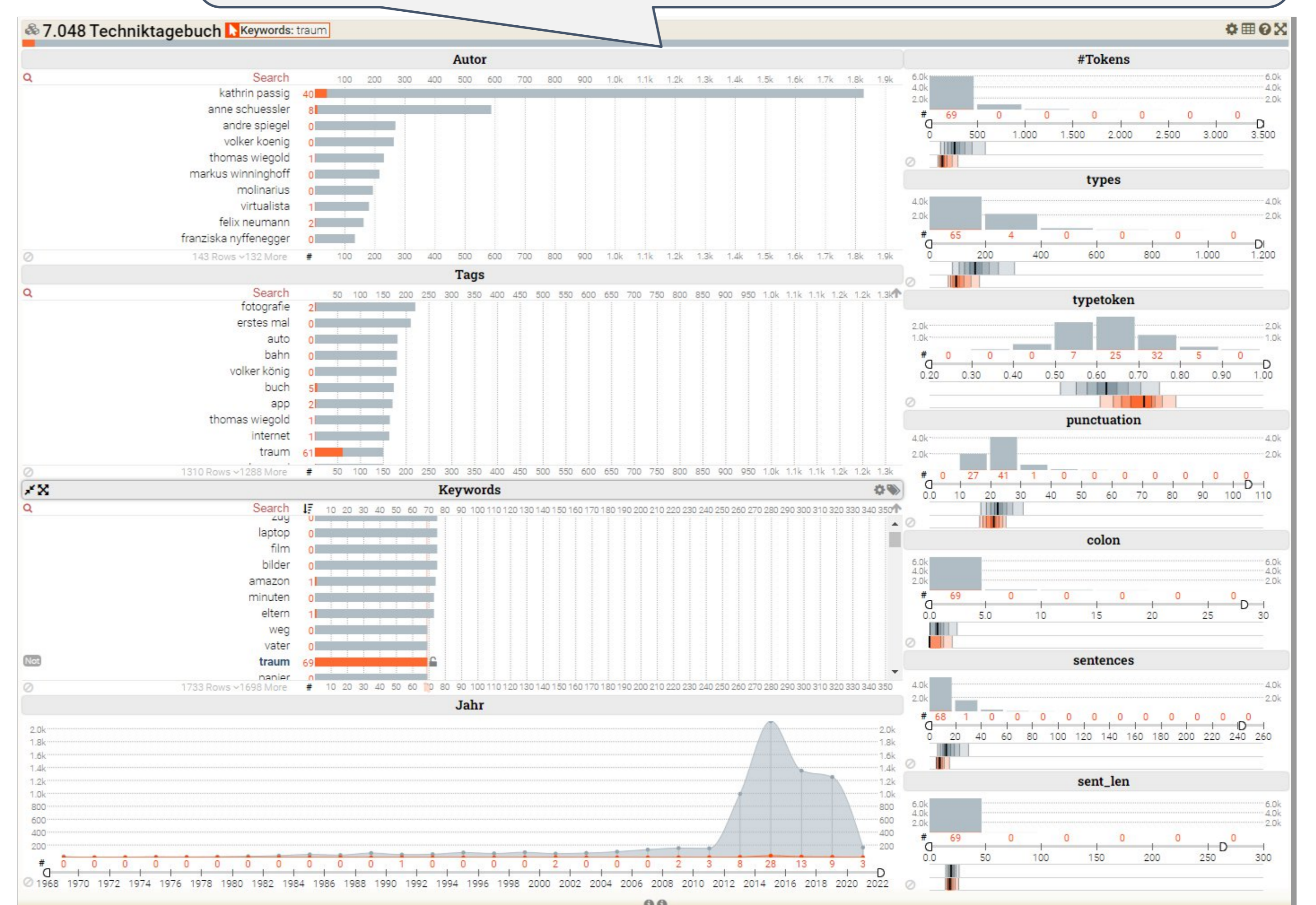
Visualisierung der bloginternen Links mittels Gephi.

Welche Wörter benutzt Kathrin Passig häufiger als die anderen TTB-Autor:innen?

Passig: unique words	others: unique words	Passig: more frequent words (minFreq=1)	others: more frequent words (minFreq=1)
1 bilch	bislang	Skypelog	Webseite
2 mjutabor	etliche	says	dennoch
3 lurch	üblicherweise	Jan	Mirus
4 Nodrog	Arbeitgeber	Touch&Travel	Neumann
5 Fatboyz	nervig	Kameramann	Rezept
6 klezk	glücklicherweise	fuer	zuhause
7 Oatridge	bisweilen	Heizdecke	rasch
8 Mewbourn	Roboter	Regensburg	Gegensatz
9 Moritz	gängig	Zettelkästen	sodass
10 Kirkuba	scottish	Leinen	anschließend
11 katzenfrett	Hörspiel	Passig	Note
12 mioumiou	fortlaufend	Tesco	deswegen
13 Habakuk	Gattin	Sascha	Drago
14 fvoerster	Theke	Crail	zuvor
15 isnowde	stoppen	Googledoc	Gaitsch
16 Munibert	mitunter	hihi	Starcevic
17 testbilch	anderweitig	Lobo	Meldung
18 IRC-Log	Sitzung	Kolumne	Alan
19 Coworkingspace	Parkschein	vorausschauend	Lied
20 Handyinternet	dummerweise	Siemens	Sohn

Wortfrequenzliste nach CQPWeb.

Lassen sich Zusammenhänge zwischen Inhalt und Form/Sprache erkennen?



Explorative Auswertung mit Keshif.