

SDC4Lit as a Hub for Digital Literature Research

Prof. Dr. Roland S. Kamzelak, Deutsches Literaturarchiv Marbach
Jan Hess, Deutsches Literaturarchiv Marbach

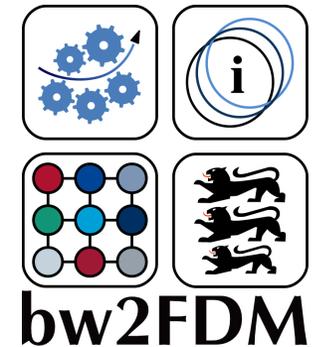
Tübingen Research Data Management Days, July 14 and 15, 2022

Science Data Centers in BW

Science Data Centers in Baden-Württemberg:



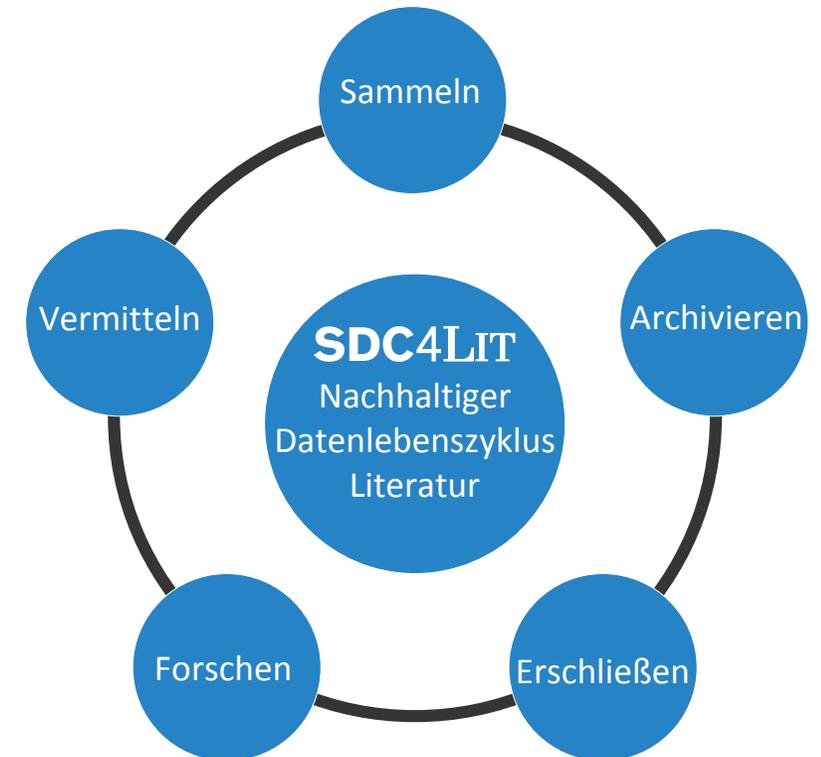
Coordination of the cross-sectional topics
of the four Science Data Centers in BW:



Project Goals

Collecting, archiving, indexing, researching and sharing literary sources using modern digital methods:

- Modelling born-digitals and literature on the web;
- development of computer-based analysis methods;
- research and education-oriented development;
- Integration of sources, methods, tools and best practices;
- sustainable data management according to FAIR principles.



Project Participants

German Literature Archive Marbach (DLA)

- Prof. Dr. Roland S. Kamzelak (head), Heinz Werner Kramski
- Jan Hess, Alex Holz, Nicolas Schenk, Mona Ulrich

High Performance Computing Center Stuttgart (HLRS)

- Dr. Thomas Bönisch, Dr. Andreas Kaminski
- Nina Buck, Andreas Ganzenmüller, Volodymyr Kushnarenko, Björn Schembera

Institute for Literary Studies/Department of Digital Humanities, University of Stuttgart (ILW/DH)

- Prof. Dr. Gabriel Viehhauser
- Pascal Hein, Claus-Michael Schlesinger

Institute for Natural Language Processing, University of Stuttgart (IMS)

- Prof. Dr. Jonas Kuhn
- André Blessing, Kerstin Jung

Funding

Funder:

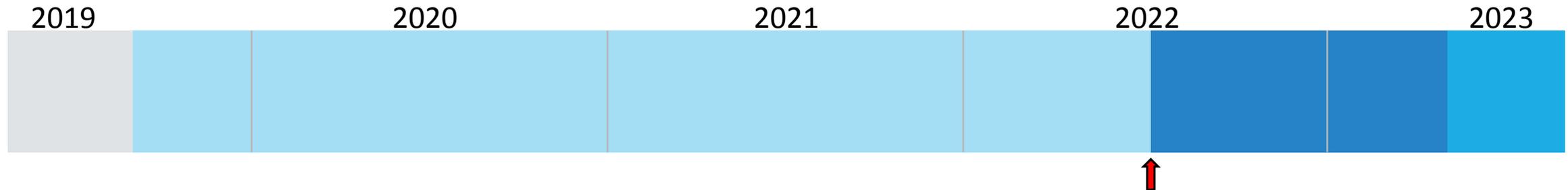
Baden-Württemberg Ministry of Science, Research and Arts

Duration of funding:

48 months

Start: 01.05.2019

End: 31.08.2023 (initially: 30.04.2023)

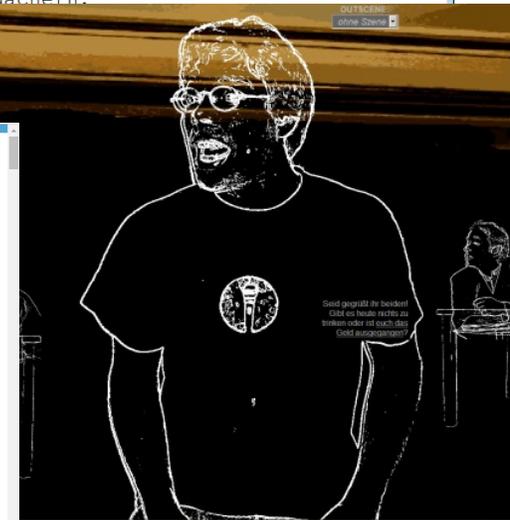
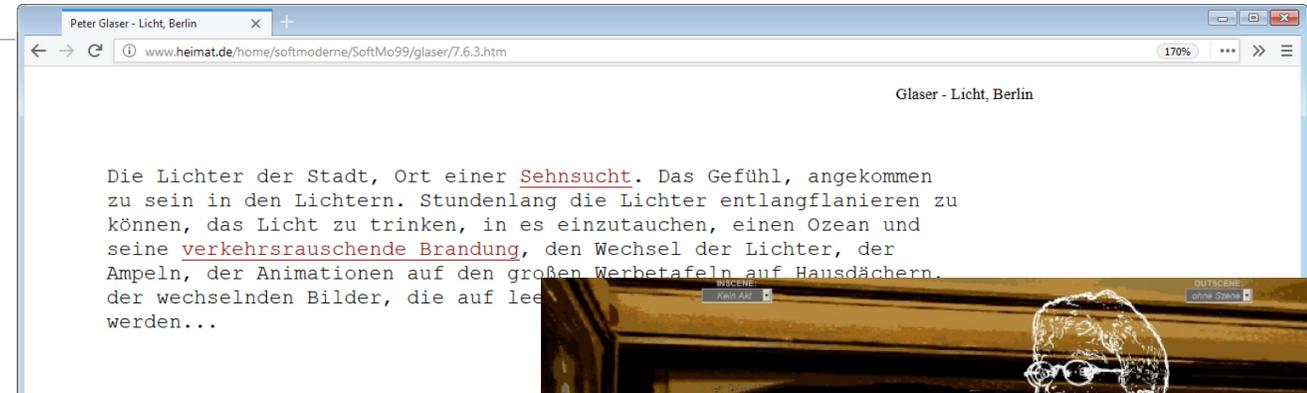


Connections to NFDI



Research Subject: Literature on the Web

- Collected by the library of the DLA;
- Web literature, literary weblogs, literary online journals;
- Currently approx. 500 objects;
- Different representations for preservation: WARC files, screenshots, screencasts, source code;
- Previously accessible via the SWBContent Literature on the web platform, soon via the SDC4Lit Repo.



Collage consisting of screenshots from:
 Peter Glaser: *Licht, Berlin*, <http://www.heimat.de/home/softmoderne/SoftMo99/glaser/7.6.3.htm>;
 Frank Klötgen: *Endlose Liebe – Endless Love*, <http://www.hirponoma.de/trashical/>;
 Techniktagebuch (Autorenkollektiv), <https://techniktagebuch.tumblr.com/>;
 (14/7/2022).

Extension Stages

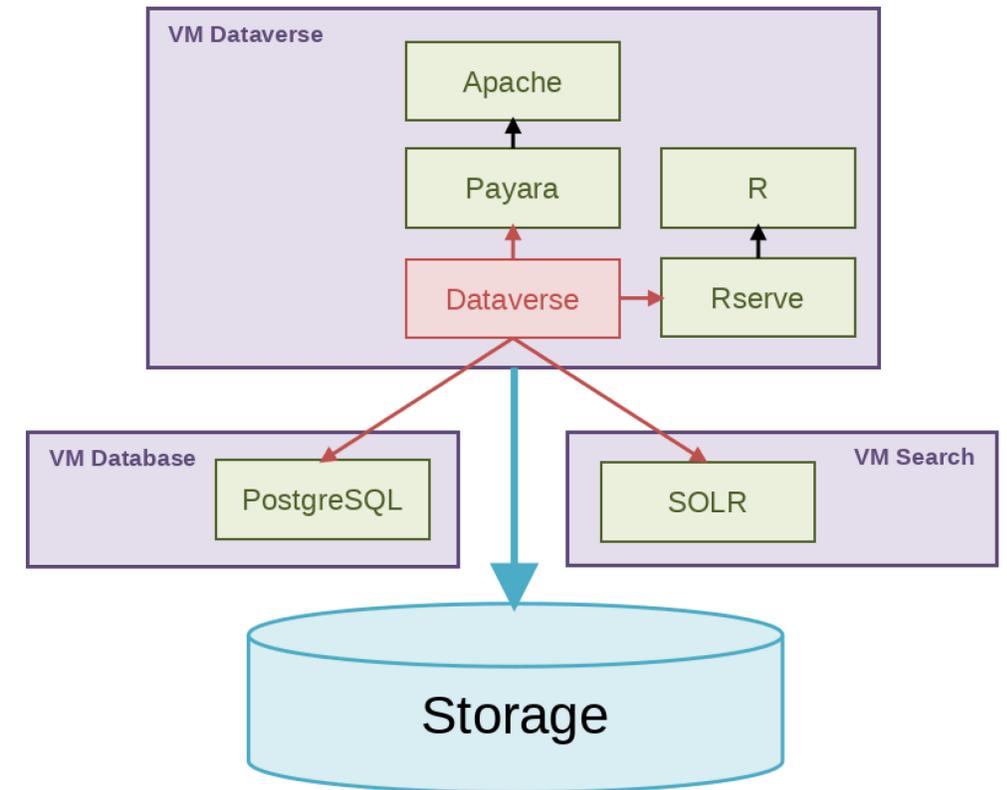
- Offline born-digitals (currently approx. 80 pre-estates and estates with digital components at the DLA);
- digitized literature;
- (literary) computer games.
- ...



Photo: DLA

SDC4Lit Repository

- Repository Software: Dataverse
- Prototypical installation
 - Physical server, virtualisation with Proxmox;
 - 18 TB RAID5 local hard disks, connection of ext. storage;
 - backup to tape storage system at HLRS.
- Several VMs for connected tools.



Welcome to SDC4Lit repository prototype!

DLA

(SDC4Lit)

SDC4Lit collections >

Contact Share

Navigation carousel with icons and labels: Computerspiele, Digitale Vor- und Nachlässe, Literatur im Netz, Literatur im Netz 20220602

Search this dataverse...



Advanced Search

Dataverses (42)

Datasets (47)

Files (330)

Dataverse Category

Department (27)

Laboratory (15)

Publication Year

2022 (53)

2021 (36)

Classification

Netzliteratur (2)

○○○○○ (4)

1 to 10 of 89 Results

Sort ▾

Licht, Berlin, Repräsentationen, 07.07.2015
Jul 6, 2022 - Licht, Berlin
Peter Glaser, 2022, "Licht, Berlin, Repräsentationen, 07.07.2015", <https://doi.org/10.82013/SDC4Lit-proto-1332>, SDC4Lit collections, V1
Repräsentationen, 07.07.2015

Licht, Berlin, Metadata
Jul 6, 2022 - Licht, Berlin
Glaser, Peter, 2022, "Licht, Berlin, Metadata", <https://doi.org/10.82013/SDC4Lit-proto-1331>, SDC4Lit collections, V1
Metadata

Licht, Berlin

SolrWayback



SOLRWAYBACK

Enter search term ? 

GROUPED SEARCH [?] IMAGE SEARCH [?] URL SEARCH [?]

[SEARCH WITH UPLOADED FILE](#) [GPS IMAGE SEARCH](#)  [TOOLBOX](#) 

[About this archive](#)

Test instance of SolrWayback
(<https://github.com/netarchivesuite/solrwayback>;
14/7/2022) with SDC4Lit test data.

SolrWayback



SOLRWAYBACK

GROUPED SEARCH [?]

IMAGE SEARCH [?] URL SEARCH [?]

FACETS

domain

- tumblr.com (152,210)
- merkur-zeitschrift.de (9,996)
- riesenmaschine.de (3,235)
- brovserboy.de (1,026)
- wordsonavatch.blogspot.de (131)
- antville.org (52)
- mequito.org (39)
- claudiakilian.de (36)
- buddenbohm-und-soehne.de (30)
- tvoday.net (30)

more domains

source_file

- web-20160428145119075-00002-20133-dl06-844... (14,797)
- web-20161104185334233-00001-1612-dl06-8444... (14,715)
- web-20171129114005303-00002-21114-dl06.bsz-... (12,951)
- web-20170814175628089-00001-6034-dl06.bsz-b... (12,798)
- web-20171129225148642-00003-21114-dl06.bsz-... (11,874)
- web-20160428034808528-00001-20133-dl06-844... (10,752)
- web-20160728205226795-00001-22818-dl06-844... (9,622)
- web-20160728071904737-00000-22818-dl06-844... (7,832)
- web-20161104082510462-00000-1612-dl06-8444... (7,676)
- web-20170815094647974-00003-6034-dl06.bsz-b... (7,440)

more source_files

RESULTS

Showing 0 - 20 of 167,058 entries matching query.

[See available export options](#)

RIESENMASCHINE - KATHRIN PASSIG [🔗](#) #1
score: 89.85333

Type: html, web page @ riesenmaschine.de
Date: 05/03-2015
Url: http://riesenmaschine.de/index.html?autor=Kathrin_Passig&seite=37

Highlighted content:
"Füssen getragen werden, denn dort muss Platz für die Five Fingers bleiben. Vorschläge? **Kathrin Passig**"

Images: showing 4 out of 10



RIESENMASCHINE - KATHRIN PASSIG [🔗](#) #2
score: 89.83753

Type: html, web page @ riesenmaschine.de
Date: 05/03-2015
Url: http://riesenmaschine.de/index.html?autor=Kathrin_Passig&seite=99

Highlighted content:
"weggeworfene Wäscheständer. **Kathrin Passig** | Dauerhafter Link | Kommentieren 05.06.2005 | 22:51"

Images: showing 4 out of 5



Test instance of SolrWayback
(<https://github.com/netarchivesuite/solrwayback>;
14/7/2022) with SDC4Lit test data.

RIESENWASCHMASCHINE » DAS BLITZSAUBERE UNIVERSUM «



26.10.2006 | 18:48 | Was fehlt | Sachen kaufen

KLEIDUNGSSCHULD UND KLEIDUNGSSÜHNE



Das Praktische und Bequeme ist, ach, nicht immer das, womit man sich auf die Strasse trauen darf. Wie schön wäre es, in den [Vibram Five Fingers](#) (in Deutschland z.B. bei Globetrotter als [Lizard Five Fingers](#) erhältlich) zu wohnen und in mindestens drei von vier Jahreszeiten Fuss- und Gummifischisten gleichzeitig glücklich zu machen ? aber es geht nicht, der Ehrverlust ist zu gross. Heimlich, im Dunkeln könnte man sie tragen und damit zu Hause auf Wänden und Decke herumlaufen, das ist aber auch keine Lösung. Was fehlt, ist ein Ausgleichkleidungsstück von solchem Adel, dass es etwa die Five Fingers, Socken in Sandalen oder auch eine Krawatte mit Cartoonfiguren ohne weiteres neutralisieren könnte wie das Magnesiumhydroxid die Magensäure. Es müsste entsetzlich teuer und ungemein vielseitig sein und darf nicht an den Füssen getragen werden, denn dort muss Platz für die Five Fingers bleiben. Vorschläge?

Bild: [vibramfivefingers](#)

Kathrin Passig | Dauerhafter Link | Kommentare (6)

25.10.2006 | 18:58 | Anderswo | Zeichen und Wunder

MISTTELEFON



Das Misttelefon (vormals [Schneetelefon](#)) ist eine vorbildliche Einrichtung der Stadt Wien. Es hat leider

● IN DER RIESENMASCHINE

- Impressum
- Die Wahrheit über die Riesenmaschine
- T-Shirts
- Riesenmaschine-Buch zum Gratis-Download
- In eigener Sache
- Alle Termine
- Kulturproduktion
- Preise
- Rundherum Supersache: RSS

● ORIENTIERUNG

- Suche:

 Senden

- Riesenmaschine Shuffle
- Alle Autoren



SolrWayback



SOLRWAYBACK

techniktagebuch

GROUPED SEARCH [?] IMAGE SEARCH [?] URL SEARCH [?]

SEARCH WITH UPLOADED FILE GPS IMAGE SEARCH TOOLBOX

FACETS

domain

- tumblr.com (997,903)
- wordpress.com (34,430)
- claudiakilian.de (16,704)
- blogger.de (3,887)
- goldfischli.blogspot.com (361)
- twitter.com (99)
- youtube.com (37)
- antville.org (26)
- buddenbohm-und-soehne.de (24)
- piaziefle.de (11)

more domains

source_file

- web-20170814175628089-00001-6034-dl06.bsz-b... (76,594)
- web-20160728071904737-00000-22818-dl06-844... (74,949)
- web-20160427145116163-00000-20133-dl06-844... (70,729)
- web-20170815094647974-00003-6034-dl06.bsz-b... (68,514)
- web-20171130114006071-00004-21114-dl06.bsz-... (68,503)
- web-20171128114010846-00000-21114-dl06.bsz-... (63,222)
- web-20161104082510462-00000-1612-dl06-8444... (61,425)
- web-20171128234118261-00001-21114-dl06.bsz-... (60,827)
- web-20171129225148642-00003-21114-dl06.bsz-... (59,296)
- web-20161104185334233-00001-1612-dl06-8444... (56,315)

more source_files

RESULTS

Showing 0 - 20 of 1,053,504 entries matching query.

[See available export options](#)

Previous 20 Next 20

TECHNIKTAGEBUCH (@TECHNIKTAGEBUCH) | ทวิตเตอร์ #1 score: 32.60351

Type: html, web page @ twitter.com
Date: 28/11-2017
Url: <https://twitter.com/techniktagebuch?lang=th>

Highlighted content:
"Techniktagebuch @techniktagebuch ทวิต ทวิต, หน้าปัจจุบัน 6,895 กำลังติดตาม กำลังติดตาม 1 ผู้ติดตาม ผู้ติดตาม"

Images: showing 4 out of 8 [See all images](#)



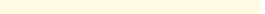
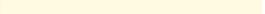
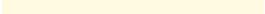
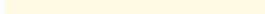
[View data fields](#)

TECHNIKTAGEBUCH (@TECHNIKTAGEBUCH) | TWITTER #2 score: 32.602608

Type: html, web page @ twitter.com
Date: 28/11-2017
Url: <https://twitter.com/techniktagebuch>

Highlighted content:
". Passwort vergessen? Neu bei Twitter? Registrieren Techniktagebuch @techniktagebuch Tweets Tweets, aktuelle"

Images: showing 4 out of 8 [See all images](#)



Test instance of SolrWayback
(<https://github.com/netarchivesuite/solrwayback>;
14/7/2022) with SDC4Lit test data.

SolrWayback

Date: 28/11-2017
Url: <https://twitter.com/technikagebuch?lang=eu>

Highlighted content:
"Technikagebuch @technikagebuch Txloak Tweets, current page. 6.895 Jarraitzen Jarraitzen 1 Jarraitzaileak"

Images: showing 4 out of 8 [See all images](#)



[View data fields](#) 

 **TECHNIKTAGEBUCH (@TECHNIKTAGEBUCH) | TWITTER** [↗](#) #5
score: 32.59433

Type: html, vweb page @ twitter.com
Date: 28/11-2017
Url: <https://twitter.com/technikagebuch?lang=ta>

Highlighted content:
"? புதுக்கணக்கை துவங்கு Technikagebuch @technikagebuch கீச்சுகள் தற்போதைய பக்கம், கீச்சுகள். 6,895"

Images: showing 4 out of 8 [See all images](#)



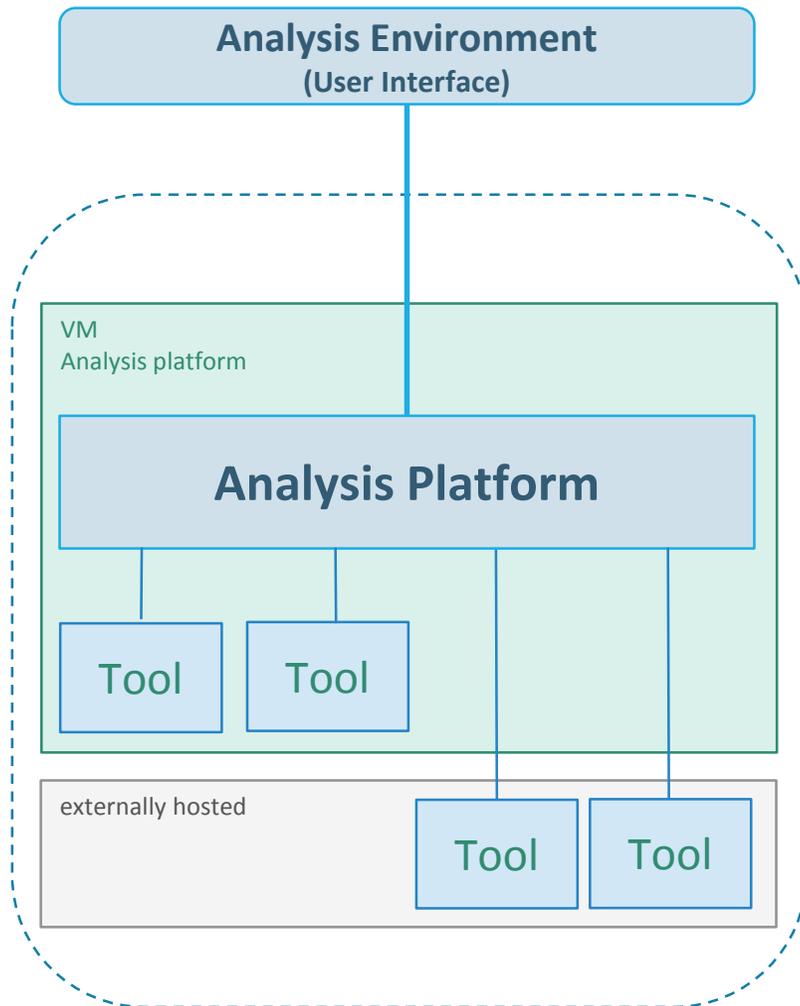
[View data fields](#) 

 **TECHNIKTAGEBUCH (@TECHNIKTAGEBUCH) | TWITTER** [↗](#) #6
score: 32.59433

Type: html, vweb page @ twitter.com
Date: 28/11-2017
Url: <https://twitter.com/technikagebuch?lang=kn>

Test instance of SolrWayback
(<https://github.com/netarchivesuite/solrwayback>;
14/7/2022) with SDC4Lit test data.

Analysis Environment



Goals of the Analysis Environment

- Testing and developing methods and tools of analysis in literary studies;
- bringing together and documenting low-threshold tools and providing them in the form of modular pipelines;
- documentation and teaching of provided tools and methodological competence.



VOR 6 TAGEN

3 NOTES

- #ANNE SCHÜSSLER
- #GLASFASER
- #INTERNET
- #DIGITALISIERUNG
- #INTERNETANBINDUNG
- #LANDLEBEN
- #STRASSENBAU

Juli 2022

Glasfaser is coming

Mitte März sind wir aufs Land gezogen. Das Land befindet sich noch am äußersten Rand des Kölner Speckgürtels, zum Beispiel fährt der Bus auch bis Mitternacht, es ist also nicht superschlimm, der nächste Supermarkt ist aber auch zwei Kilometer weit weg und wir wurden schon von einem Lieferdienst abgelehnt, weil wir zu weit weg wohnten (der Ort hat nur eine Postleitzahl, ist aber flächenmäßig sehr ausgedehnt, zu einer Freundin, die im gleichen Ort, aber am entgegengesetzten Ortsende wohnt, braucht man 15 Minuten mit dem Auto).

Als wir hinziehen, haben wir uns Internet eines sehr lokalen Internetanbieters organisiert, denn das große Magentaunternehmen bietet nur unterdurchschnittliche Downloadkapazitäten an. Mit dem Internet sind wir weitgehend zufrieden, man kann damit alles machen, inklusive übers Internet fernsehen, mehr brauche ich eigentlich nicht.

Im Juli stehen auf einmal Bagger in der Straße, reißen die Straße auf und verlegen dicke orangene Kabel. GLASFASER IS COMING!



Wenige Tage später steht ein Mann vom Straßenbau vor der Tür, wo denn bei uns der Anschluss gelegt werden soll. Wir wissen von nichts, weil wir auch nichts von der Telekom oder irgendwem gehört haben, äußern aber generelles

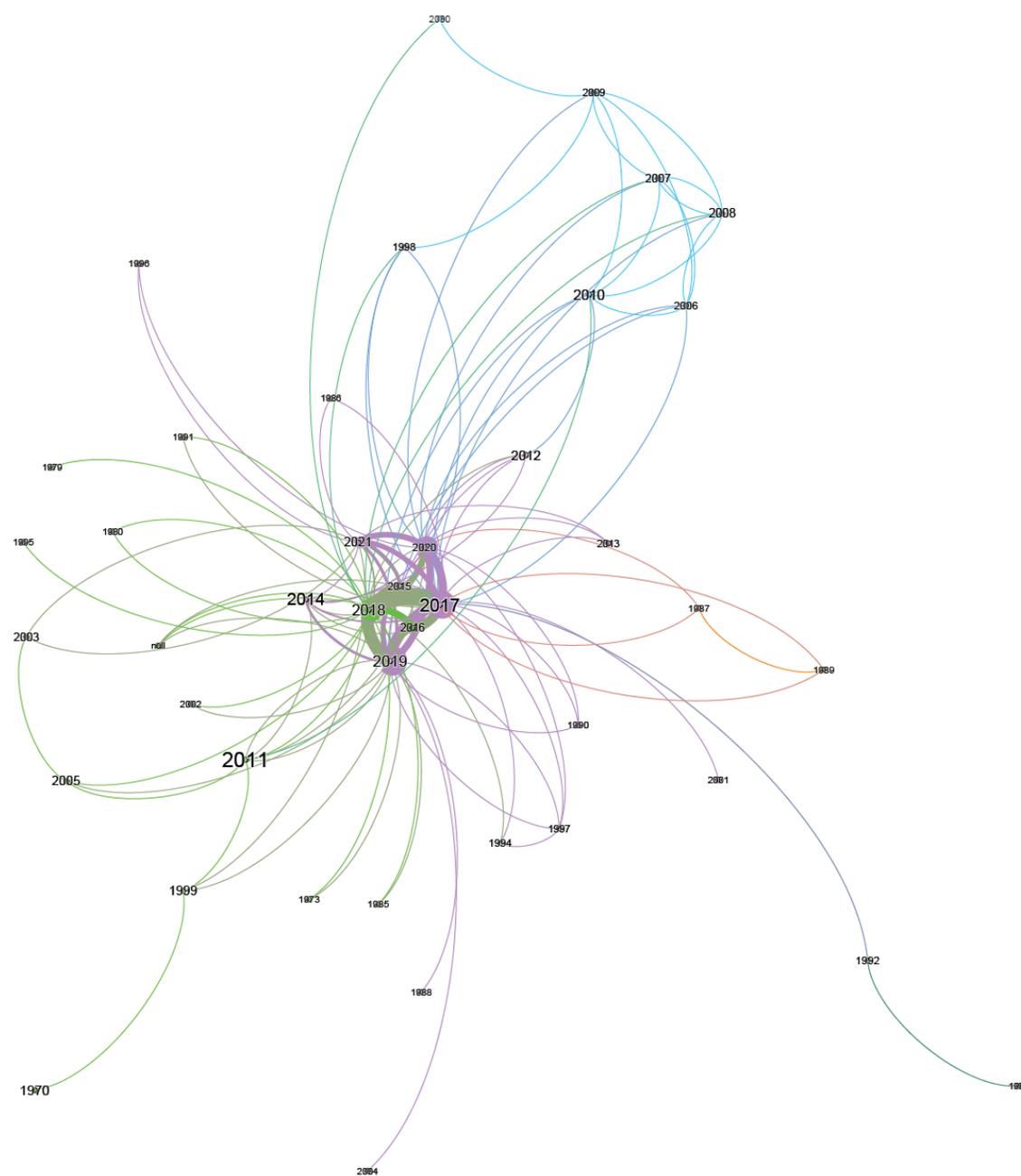


Ja, jetzt ist das langweilig.
Aber in zwanzig Jahren!

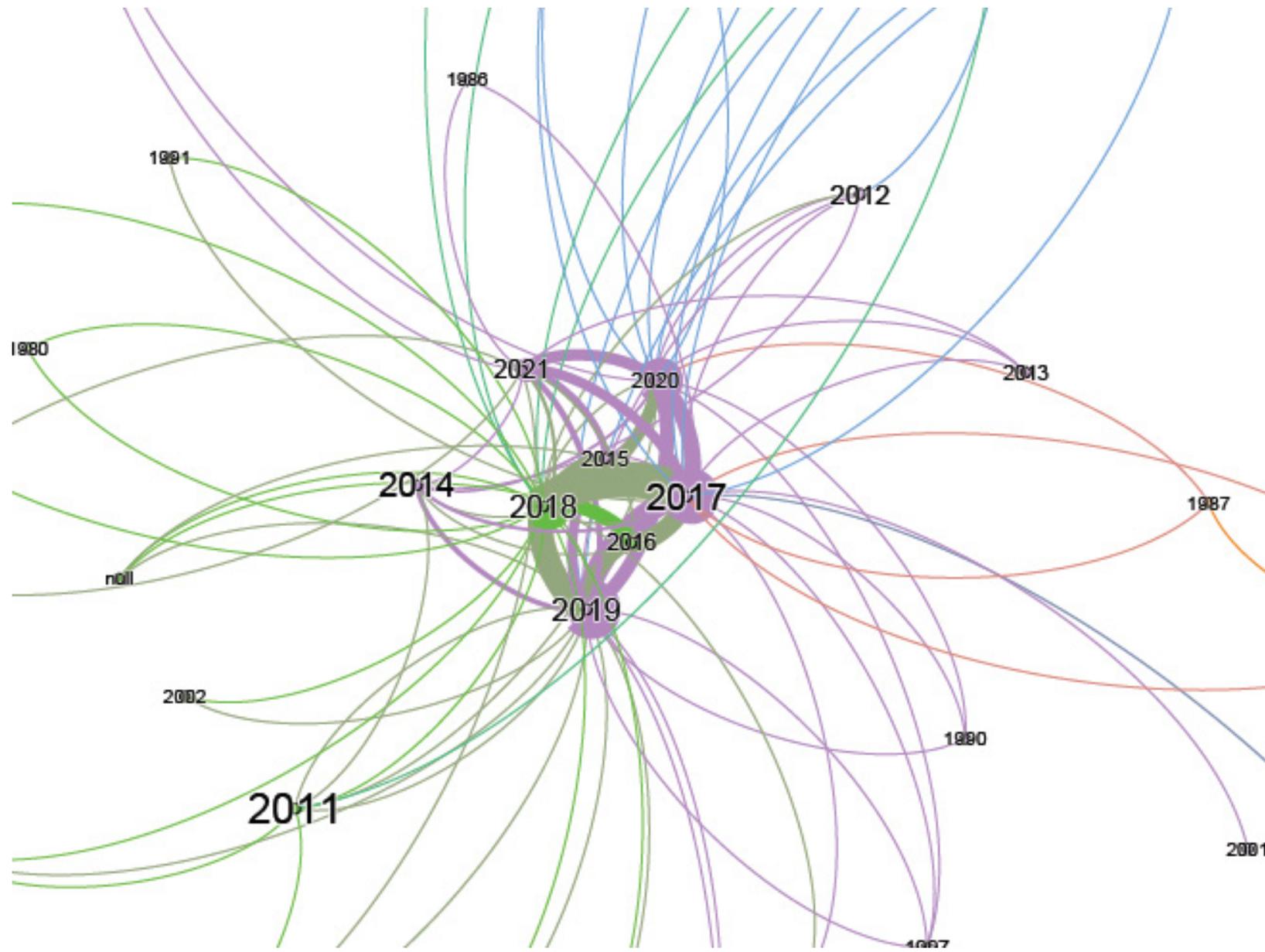
- ÜBER DAS
TECHNIKTAGEBUCH
- STICHWORTVERZEICHNIS
- AUSGEZEICHNET MIT
EINEM GRIMME ONLINE
AWARD 2019
- TECHNIKTAGEBUCH ALS
BUCH
- T-SHIRTS
- DATENSCHUTZ
- ➔ **GASTBEITRAG
SCHREIBEN**
- 📡 RSS
- 📅 ARCHIV
- 📅 ZUFALLSBEITRAG
- 📅 BEST OF

11.7.2022 Das Internet ist doch schlauer als die Menschen im Internet Ich bin bei der Physio und als ich an der Rez...
<https://t.co/gJUck05WwA>

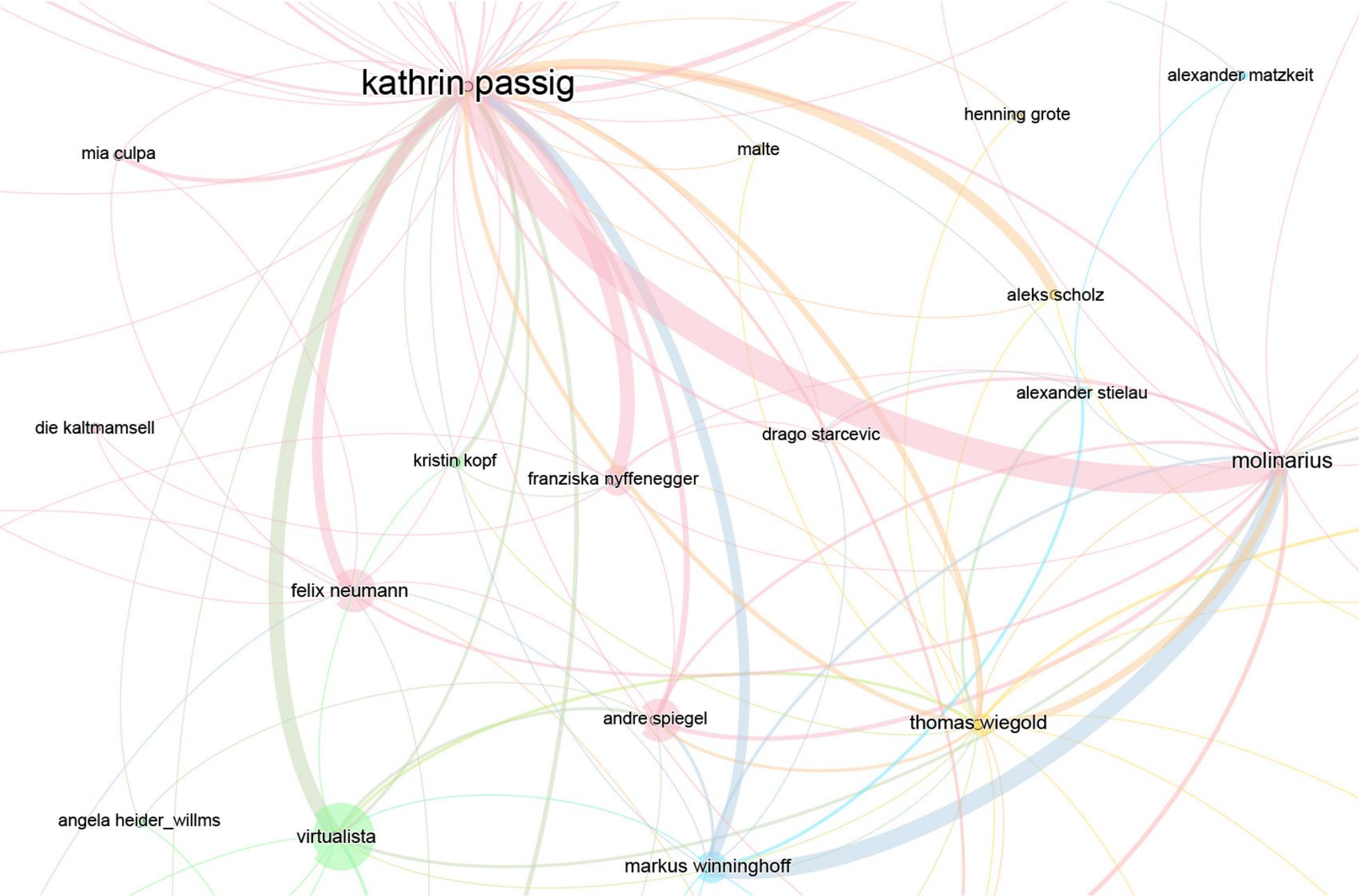
@TECHNIKTAGEBU...



Direct references (links) between TTB blogposts by reference year, visualized with Gephi (<https://gephi.org/>; 14/7/2022).



Direct references (links) between TTB blogposts by reference year visualized with Gephi (<https://gephi.org/>; 14/7/2022).



Direct links between TTB blogposts by author, visualized with Gephi (<https://gephi.org/>; 14/7/2022).



VOR 24 JAHREN

3 NOTES

#PC
#TURBOTASTE
#HUMAN ERROR
#KATHRIN PASSIG
#TOASTER
#AUFKNUSPERTASTE
#BEST OF

Irgendwann zwischen 1995 und 2000

Mein Rechner hat eine Aufknuspertaste

Mein Rechner hat eine Turbotaste zur Umschaltung des Prozessortakts von 8 MHz auf 16 MHz. [In der Wikipedia steht dazu](#): "Es kam vor, dass einige ältere Spiele auf seinerzeit modernen Computern zu schnell liefen oder Verzögerungsschleifen zu schnell abgearbeitet wurden. Die Turbo-Taste sorgte hier für Abhilfe." Da die Defaulteinstellung "schnell" ist, müsste die Turbo-Taste ja eigentlich eine Schnarch-Taste sein, aber weil das nicht so gut klingt, ist es umgekehrt, sie muss also gedrückt sein, wenn man die 16 MHz haben will.

Ich habe eigentlich keine Verwendung für die Turbotaste beziehungsweise ihre Schnarchfunktion, aber sie ist nun mal an diesem Rechner angebracht. [Ungefähr so muss man sich das vorstellen](#). Mein Rechner ist aber nicht der abgebildete, sondern hat ein Towergehäuse, das heißt, man stellt ihn hochkant unter den Schreibtisch. Mein Freund hat die Taste mit "Aufknuspertaste" beschriftet, eine Aufknuspertaste ist eine relativ aktuelle Neuerung auf dem Toastermarkt. Auf der im verlinkten Bild zu erahnenden [Segmentanzeige](#) sieht man, ob der Rechner gerade mit 8 oder 16 MHz arbeitet.

Ich klage seit ungefähr einem Jahr darüber, dass mein Rechner unerträglich lahm geworden ist und trage verschiedene Theorien vor, warum das so ist. Bis eines Tages ein aufmerksamer Freund unter meinem Schreibtisch guckt und "Wieso ist denn deine Turbotaste nicht gedrückt?" sagt. Ich muss wohl irgendwie mit dem Fuß an den Schalter gekommen sein, vor einem Jahr.

(Kathrin Passig)



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

ÜBER DAS
TECHNIKTAGEBUCH

STICHWORTVERZEICHNIS

AUSGEZEICHNET MIT
EINEM GRIMME ONLINE
AWARD 2019

TECHNIKTAGEBUCH ALS
BUCH

T-SHIRTS

DATENSCHUTZ

➔ [GASTBEITRAG
SCHREIBEN](#)

📡 RSS

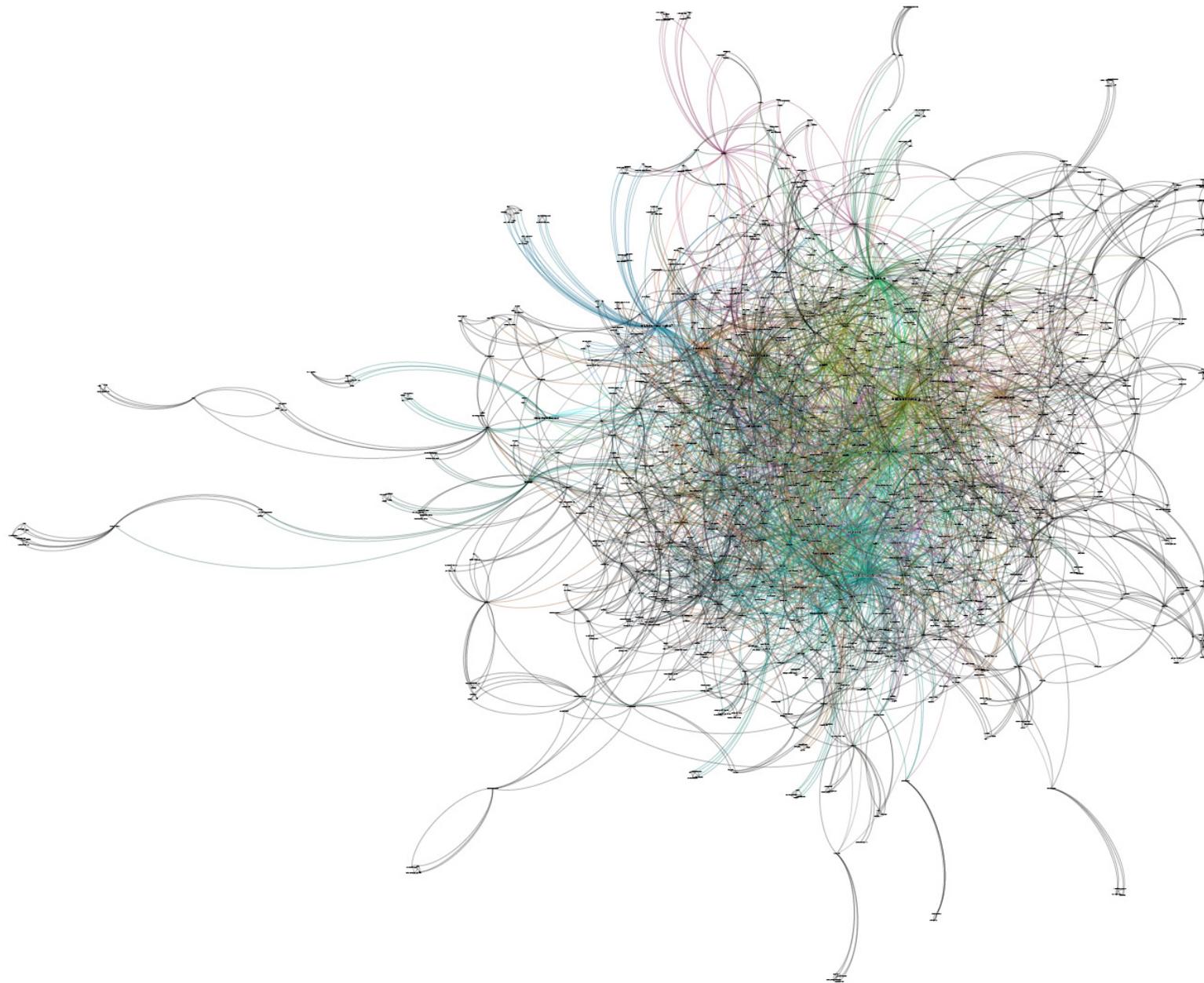
🗄️ ARCHIV

🎲 ZUFALLSBEITRAG

🏆 BEST OF

11.7.2022 Das Internet ist doch schlauer als die Menschen im Internet Ich bin bei der Physio und als ich an der Rez...
<https://t.co/gJUck05WvA>

@TECHNIKTAGEBU...



Author tags of the TTB blogposts,
visualized with Gephi (<https://gephi.org/>; 14/7/2022).

Topic Model Browser

Overview
Topic ▾
Document
Word
Bibliography
Word index
Annotations
Data: w75 ▾
Settings
About



Test instance of the Topic Model Browser
 (<https://agoldst.github.io/dfr-browser/>;
 14/7/2022) with SDC4Lit test data.

Topic 29

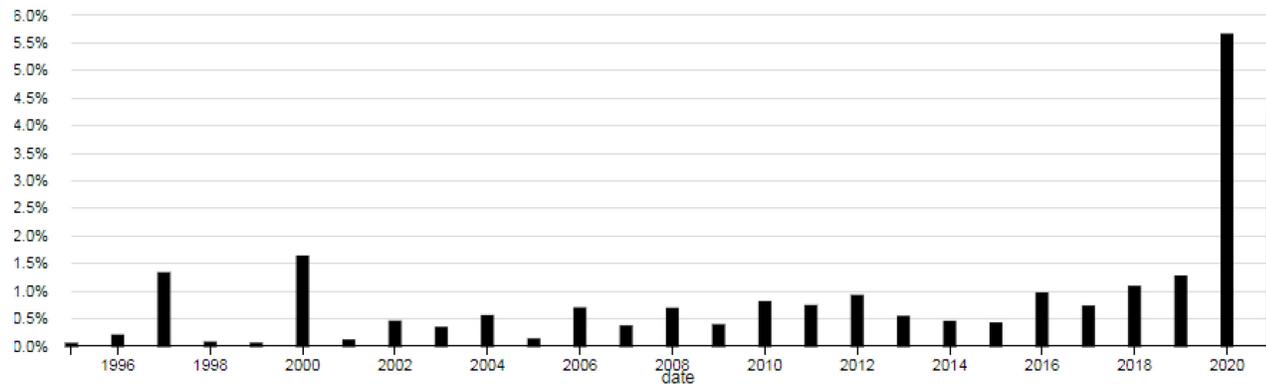
Top words

Word	Weight
vortrag	
gut	
kamera	
video	
halten	
hören	
teilnehmer	
publikum	
sprechen	
gemeinsam	
bild	
verwenden	
raum	
stellen	
leute	
zeigen	
frage	
stattfinden	
treffen	
tag	
pandemie	
gespräch	

Conditional probability of words in topic

Click a bar to limit to the documents it represents

clear selected



Top documents

Document	%	Tokens
Ab Anfang März 2020 https://techniktagebuch.tumblr.com/post/612772018510839808/ab-anfang-m%C3%A4rz-2020 2020 ["coronavirus,covid 19,voip,google hangouts,screensharing"] Die Arbeit in den Zeiten der COVID-19 An meiner Arbeitsstätte wird binnen weniger Tage von 11. bis 13. März 2020 die Reaktion auf die COVID-19-Pandemie eskaliert; von „Department-Meeting vorläufig... 1808	61.1%	44
November 2018 https://techniktagebuch.tumblr.com/post/180202625902/zuhause-in-ruhe-doppelt-so-schnell 2018 ["vortragstechnik,aufzeichnung,youtube,workaround,esther seyffarth"] Zuhause in Ruhe doppelt so schnell Ich höre einen inhaltlich sehr spannenden Vortrag. Leider spricht die Vortragende sehr lange und sehr langsam, sodass ich nicht allem, was sie sagt, folgen kann. Zum... 18592	60.9%	14

Keshif



Test instance of Keshif
(<https://github.com/adilaycin/Keshif>;
14/7/2022) with SDC4Lit test data.



Test instance of Keshif
<https://github.com/adiljalcin/Keshif>
 14/7/2022) with SDC4Lit test data.

Prospect

SDC4Lit Portal

- Integration of repository, analysis tools, search, etc.

Documentation/Education

- Documentation of the analysis tools, software and infrastructure for web archiving.

Contact

<https://sdc4lit.de/>



The screenshot shows the SDC4LIT website with a blue navigation bar containing links for 'SDC4Lit', 'Projekt', 'Veranstaltungen', 'Kontakt', and 'Impressum'. Below the navigation bar is a large graphic with the SDC4LIT logo and a network diagram. The main content area features the title 'Science Data Center for Literature' and a subtitle 'Aufbau eines nachhaltigen Datenlebenszyklus für Literaturforschung und -vermittlung'. The text describes the center's mission and lists project partners: Deutsches Literaturarchiv Marbach, Höchstleistungsrechenzentrum Stuttgart, Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart, and Abteilung Digital Humanities der Universität Stuttgart. On the right side, there are logos for 'deutsches literatur archiv marbach', 'HLRIS High-Performance Computing Center | Stuttgart', 'Universität Stuttgart', 'Baden-Württemberg MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST', and 'digital@bw'.

SDC4Lit Projekt Veranstaltungen Kontakt Impressum

SDC4LIT

Science Data Center for Literature

Aufbau eines nachhaltigen Datenlebenszyklus für Literaturforschung und -vermittlung

Das interdisziplinäre Science Data Center für Literatur (SDC4Lit) hat sich zum Ziel gesetzt, einen Datenlebenszyklus für Digitale Literatur nachhaltig umzusetzen: Archivierung, Erforschung und Vermittlung werden systematisch reflektiert und entsprechende Lösungen umgesetzt. Dabei soll insbesondere der Tatsache Rechnung getragen werden, dass Modellierung, Aufbewahrung und Analyse aufgrund der besonderen Erscheinungsform digitaler Literatur nicht zu trennen und Forschungsfragen von den Fachcommunities her zu denken sind.

In SDC4Lit werden verteilte langzeitverfügbare Repositories für Digitale Literatur aufgebaut und über eine Forschungsplattform zugänglich gemacht. Die Repositories werden regelmäßig erweitert, die Forschungsplattform bietet die Möglichkeit, computergestützt mit den Daten zu arbeiten.

Projektbeteiligte:

- Deutsches Literaturarchiv Marbach
- Höchstleistungsrechenzentrum Stuttgart
- Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart
- Abteilung Digital Humanities der Universität Stuttgart

deutsches literatur archiv marbach
HLRIS
High-Performance Computing Center | Stuttgart
Universität Stuttgart
Baden-Württemberg
MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST
digital@bw