

NLP @ ARGUS DATA INSIGHTS

WHAT'S HOT IN THE NEWS TODAY?

Claudia Schulz

SwissText 2021, 15th June 2021



EVERYTHING FROM A SINGLE SOURCE Cover the entire communication cycle

MEDIA NEWS
Media monitoring
Customised company news
Newsletters (on specific topics)
Individual media review
Social media reports and listening
Newsroom implementation



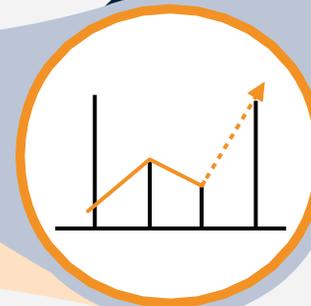
MEDIA SERVICES
Consulting & workshops
Strategic communications planning
ARGUS Sprachmanufaktur



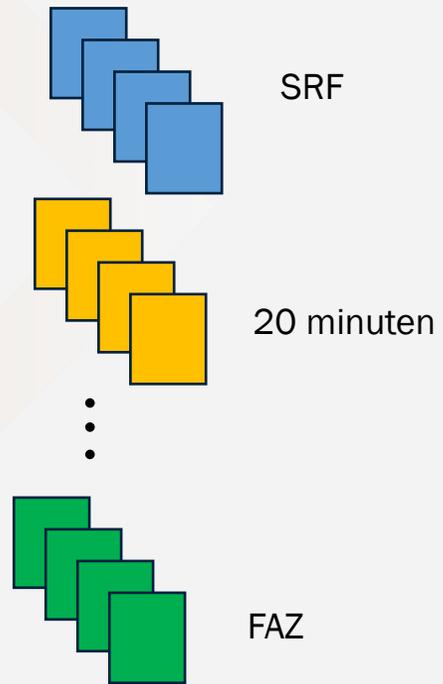
MEDIA ENGAGEMENT
Communication/event workflow management
MediaContacts database
Media distribution



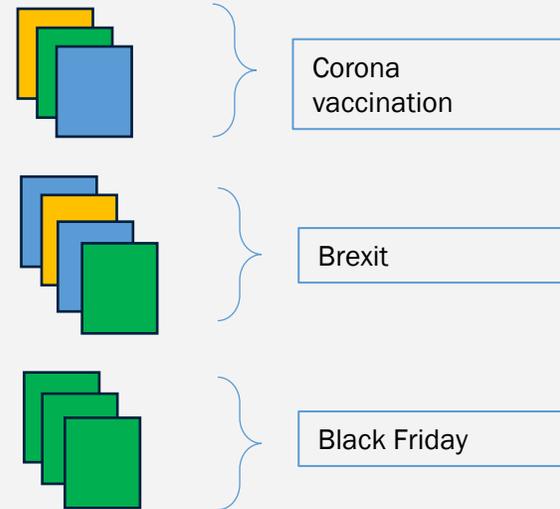
MEDIA ANALYTICS
KPI & controlling Reports
Media value
PR evaluation, event analyses, benchmarking
Social media analysis
Reputation analyses



Today's Articles



Top News Headlines



BERTopic

BERTopic is a topic modeling technique that leverages transformers and c-TF-IDF to create dense clusters **allowing for easily interpretable topics** whilst keeping important words in the topic descriptions. It even supports visualizations similar to LDAvis!

Corresponding medium posts can be found [here](#) and [here](#).



<https://maartengr.github.io/BERTopic/>

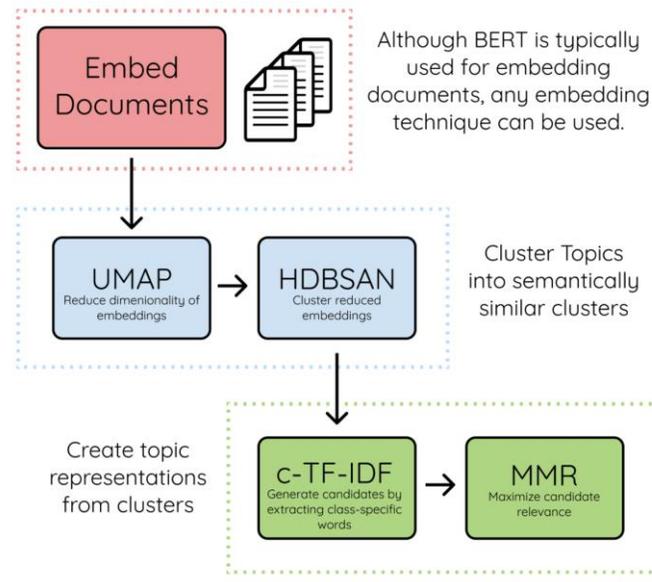
Top2Vec

Top2Vec is an algorithm for **topic modeling** and **semantic search**. It automatically **detects topics present in text** and generates jointly embedded topic, document and word vectors. Once you train the Top2Vec model you can:

- Get number of detected topics.
- Get topics.
- Get topic sizes.
- Get hierarchical topics.
- Search topics by keywords.
- Search documents by topic.
- Search documents by keywords.
- Find similar words.
- Find similar documents.

<https://github.com/ddangelov/Top2Vec>

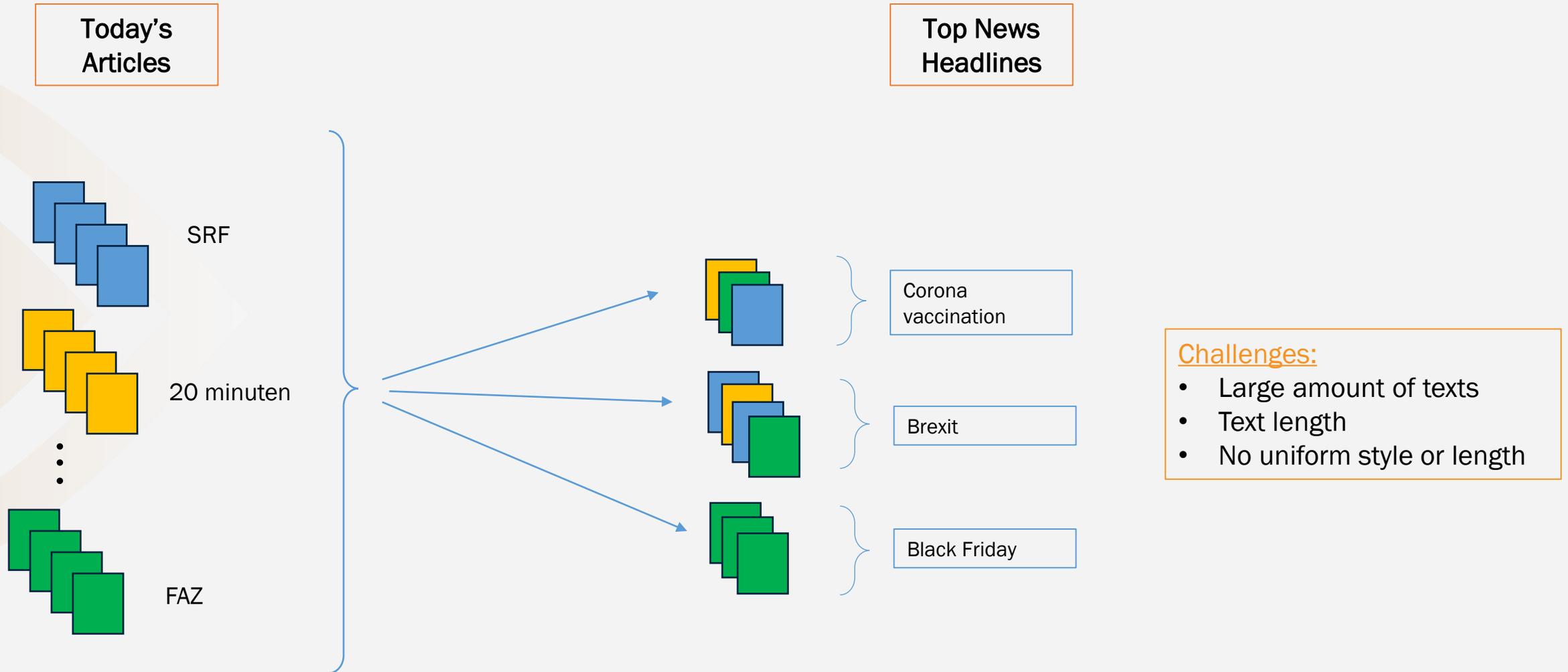
BERTopic

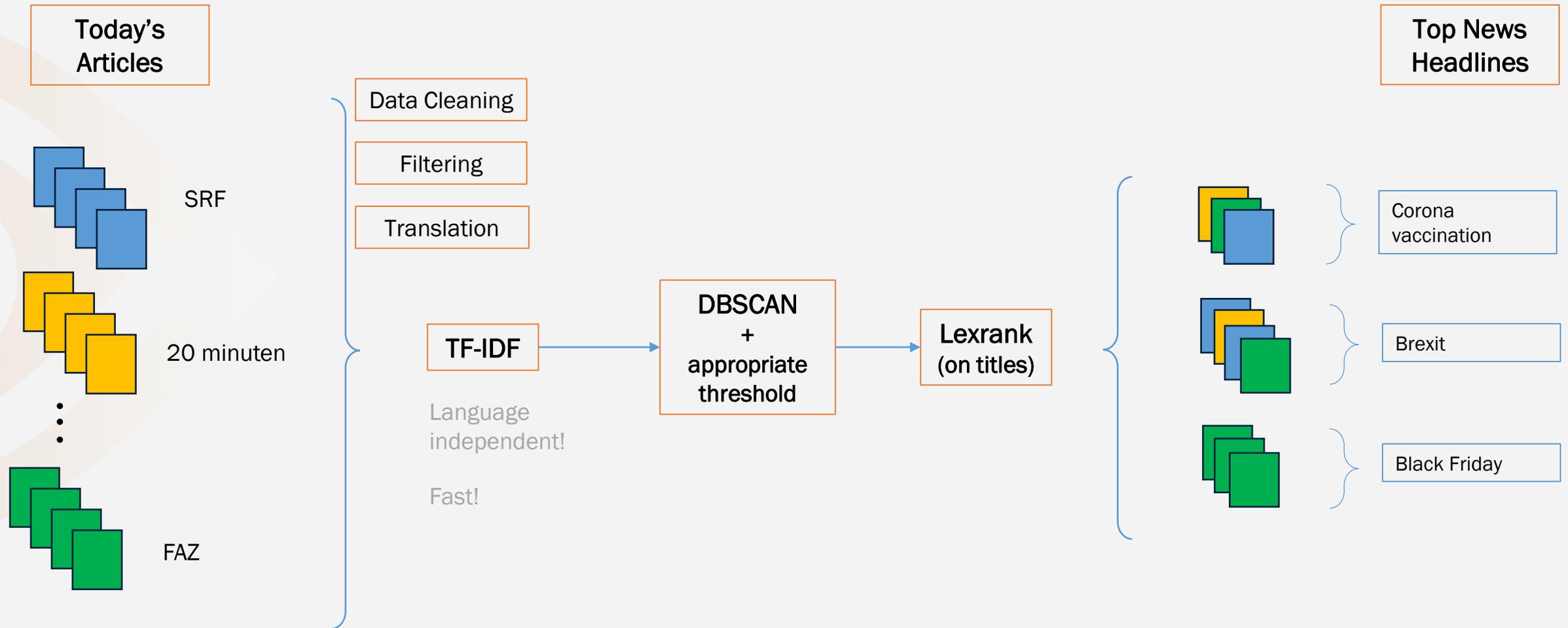


Carrot2

Carrot2 is a programming library for clustering text. It can automatically **discover groups of related documents** and label them with short key terms or phrases.

<https://github.com/carrot2/carrot2>





Hot Topics Zusammenfassung



Politik: Erster Nato-Gipfel mit Biden steht an.

China – Die Staats- und Regierungschefs der 30 Nato-Staaten kommen an diesem Montag zu einem Gipfeltreffen in Brüssel zusammen. Ein zentrales Thema ist die Frage, wie das Bündnis mit dem Aufstieg Chinas zu einer militärischen Weltmacht umgehen soll.



Politik: Israel vereidigt neue Regierung.

Eine hauchdünne Mehrheit der Abgeordneten im israelischen Parlament hat am Sonntag für die neue Regierung gestimmt. 60 von 120 Knesset-Mitgliedern votierten nach stürmischen Debatten für das Acht-Parteien-Bündnis unter Führung von Naftali Bennett von der ultrarechten Jamina und Jair Lapid von der Zukunftspartei. 59 stimmten dagegen, es gab eine Enthaltung.



Sport: Der Unbesiegbare.

Novak Djokovic besiegt Stefanos Tsitsipas nach 0:2-Satzrückstand und gewinnt zum zweiten Mal die French Open. Wenn es noch den geringsten Zweifel zur Widerstandskraft von Novak Djokovic gegeben haben sollte, so dürften dieser endgültig ausgeräumt sein. Der 34-jährige Serbe hat schon Grand-Slam-Finals gewonnen, in denen er Matchbälle abwehren musste, doch bei den French Open gelingt ihm etwas, das selbst für ihn Neuland ist:



Unbekannt: Berichte: London will Maßnahmen verlängern.

Grossbritannien – Der britische Premierminister Boris Johnson will Medienberichten zufolge die Corona-Massnahmen in England wegen der Ausbreitung der zunächst in Indien entdeckten Delta-Variante verlängern. Der bislang für den 21. Juni geplante "Tag der Freiheit" solle um bis zu vier Wochen aufgeschoben werden, meldeten verschiedene britische Medien am Montag.



Politik: FDP-Präsidentin Petra Gössi tritt spätestens Ende Jahr zurück.

FDP-Präsidentin Petra Gössi tritt zurück. Sie will ihr Mandat spätestens Ende Jahr niederlegen. Das teilte Gössi am Montag auf dem Kurznachrichtendienst Twitter mit.

A NOTE ON EVALUATION



Top Topics of the Day

07.01.2020

- USA: Zertifizierung des Wahlergebnisses
- USA: Sturm des Kapitolgebäudes in Washington
- USA: Senatswahlen Georgia: Beide Sitze gehen an Senatoren der «Democrats»
- Schweiz: Verlängerung der Anti-Corona-Massnahmen, Vorbereitungen für einen «Lockdown»
- Schweiz: Gewerbeverband und Gastroverband protestieren gegen die beschlossenen Corona-Massnahmen

08.01.2021

- USA: Trump sichert geordnete Amtsübergabe zu und verurteilt den Sturm auf das Kapitol
- Schweiz: Lobbying-Affäre um Christa Markwalder und Christian Miersch: Kasachischer Ex-Politiker Wiktor Chrapunow erhält Flüchtlingsstatus in der Schweiz
- Schweiz: Ski Weltcup Adelboden auf Corona-Hilfsgelder angewiesen

11.01.2020

- USA: Zweites Amtsenthebungsverfahren gegen Trump
- USA: Nach Löschung des Twitter-Accounts von Trump: «Parler» nicht mehr erreichbar
- Schweiz: Zulassung weiterer Impfstoffe stehen bevor
- Schweiz: Diskussion um Corona-Massnahmen und Entschädigungen
- Schweiz: Corona-Situation in Wengen/ Bern Diskussion um Durchführung der Lauberhorn-Rennen

13.01.2021

- Schweiz: Diskussion der möglichen Verschärfung der Corona-Massnahmen
- Italien: Regierungskrise – Premier Conte vor Verlust der Parlamentsmehrheit
- Schweiz: T. Burckhard (FDP) fordert Abbruch der Verhandlungen zum Rahmenabkommen mit der EU

C1	B	C	D	E	F	G	H	I	J	K	L
Size: 248 Sturm aufs Kapitol Biden Wahlsieg	Size: 109 Stocks	Size: 95 Corona Impfstoff (Zulassung)	Size: 92 Mix (not meaningful)	Size: 91 USA ticker (+++ format)	Size: 79 Facebook & Twitter sperren Trump account	Size: 71 Bitcoin mehr als 1Mio wert	Size: 69 Bundesrat verschärft Massnahmen	Size: 58 Assange bleibt in Haft	Size: 56 Parteitag Nordkorea: Kim Jong Un räumt Fehler ein, Militär wird ausgebaut	erstes Cluster sollte geteilt sein? fehlt: Senatswahlen Georgia fehlt: Gewerbeverband protestiert gegen Massnahmen	
Size: 783 Sturm aufs Kapitol Senatswahl Georgia Biden Wahlsieg	Size: 373 Mix (not meaningful)	Size: 221 Bundesrat verschärft Massnahmen	Size: 120 Corona Impfstoff (Zulassung)	Size: 117 Stocks	Size: 91 USA ticker (+++ format)	Size: 81 Bitcoin mehr als 1Mio wert	Size: 81 verschiedene Corona Artikel	Size: 79 Watt D'Or (Energiepreis)	Size: 78 Sportler in Quarantäne		
Size: 640 Sturm aufs Kapitol Senatswahl Georgia Biden Wahlsieg	Size: 241 Bundesrat verschärft Corona Massnahmen und denkt über Lockdown nach	Size: 110 Stocks	Size 103: Corona Impfstoff (Zulassung)	Size: 101 Mix (not meaningful)	Size: 91 USA ticker (+++ format)	Size: 76 Watt D'Or (Energiepreis)	Size: 71 Bitcoin mehr als 1Mio wert	Size: 43 LafargeHolcim kauft Dachbaufirma auf	Size: 59 Assange bleibt in Haft	erstes Cluster sollte geteilt sein, nützlich für subclustering fehlt: Gewerbeverband protestiert gegen Massnahmen	
Size: 556 Sturm aufs Kapitol Senatswahl Georgia Biden Wahlsieg	Size: 95 Mix (not meaningful)	Size: 93 Bundesrat verschärft Massnahmen	Size: 93: Corona Impfstoff (Zulassung)	Size: 83 Bundesrat verschärft Massnahmen	Size: 76 Watt D'Or (Energiepreis)	Size: 58 Assange bleibt in Haft	Size: 58 Bitcoin mehr als 1Mio wert	Size: 56 Sozialhilfe Konferenz	Size: 56 Parteitag Nordkorea: Kim Jong Un räumt Fehler ein, Militär wird ausgebaut		
Size: 317 Sturm aufs Kapitol Senatswahl Georgia Biden Wahlsieg	Size: 121 Bundesrat verschärft Massnahmen	Size: 91 USA ticker (+++ format)	Size: 86 Corona Impfstoff (Zulassung)	Size: 84 Stocks	Size: 76 Facebook (& Twitter) sperren Trump account	Size: 71 Bitcoin mehr als 1Mio wert	Size: 79 Mix (not meaningful)	Size: 59 Watt D'Or (Energiepreis)	Size: 56 Parteitag Nordkorea: Kim Jong Un räumt Fehler ein, Militär wird ausgebaut	erstes Cluster sollte geteilt sein fehlt: Gewerbeverband protestiert gegen Massnahmen Manche Cluster sind kleiner als bei vorherigem Ansatz: es fehlen Artikel?!	
Size: 203 Sturm aufs Kapitol Biden Wahlsieg	Size: 91 USA ticker (+++ format)	Size: 74 Facebook & Twitter sperren Trump account	Size: 64 Bitcoin mehr als 1Mio wert	Size: 56 Senatswahl Georgia Tump Anhänger und Sturm aufs Kapitol	Size: 55 Parteitag Nordkorea: Kim Jong Un räumt Fehler ein, Militär wird ausgebaut	Size: 54 Assange bleibt in Haft	Size: 53 Sozialhilfe Konferenz	Size: 47 Stocks	Size: 46 Kühne+Nagel verantwortlich für Moderna Impfstoff Vertrieb	Cluster scheinen zu klein (Artikel fehlen) fehlt: Verschärfung der Corona Massnahmen	
Size: 42 Infizierte Kinder gefährden Eltern	Size: 36 Watt D'Or (Energiepreis)	Size: 38 Keine systematischen Tests bei Einreise	Size: 41 Facebook & Twitter sperren Trump account							Cluster viel zu klein, Artikel sind fast immer genau die gleichen	
Size: 58 US Ticker	Size: 42 Infizierte Kinder gefährden Eltern									Cluster zu klein, Artikel sind immer sehr sehr ähnlich	
Size: 63 US Ticker	Size: 61 USA ticker (+++ format)	Size: 42 Infizierte Kinder gefährden Eltern								Cluster zu klein, Artikel sind immer sehr sehr ähnlich	
Size: 124 Sturm aufs Kapitol Biden Wahlsieg	Size: 91 USA ticker (+++ format)	Size: 85 US Ticker some other US articles	Size: 77 Facebook & Twitter sperren Trump account	Size: 64 Mix (not meaningful)	Size: 63 Stocks	Size: 62 Corona Impfstoff (Zulassung)	Size: 59 Assange bleibt in Haft	Size: 57 Bundesrat verschärft Massnahmen	Size: 56 Parteitag Nordkorea: Kim Jong Un räumt Fehler ein, Militär wird ausgebaut	fehlt: Senatswahlen Georgia fehlt: Gewerbeverband protestiert gegen Massnahmen Clusters seem a bit small	
Size: 91	Size: 82 US Ticker some other US	Size: 65 Facebook & Twitter sperren	Size: 64 Mix (not meaningful)	Size: 62 Sturm aufs Kapitol	Size: 58 Assange bleibt in	Size: 58 Corona Impfstoff	Size: 56 Parteitag Nordkorea: Kim Jong Un räumt Fehler ein, Militär	Size: 56	Size: 53	Cluster zu klein - es fehlen Artikel fehlt: Bundesrat verschärft Massnahmen fehlt: Senatswahlen Georgia	



- **Text encoding:** BERT is not always superior – don't disregard “old-school” methods
- **Off-the-shelf algorithms:** useful as building blocks (e.g. translation), but rarely applicable as is
- **Data cleaning:** don't underestimate the impact of data processing, cleaning, and filtering