



## *Content analysis report*

### Deliverable 2.1



## Project Information

Project acronym	COMMIT
Project title	COMMunication campaign against exTremism and radicalisation
Reference number	867019
Programme	ISFP-2018-AG-CT-CSEP
Coordinator	Centro per lo Sviluppo Creativo Danilo Dolci
Project website	<a href="https://www.commitproject.eu/">https://www.commitproject.eu/</a>

## Deliverable Details

Deliverable n.	<i>D2.1</i>
Title	Content analysis report -Report on push & pull factors and root causes of radicalisation
WP	WP2 - Campaign research and communication strategy
Lead beneficiary	UNIPA ( <i>Report developed by Textgain</i> )
Dissemination Level	<i>Public</i>
Format	<i>Electronic</i>
Document version	<i>1 – September 2020</i>



## CONTENTS

<b>The Project and its objectives .....</b>	<b>4</b>
<b>Introduction to the report .....</b>	<b>6</b>
<b>Insights from Social Media Harvesting and Data Analysis .....</b>	<b>6</b>
Data harvesting .....	7
Data Harvesting for the COMMIT-project .....	8
Bootstrapping Greek and Italian POW-analyses using seed words .....	11
General findings .....	12
Tracking Micro Issues.....	13
Macro Issues .....	16
Topic Modelling.....	21
German.....	21
Greek.....	25
Italian .....	27
Dutch.....	29
Related terms.....	31
Don't be afraid to ask.....	32
Greek .....	32
Italian .....	32
German.....	33
Dutch.....	33
Islamic Radicalization.....	35



## The Project and its objectives

During the past years, Europe has been facing alarming tendencies toward increasing extremist, radical & terrorist propaganda. Extremists use widely online means to spread hate, encourage violence, radicalise and recruit followers, prepare and incite attacks and then claim credit for them. The online activities are highly concentrated on social media.

According to Europol, over 150 social media platforms are used by terrorists for propaganda. Such content has proven to accelerate the radicalisation processes.

However, it is not only terrorists who promote violence & push audience towards radicalisation. Many European extreme right- left/-wing groups are mobilising their followers, communicating about Daesh brutality as justification for their own extreme response. The migration crisis only fuelled this trend.

Besides, media (willingly or not) support the extremist' and radicals' efforts by negative hostile media coverage, incorrect, selective or misleading reporting opening the door to populism, right-/left-wing extremism, intolerance & violence. Education around the issue for future media specialists is a self-evident need, however is seldom part of the standard curricula.

Extremists, radicals & terrorists are populating the spaces where young people consume & share information, socialise & are socialised which makes them a highly vulnerable group to terrorist/radical/extremist propaganda & recruitment attempts. Young people are therefore one of the main groups that would benefit from learning skills improving their capacity to detect propaganda, fake information & extremist content, and being trained about how to develop & disseminate positive messages among their peers that can be a real alternative to extremist content online. To help them resist indoctrination & radicalisation, their critical thinking & media literacy must be improved & the internalisation of democratic values supported. COMMIT does so via offline activities such as workshops and online communication on platforms usually used by the young.

Although online communication activities are not deradicalisation tools in themselves, they can be used as preventive tools, offering different perspectives on societal challenges and, directly/indirectly, challenging extremist ideas via counter & alternative narratives. Counter narratives aim to "deconstruct, discredit & demystify extremist propaganda while alternative narratives undercut it by focusing on what we are 'for' rather than 'against', or by putting forward a positive story about social values, such as tolerance, freedom & democracy". The development of such narratives as well as strategic communications are considered as a priority<sup>7</sup> in order to prevent terrorist propaganda from having success<sup>8</sup>.

Considering all of the latter, COMMIT aims to prevent & dissuade vulnerable young people (13 – 25) in 4 partner countries from extremism, radicalism & terrorism providing them with skills relevant to co- create counter narratives challenging extremist online propaganda and alternative narratives



promoting democratic values, tolerance & cooperation, and to identify & resist extremist online content.

COMMIT adopts a trans-medial approach combining online campaigning (3 campaign will be developed addressing: fake news, hate speech, populistic propaganda; right-/left-wing extremism; Islamic radicalism & terrorism) & face to face activities (workshops, contest, events). COMMIT will also improve the capacity of university students, media professionals, Civil Society Organisations, Internet companies, stakeholders to meet the new challenges linked to extremist propaganda online & radicalisation, training them in alternative & counter narratives and their use in prevention of radicalisation through a capacity building programme.

## **PROJECT'S ACTIVITIES**

- Campaign research (content analysis and target audience mapping) and communication strategy for the campaigns' development;
- Capacity Building for:
  - 1) internal staff of the project;
  - 2) university students, civil society organisations and local stakeholders (10 training sessions for a total of 30h);
  - 3) young people (approx. 26h of workshops & contest);
- Development and co-creation of 3 communication campaigns addressing:
  1. fake news, hate speech, populistic propaganda;
  2. right-/left-wing extremism;
  3. Islamic radicalism & terrorism;
- Development of COMMIT Declaration;
- Creation of a local network.
- Setting up and implementation of intervention support via dedicated online channels to support people affected by/close to Islamic radical and terrorist groups.

The project activities are developed throughout 7 working packages:

WP1 - Management and Coordination of the Action

WP2 - Campaign research and communication strategy

WP3 - Internal capacity building and campaign evaluation

WP4 - Capacity Building for university students, civil society organisations and local stakeholders

WP5 - CSL1 - Campaign service line addressing fake news, hate speech, populistic propaganda online

WP6 - CSL2 – Campaign addressing right- and left-wing extremism

WP7 - CSL3 – Campaign addressing Islamic radicalism and terrorism



## Introduction to the report

The current document is a technology-supported online content analysis developed by Textgain (Belgium) under the coordination of UNIPA (Italy) to identify the push and pull factors and root causes bringing to radicalization. It is meant to be read together with **D2.2 – Target Audience Mapping report**, developed parallelly. Both reports are crucial in the scope of the project COMMIT, since they provide the coordinates and the basis for the design and development of the campaigns to be spread in the later stage of the project. Both this report and the target audience mapping try to focus on the causes and identify the target to whom the partnership will address their future activities. In the next pages, both the methods and the results of this analysis will be introduced, providing an overview on all the countries involved in the research (namely, Italy, Greece, The Netherlands, Austria).

## Insights from Social Media Harvesting and Data Analysis

Textgain develops social media monitoring tools that collect data from social media and then analyses that data using Artificial Intelligence technology from the field of Natural Language Processing (NLP). NLP is technology that is able to automatically extract information from large document streams through automated text analysis. This provides a quantitative perspective on the data that is able to support the qualitative analyses conducted by the other project partners.

In the COMMIT-project we apply both top-down and bottom-up approaches to the automated analysis of text:

- **top-down:** [pretrained linguistic classification engines](#) that can extract information such as grammatical features, keywords and the demographic features of the author of a text through computational stylometry. In the context of the detection of polarization and radicalization, we make use of our [POW<sup>1</sup>-analysis](#) tool: a lexicon-based approach that scans documents for linguistic markers of toxicity and assigns an overall toxicity score to a document, along with its toxicity dimensions.
- **bottom-up:** we also employ machine learning technology to *let the data speak for itself*. Given enough data, our tools can automatically identify interesting linguistic patterns that arise from the data without any supervision. This is particularly useful when inspecting subsets of the data to detect linguistic patterns that are specific to a region, time frame or topic.

After the data analysis, the data and its metadata are presented in a [dashboard](#) that allows the users to inspect and interact with the collected data and identify additional microtrends. This report contains a description of our initial findings. The dashboard consequently serves as a living document that allows the project partners to continuously monitor the social pulse for the topics under investigation in the COMMIT-project.

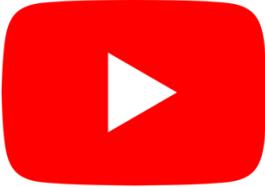
---

<sup>1</sup> Profanity and Offensive Words



## Data harvesting

We currently have technology to monitor these three major social media platforms:

		
<b>Twitter</b> is the most straightforward platform from which to harvest data. It offers programmatic access to tweets through an API: queries can be sent to an endpoint in the form of <i>search terms</i> , which then returns the tweets that match them. A free endpoint is available that yields a subset of the tweets available on the platform.	<b>YouTube</b> likewise offers an API for data harvesting. Queries can be sent containing search terms, after which a list of relevant videos is returned. Additional API-calls can then return the comments for said videos. Access to the YouTube-API is limited however and harvesting comments is costly.	<b>Facebook</b> has severely limited access to its content through its GRAPH-API, following the Cambridge Analytica controversy. Programmatic access to content is now controlled by page and group owners. We therefore are only able to collect data from this platform that is made publicly available for Facebook-users and non-Facebook users alike. Rather than searching for keywords, Facebook data is harvested on the level of the page or group.

Other platforms that were considered, but not retained for the COMMIT project:

- **reddit** offers full access to its content, but is very Anglo-Saxon centered and therefore less suited for consideration in the COMMIT-project.
- **Instagram** offers programmatic access to its content, but only through hash tag searches, severely limiting the range of the available content streams. Furthermore, Instagram is mostly an image-based platform. The current limits in the state-of-the-art in automated image processing prevent the reliable use of automated content analysis techniques.
- **TikTok** has a very closed ecosystem and only allows for a small subset of its data to be harvested. The platform is under scrutiny for its abuse of data and may implement more transparent data access in the future.



- **Telegram** is prominently known for fostering toxic groups. It is however not easy to mine and doing so would require undesirable infiltration practices that fall beyond our mandate.

### Data Harvesting for the COMMIT-project

The COMMIT-project is organized around a few central themes: populist, right-/left-wing extremist, Islamic radical and terrorist online propaganda and fake news. Working with broader themes prevents us from initiating our data harvesting on Twitter and YouTube, as we don't have yet have an accurate set of search terms to trigger the right queries to the API. In order to quickly collect in-domain data, we therefore asked the project partners to identify relevant Facebook pages and groups. The data collected from these sources allows us to identify linguistic patterns that are specific to the respective themes that are being discussed on said pages and groups.

The following table provides an overview of the sources identified by the project partners and that are being targeted in the current data harvesting cycles. Some resources are not (yet) available for harvesting, the reasons for which are indicated in red<sup>2</sup>.

Language	Page name	Themes	Status
Greek	Ελληνικά Hoaxes	Fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	Αρτέμης Σώρρας Υπ. Βουλευτής Β' Πειραιώς	Fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	Σουρεάλ Ελλαδιστάν	Fake news, hate speech, populist propaganda	unreachable
	STOP Islam in Greece	Fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	ΠΟΤΕ ΞΑΝΑ ΑΡΙΣΤΕΡΑ	Right and left wing extremism	<input checked="" type="checkbox"/>
	Ρουβίκωνας	Right and left wing extremism	removed
	Ενάντια στο σκοτάδι του φασισμού	Right and left wing extremism	community
	ΑΝΤΙΦΑΣΙΣΤΙΚΗ ΔΡΑΣΗ	Right and left wing extremism	<input checked="" type="checkbox"/>
	islam.gr	Islamic radicalism and terrorism	<input checked="" type="checkbox"/>
	ΠΑΤΡΙΩΤΙΣΜΟΣ ΚΑΙ ΟΡΘΟΔΟΞΙΑ	Islamic radicalism and terrorism	unreachable
	ΙΣΛΑΜ ΛΑΘΡΟ ΕΙΣΒΟΛΗ - INVASION	Islamic radicalism and terrorism	unreachable
	Χριστιανισμός & Ισλάμ είναι Κατασκευασμένα Ιουδαϊκά Ψεύδη	Islamic radicalism and terrorism	community
	Έλληνες Ξεσηκωθείτε Κατά της Τουρκοποίησής Σας	Islamic radicalism and terrorism	no_comments
German (Austria)	Kronen Zeitung	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	unzensiert.at	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>

<sup>2</sup> Note that there are very few resources that explicitly deal with Islamic radicalism. This is due to the fact that such pages are proactively removed from Facebook as soon as they appear on their radar.



	Felix Baumgartner	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	FPÖ	right and left wing extremism	<input checked="" type="checkbox"/>
	Norbert Hofer	right and left wing extremism	<input checked="" type="checkbox"/>
	Wir wollen kein Islam in Deutschland und Österreich	right and left wing extremism	<input checked="" type="checkbox"/>
	KPÖ	right and left wing extremism	<input checked="" type="checkbox"/>
	autonome antifa w	right and left wing extremism	<input checked="" type="checkbox"/>
	KJÖ – Kommunistische Jugend Österreichs	right and left wing extremism	<input checked="" type="checkbox"/>
Italian	Matteo Salvini	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	Alessandro Di Battista	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	Giorgia Meloni	right and left wing extremism	<input checked="" type="checkbox"/>
	Potere al Popolo	right and left wing extremism	<input checked="" type="checkbox"/>
	Islamici Sunniti in Italia	islamic radicalism and terrorism	community
Dutch	Zwarte Piet is Racisme	right and left wing extremism	<input checked="" type="checkbox"/>
	Zwarte Piet Blijft	right and left wing extremism	<input checked="" type="checkbox"/>
	Welkom Zwarte Piet	right and left wing extremism	<input checked="" type="checkbox"/>
	Omroep Brabant	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	Filip Brusselmans	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	Schild & Vrienden	right and left wing extremism	<input checked="" type="checkbox"/>
	Vlaams Belang	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	Sam van Rooy	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	Tom Van Grieken	right and left wing extremism	<input checked="" type="checkbox"/>
	VRT NWS	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	HLN.be	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	De waarheid voorbij het nieuws	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	TScheldt	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	SCEPTR	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	5G-Plein	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	Stop 5G (België)	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>
	Stop 5G Nederland	fake news, hate speech, populist propaganda	<input checked="" type="checkbox"/>



## Statistics

The following table gives an overview of the quantities of harvested data as of mid-September 2020.

Page/Group	Posts	Comments
<b>ΠΟΤΕ ΞΑΝΑ ΑΡΙΣΤΕΡΑ</b>	16287	83279
<b>De waarheid voorbij het nieuws</b>	16246	97476
<b>Kronen Zeitung</b>	11685	607684
<b>HLN.be</b>	10649	4435
<b>VRT NWS</b>	7829	548030
<b>Καταγραφή Γεγονότων (Ομάδα)</b>	6114	7529
<b>Matteo Salvini</b>	4523	1552422
<b>Omroep Brabant</b>	3078	277020
<b>5G-Plein</b>	2680	24120
<b>SCEPTR</b>	2573	4553
<b>unzensuriert.at</b>	2266	62262
<b>Giorgia Meloni</b>	1649	744294
<b>Potere al Popolo</b>	1595	56300
<b>FPÖ</b>	1219	102432
<b>Sam van Rooy</b>	1148	8036
<b>Norbert Hofer</b>	908	106943
<b>Tom Van Grieken</b>	787	176288
<b>TScheldt</b>	743	11145
<b>Ομάδα της ιστοσελίδας ellinikahoaxes.gr</b>	666	10993
<b>Vlaams Belang</b>	561	95931
<b>autonome antifa w</b>	516	4001
<b>Ελληνικά Hoaxes</b>	338	32059
<b>KPÖ</b>	290	1756
<b>IZLAM GR</b>	286	288
<b>Filip Brusselmans</b>	179	7339
<b>KJÖ – Kommunistische Jugend Österreichs</b>	155	811
<b>Felix Baumgartner</b>	132	5059
<b>Zwarte Piet onder Vuur</b>	121	1234
<b>Stop 5G (België)</b>	111	2412
<b>Alessandro Di Battista</b>	79	56309
<b>Schild &amp; Vrienden</b>	67	1005
<b>Stop 5G Nederland</b>	27	3
<b>Zwarte Piet Niet</b>	16	0
<b>STOP Islam in Greece</b>	15	0
<b>Welkom Zwarte Piet</b>	9	0
<b>Wir wollen keinen Islam in Deutschland und Österreich</b>	1	0
<b>Zwarte Piet Blijft</b>	1	0
<b>TOTAL</b>	<b>99549</b>	<b>4693448</b>



### Privacy Concerns

When harvesting data, we go to great lengths to anonymize the data. We remove any references to personally identifiable data (e.g. mentions) and remove any additional available metadata. What is left is the text itself, its time of publication and the unique ID of the post and the comment (or reply to the comment). The IDs are kept in case the comment needs to be studied in context. Should the submitter of the content have removed the comment in the meantime, the original message cannot be retrieved, thereby safeguarding the person's right to be forgotten.

### Bootstrapping Greek and Italian POW-analyses using seed words

One of the key features of our approach is the POW-analysis, which provides a relative estimate of the toxicity level of user-generated content. At the start of the project we had modules available for Dutch, English, French and German. To bootstrap POW-analyses for Greek and Italian, we consulted our partners to establish a *seed list* of toxicity terms. To create a seed list, we used Google Translate to translate the English list of toxicity terms to Greek and Italian and selected 150 of the most toxic words for manual inspection.

The respective partners were presented with the following interface:

Phrase	Source	Toxicity level	HATE	SHIT	SCUM	FOOL	SLUT	FUCK	GOOK	HEIL	HELL
αράπης	en: nigger	😢	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
γαμώ	en: fuck	😢	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
σκατά	en: crap	😢	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
γαμημένος	en: fucking	😢	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
καλούπι	en: die	😢	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ρατσιστής	en: racist	😢	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
εχθρός	en: enemy	😢	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The interface allows for erroneous translations to be corrected and for the following parameters to be adjusted for the target language:

- The toxicity level of the term in isolation
- The toxicity dimensions of the word (we refer the reader to [our technical report](#) for more info on what these dimensions entail)

This small subset functions as a seed list: a gold-standard of toxicity patterns in the target language. After data harvesting, word embedding techniques can be used to identify other toxic terms, as well



as their related concepts. This automatically expanded list will consequently be used to harvest data using search-term based APIs for platforms such as Twitter and YouTube.

### General findings

In previous research efforts, we observed that polarized individuals certainly don't keep to one topic or theme when venting their discontent on social media, but that a common framework of norms and values has long since been established that shapes the individual's opinion on a wide variety of issues. People who post anti-EU content, also tend to be climate sceptics and have negative feelings towards immigration, Islam and multiculturalism in general. Likewise, they are generally positive towards the policies of US President Trump and other conservative state leaders such as Orban and Bolsonara. This common framework was seemingly established dynamically: there is no specific reason why climate sceptics are typically also against abortion, but the issues have become consolidated as a unit. Sometimes a theme arises for which there is less disagreement: for some actors on the extreme-right, COVID-19 is considered as a hoax, by others as a serious biological attack on society, instigated by China, Bill Gates or the Deep State. Regardless, there is still a distinct common ground, i.e. the fact that local governments are to blame, either for implementing exaggerated measures, or by cohorting with the enemy.

This general common ground is also fostered for seasonal issues on the hyperlocal level: there is typically a strong agreement on themes such as Black Pete in the Dutch language area (*zwarte piet*; June 2020), the smoking ban in Austria (*smoke-abstimmung*; June 2020), the mobilization of refugees at the Greek border (*επιστράτευση*; July 2020) or even minor issues, such as the re-opening of discos in Italy (*discothece*; July 2020).

It can be observed that seemingly small issues, such as the discussion on the alcohol ban at Vienna's Christmas market (discussed in September 2020) escalate quickly. Such issues are typically introduced in the public debate by thought leaders (in this case, the FPÖ), causing a ripple effect of discontent to the page's members. Important to note is that the open nature of these page's comments sections, causes the story to appear on the feeds of the connected friends of the members as well, adding to the ripple effect of exposure.



Mein Weihnachten mit meiner gesamten Familie lass ich mir sicher nicht verbieten. Es hat schon gereicht das Ostern nichts geworden ist, und danach hätten wir es ja falsch verstanden. Ich bin doch keine Marionette!

Ein Wahnsinn, die Diktatoren der Regierung wollen uns das Recht auf ein ordentliches Vollbrett nehmen?!?  
Nicht mit uns!! So 8 bis 11 Glühwein am Punschstand sind eine Tradition, die wir uns bestimmt nicht verbieten lassen werden!!11!

09:42 · Sep 6, 2020

<https://www.facebook.com/fpoe/posts/1765808106893784>

This serves to recruit additional supporters to a more general cause: an anecdotal issue such as Gluhwein can attract attention from *bystanders* and draw them into a more general narrative that



conforms to the common framework of norms and values. Once an individual has engaged positively with content from a specific page, the social media content delivery algorithms take over and ensure that similar content will appear more frequently in their feed.

As such, social media have become an incredibly powerful grooming tool for populist politicians: it allows them to broadcast opinions on a wide range of (often micro-) issues, draw in additional supporters and consequently expose them to increasingly polarizing content. Step two is to attempt to discredit alternative voices: a common tactic is the projection of mainstream media (commonly referred to as *the MSM*) as leftist organizations pushing a fake, pro-establishment narrative. This serves to alienate the reader from more traditional, nuanced content and stimulate them to consume alternative media, such as blogs and YouTube videos that push a specific narrative.

It is no surprise that many radicalized individuals become highly susceptible to the conspiracy theories that propagate through the filter bubbles they have entered. Outside voices that attempt to dispute (parts of) the framework consequently face an uphill battle, as they are considered to be agents of the conspiracy that is keeping the *real truth* hidden.

### Tracking Micro Issues

Once an individual has become a consolidated member of the filter bubble, it becomes very hard to reach them with counternarratives. An individual may start his/her journey by lamenting the lack of Gluhwein at the Christmas market and then continue down a rabbit hole that has them eventually reject multiculturalism on the whole. It is therefore crucial to reach people as early as possible in the radicalization process, starting at the level of micro-issues.

But while counternarrative practitioners, policy makers and scholars have a solid grasp on the general framework of norms and values that circulate within specific filter bubbles, hyperlocal and temporary issues are difficult to track manually. Automated social media monitoring provides a solution: content scraping of social media data and subsequent information extraction techniques can continuously monitor filter bubbles and detect new trends as they arise and dissipate.

To this end, we developed a machine learning based approach to analyze the content streams of the COMMIT-project. The algorithm called Meerkat considers  $n$  number of data slices as independent sets of documents to be differentiated, and their linguistic patterns as the markers that differentiate them. The technique ignores patterns that are common to the respective data sets and searches for those linguistic patterns that are indicative of their set. The advantage of the approach is that it is language-agnostic, as well as task-agnostic. The different data slices can represent ideologies (e.g. FPÖ vs ANTIFA), stances (e.g. pro-immigration vs anti-immigration), but also time frames. The latter case provides the entrance into the temporal trends.



We illustrate the output on data of the Matteo Salvini-page for the year 2020.

1/2020	<i>plastica, bonaccini, buonanno, bonanno, maranello, gaetano pizzo, lucia, angelo pietro, complici, 26, william, borgonzoni, romagnoli, emilia, soufiane, libici, donazioni, gianluca, mirco casa, 49 euro, ferrari, tuslo, karma, materiali, cesenatico, palmiro, sardine, libia, togliere parola, carmela boccia, carmela, spingete, defunti, veruccchio, maria bari, meravigliose, comprensibile, calabresi, aria fresca, umbria</i>
2/2020	<i>cappuccino, capuozzo, eleonora, benetton, ponte, francesco bruno, inglesi, brexit, devi sveglio, lettera, investitori, cambiare europa, tassi salvatore, concessioni, lira, boris, inglese, lamorgese conte, maria albanese, idiota, sterlina, 02, padania, vignola, tessera, agenore, mojito, giustizia funziona, pozzallo, toscani, tennenini, marco tennenini, caffè, condannarti, gaglano, sari, francia germania, cambiaria</i>
3/2020	<i>carabiniere, pistola, rapinatore, alimentari, chiudere, stefano pastore, bel funerale, michela bruno, bollettino, rapina, indispensabili, farmacie, giovanni bianchi, lavanderie, pronto soccorso, lineare, colpa genitori, giordano, ucciderlo, persona sana, fabbrica, conosco tante, supermercati, sciare, messo rischio, casa figli, governatrice, tamponi, smile, schierarsi, uguale pd, tosto, meccanici, mirato, tira palle, carabinieri, eccesso, raccontando, aperti</i>
4/2020	<i>bollette, bolletta, de luca, compassione, consumo, tommaso benedetto, cateno, deluca, matteini, oneri, consumi, lucia brambilla, raffaella, matteo kurba, kurba, seconda casa, contatore, sindaci, cose devono, 48, signora dovrebbe, salvini famiglia, mancia, propositi, wood, el wood, messinesi, arzente, salvini rosicone, ritardi, governo potevi, sfiducia, vedi foto, calabro, milena calabro, cristiano greco, lavoro sporco, maglietta, sindaco vuole</i>
5/2020	<i>vauro, saviano, commercialisti, plasma, vignette, cinesi, vignetta, scorta, satira, vignettista, luglio, ornella saba, saba, fazio, commercialista, petto, disegnato, cina, corrotti sinistra, ornella, vuol lavorare, omosessuali, disegnatore, disegna, volgarità, squallido, cinese, domenico vuoto, urlato, qnt, italiani chiedono, misiani, cartello, laboratorio, pericolo giallo, fiore, autoritratto, londra, giugno</i>
6/2020	<i>briatore, prete, mes, lega votato, zanda, don, costantino, csm, gianluca palumbo, vitalizio, generali, preti, cossiga, commissione, specchi, radiato, presepe, laura viole, vitalizi, correnti, obbedisce, magistrato, argyris, irene argyris, fabio giordano, vasto, pechino, sabrina, governo stelle, milazzo, confessionale, giletti, benvenuto, 20 luglio, durso, btp, salvini sicilia, parrocchiani</i>
7/2020	<i>capezzone, vincenzo palazzo, bonghi, lampedusani, alessandro pepe, animalisti, magdi, sabrina squitieri, isolani, sindaco lampedusa, alessandro giuseppe, tommaso natale, buon luglio, caro sindaco, thatcher, usanze, ficco, luca ficco, pecora, sindacalista, sparascio, nuda, selvaggi, paccheri, audi, provocarci, mandolino, mese luglio, foresta, marisa martella, bongo, barchini, calenda, lucarelli, agapanthus, targhe, macellazione, pallottoliere</i>
8/2020	<i>arenzo, tunisia, ninfa, tunisi, libano, arabo, caserta, andrà finire, parassiti complimento, sopportarli, susanna, buon venerdì, sbarchi zero, gatto volpe, morte rossi, franco alberto, conto rovescia, svenduto, patria matteo, legaaaa, scodinzolare, miriano, derisa, casertano, incoerenti, grandi valori, brutti ricordi, pane arabo, campania bisogno, scuola capisce, scusi posso, telese, manto, mattinate, lecco, buon ferragosto, ammalano, bannon, ditino</i>
9/2020	<i>fusani, giornaliste, aggressione, 15 regioni, vinto, berlinguer, pio, fitto, violenza, fusa, verticale, bianchina, lucci, azzolina, orizzontale, senato, francesco cerino, aggredito, aosta, viva amore, congratulazioni, puntini, imprenditrice, willy, marchigiani, autocontrollo, bacia, marche, pugliese, emergenza, rai, fidanzata, befane, vittorie, toti, consiglieri, aggressioni, rosi, santuario</i>

Since the technique is task-agnostic, we can apply it to shorter time frames as well, zooming on the level of the month or even the day. Since it is fast enough to work in real time, we implemented Meerkat as a functionality in the COMMIT-dashboard.



We present an operational example of a temporary micro-issue for the Greek Facebook-group ΠΟΤΕ ΞΑΝΑ ΑΡΙΣΤΕΡΑ (Never Again Left). The top indicator for March 7 is the word δημοτικά (*primary school*). These comments can be found on a post on leftist practices at Universities, which is consequently railroaded towards similar practices happening much earlier in the educational system. It illustrates the hive mind and common framework of beliefs that counternarrative campaigners and policy makers should be aware of when implementing their actions.

ΧΙΛΙΑΔΕΣ ΕΛΛΗΝΟΠΟΥΛΑ ΕΓΙΝΑΝ ΑΝΘΕΛΛΗΝΕΣ ΣΤΑ ΔΗΜΟΣΙΑ "ΠΑΝΕΠΙΣΤΗΜΙΑ" ΑΠΟ ΤΟΥΣ ΑΡΙΣΤΕΡΟΥΣ ΚΑΘΟΔΗΓΗΤΕΣ ΠΟΥ ΔΕΝ ΑΦΗΝΑΝ ΝΑ ΚΑΝΟΥΝ ΜΑΘΗΜΑ ΟΙ ΚΑΘΗΓΗΤΕΣ... ΕΤΣΙ ΑΝΤΙ ΤΑ ΠΑΙΔΙΑ ΝΑ ΜΑΘΟΥΝ ΓΡΑΜΜΑΤΑ ΕΜΑΘΑΝ ΝΑ ΚΑΝΟΥΝ ΠΟΡΕΙΕΣ ΝΑ ΚΑΙΝΕ ΤΟΝ ΚΟΣΜΟ ΜΕ ΜΟΛΟΤΟΦ ΚΑΙ ΝΑ ΚΑΙΝΕ ΤΗΝ ΕΛΛΗΝΙΚΗ ΣΗΜΑΙΑ !!! ΚΑΤΙ ΠΟΛΥ ΣΑΠΙΟ ΣΥΝΤΕΛΕΙΤΑΙ ΣΤΑ ΔΗΜΟΣΙΑ ΠΑΝΕΠΙΣΤΗΜΙΑ ΕΔΩ ΚΑΙ ΧΡΟΝΙΑ !!! ΝΑ ΣΤΟΧΕΥΕΤΕ ΣΤΗΝ ΡΙΖΑ ΤΟΥ ΚΑΚΟΥ !!!

[See Translation](#)

274

59 Comments 56 Shares

Το κακό άρχισε νωρίτερα. Από τα δημοτικά, με κάτι αληταμπουρες ανταρσυες και κουκουεδες, όπως αυτόν στην Κέρκυρα αν θυμάστε.

8

Όχι μόνο στα πανεπιστήμια. Αυτο το πολύ σάπιο ξεκινά από τα δημοτικά σχολεία με δασκάλους άχρηστους αριστερούς και συνεχίζεται σε γυμνάσια και λύκεια από αριστεροφασιστες καθηγητές.

13

Το κακό ξεκινάει ακόμα κι από τα δημοτικά Σχολεία από επίορκους δασκάλους!!! Η ιστορίδωση των πάντων μετά το 1974 !!! Κι Η μεγάλη ευθύνη του Ανδρέα Παπανδρέου και του ΠΑΣΟΚ Που εξιλέωσε την Αριστερά στην κολυμπήθρα του Σιλωαμ!!!

3

The Meerkat-method can also discover specific terminology that is used within communities. Understanding the vocabulary of a community can help the design of more effective counternarratives using a common vocabulary. For the Dutch Facebook group 5G-plein, which shares (mostly) false information on the dangers of 5G-technology, the word “trigger” is often used, in a variety of contexts. A person can for instance be “triggered” to finally understand the truth about 5G, but the term is mostly used to refer to the health problems that 5G causes in your body (e.g. activating corona) or the government mind control that 5G is able to *trigger* through a yet to be implanted microchip. A counternarrative towards this content is advised to re-use such terms, as a common vocabulary will help send the message across.

welke trigger heeft je wakker gekregen?

[Like · 32w](#)

dat is trouwens ook weer niet helemaal waar.... het verhaal zou zijn dat via 5g vatbaarder gemaakt word of kan worden voor het corona virus.een soort trigger

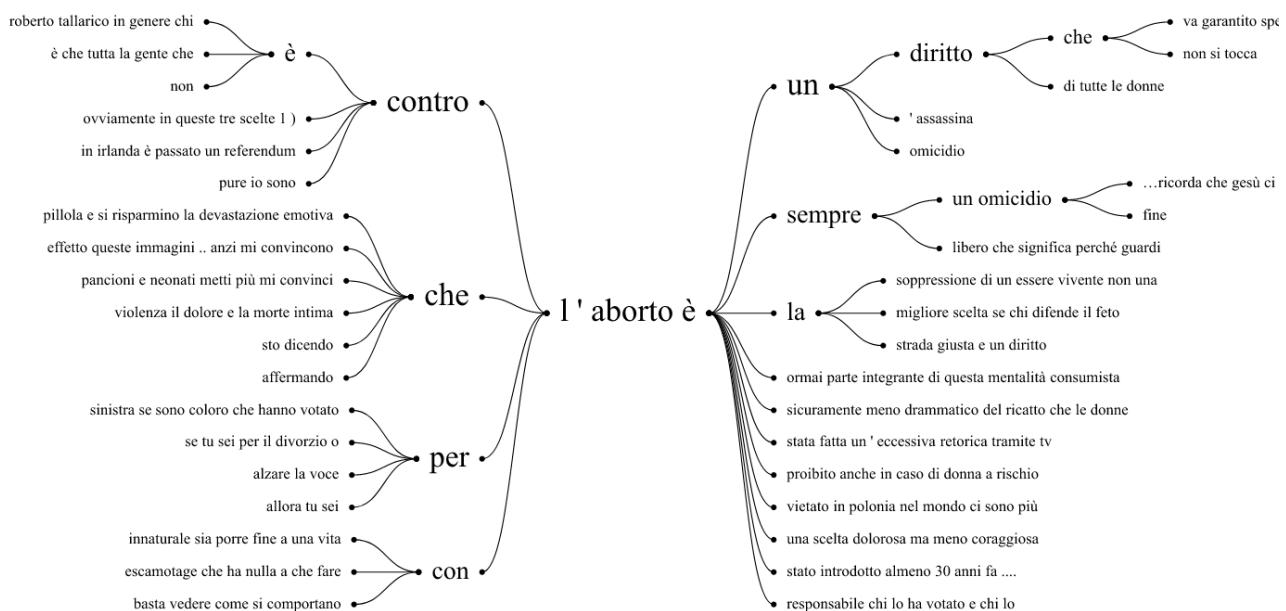
[Like · 30w · Edited](#)



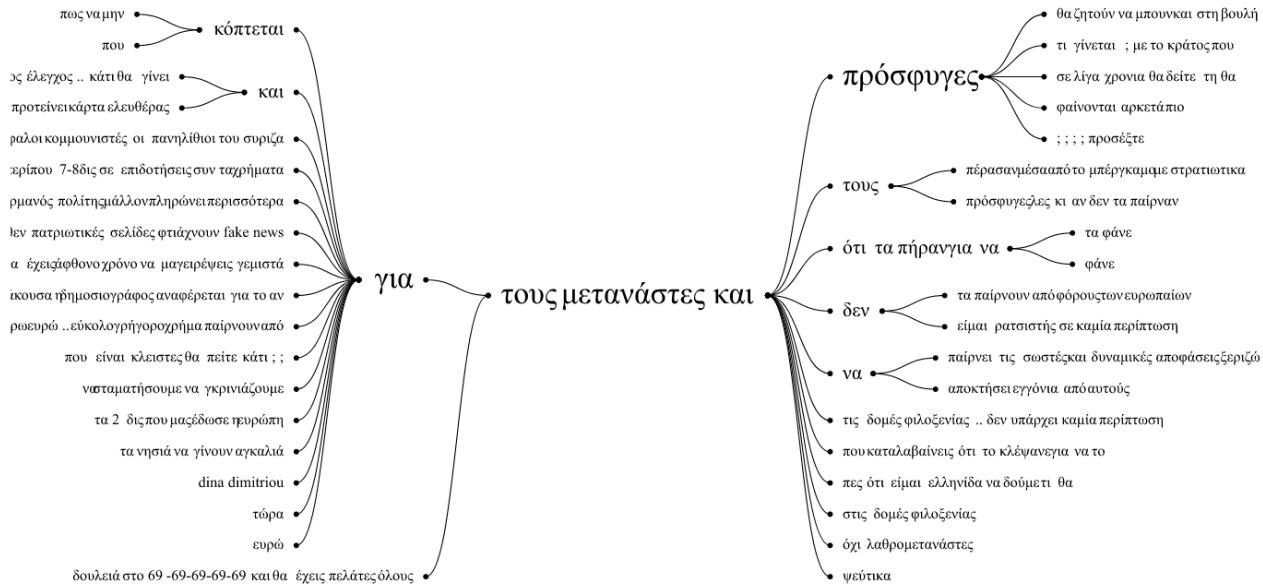
## Macro Issues

Larger political and ethical themes obviously give rise to polarized debate as well. Such themes are a lot more prominent in the data than the aforementioned micro-trends and therefore require a different approach to study. Traditional collocations can present the word in its context, but still require a lot of manual inspection. We suggest the use of word trees to study these concepts: we pre-filter the statements in which the word occurs according to their content and then present the collocations in a tree structure that summarizes the content.

### *Italian: aborto (abortion)*



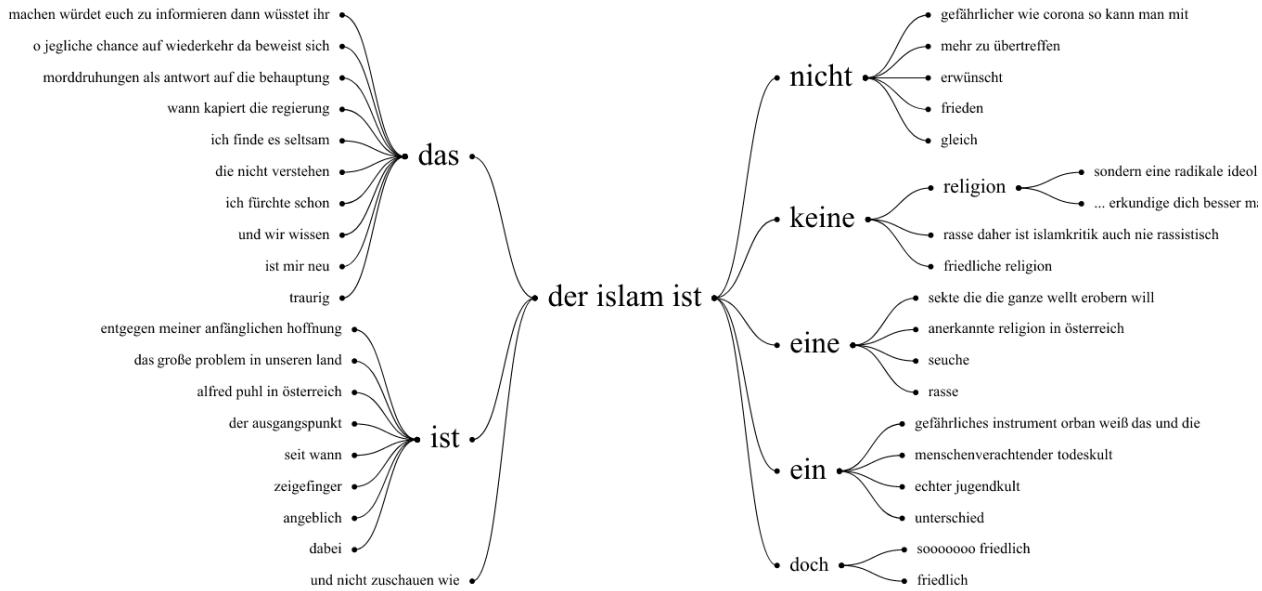
### *Greek: Μετανάστες (Immigrants):*



Dutch: 5G

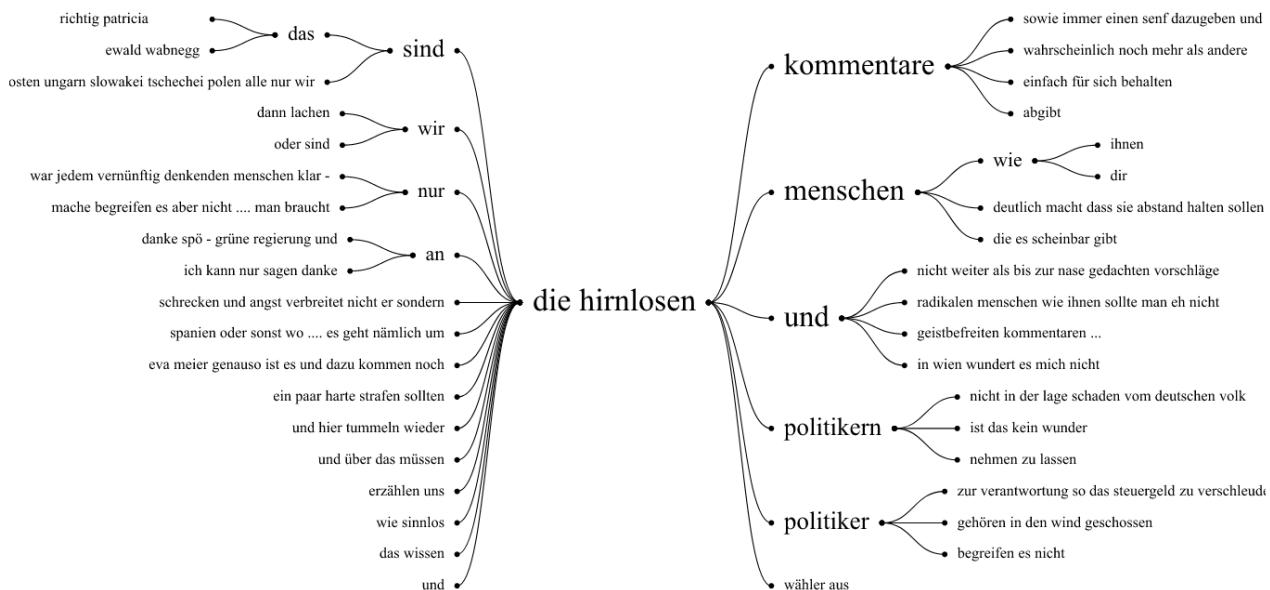


### German: Islam

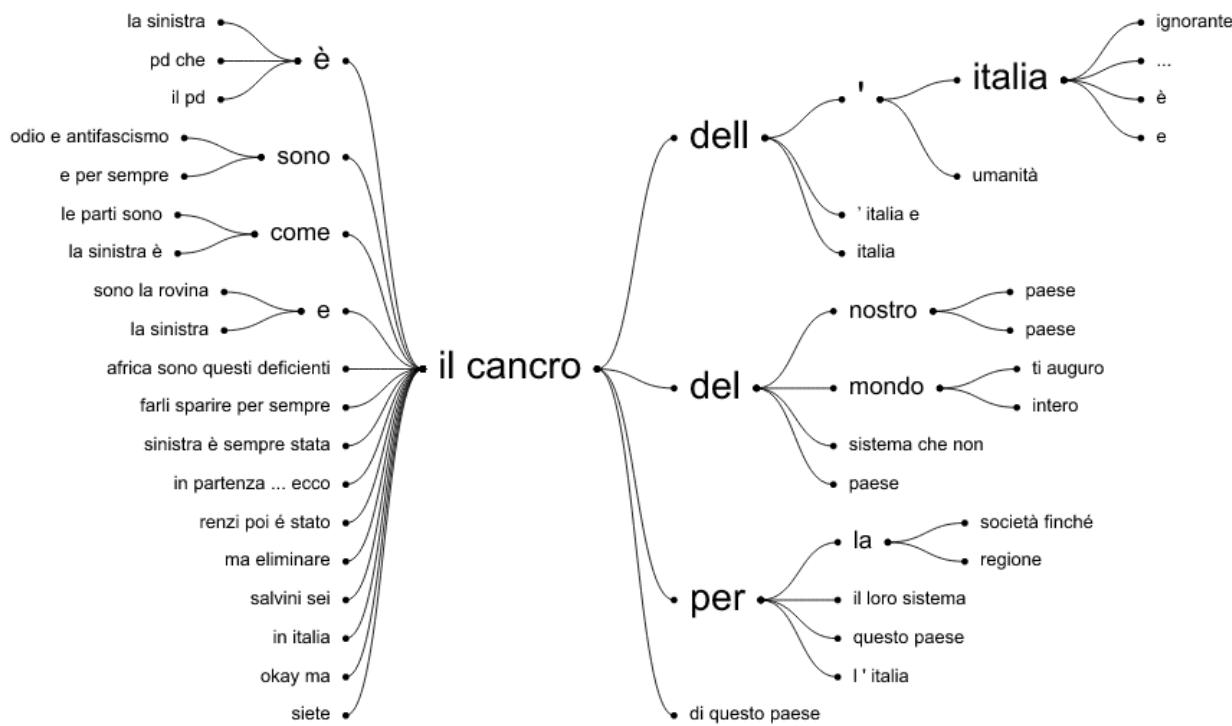


The technique can also provide insight into toxicity patterns, as exemplified in the following examples.

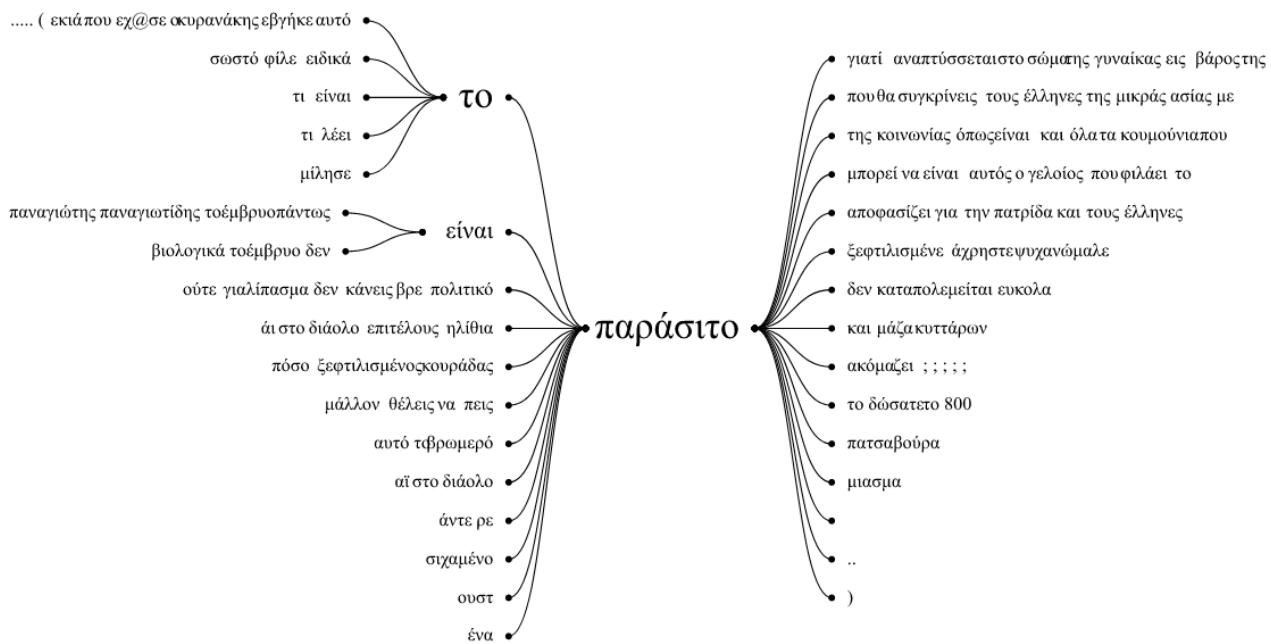
### German: hirnlosen(brainless)



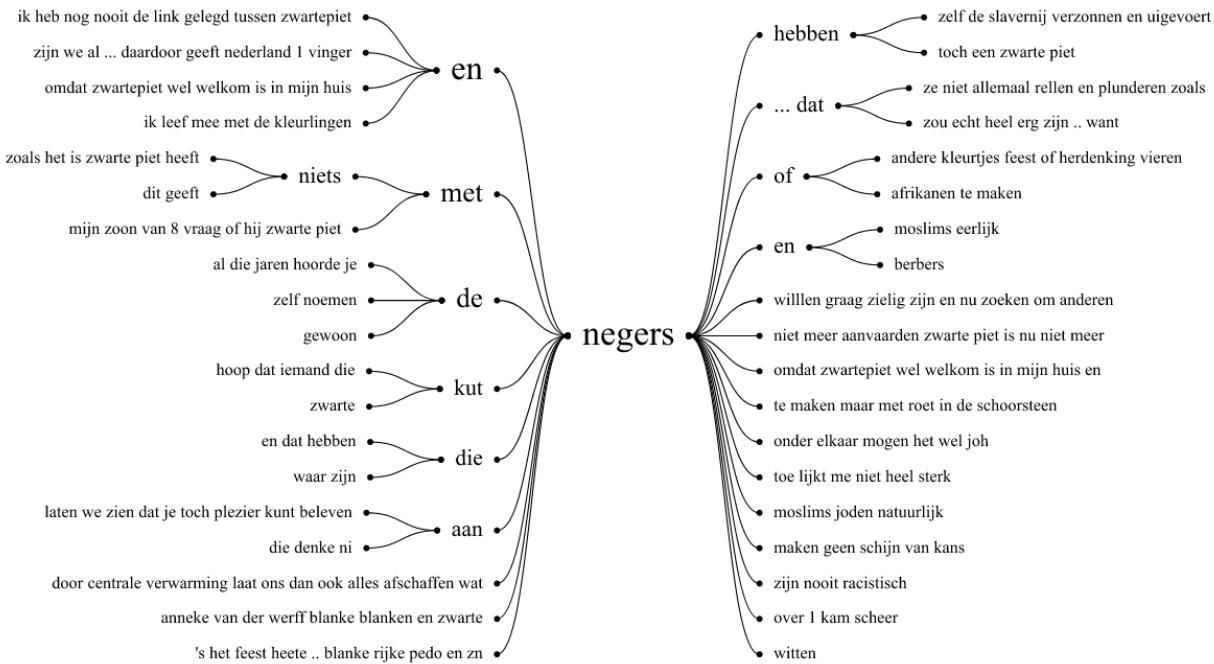
*Italian: cancro (cancer)*



*Greek: παράσιτο (parasites)*



### Dutch: negers (negros)



## Topic Modelling

The word tree approach provides a top-down method of perusing the data streams and gain some insight into the patterns surrounding known specific thematic or toxic words. Another way to find subthemes in the data is through the use of topic modelling. By observing the distributions of linguistic tokens over the documents, topic modelling attempt to automatically infer latent themes that may be present in a collection of documents. The advantage of topic modelling is that it is completely unsupervised: no predefined assumptions are made about the data. The disadvantage is that topics can only be represented and interpreted as “bags” of words. Often topic clusters are formed that function as “garbage” collections of words that do not fit in with any of the other topics, but have no internal semantic consistency otherwise. Nevertheless, such topic models enable a serendipitous approach to identifying terms for further scrutiny.

We compiled topic models for each of the datasets, requesting 20 topics per dataset. We display the results below: the 20 inferred topics for the respective datasets and their top-10 most indicative terms. Rather than temporal analysis of *trending words*, these tables provide insight into words that are important to the respective subthemes that have been discovered.

### German

#### Kronen Zeitung

	0	1	2	3	4	5	6	7	8	9
0	mal	schau	schauen	sehen	ma	gespannt	sag	lesen	reden	nachdenken
1	danke	info	herr	schön	china	griechenland	gesund	ok	einsatz	polizei
2	sicher	einzelfall	impfen	österreicher	net	ned	lass	lasse	impfung	geh
3	genau	meinung	jo	wissen	humor	sehe	weiß	lesen	martin	christian
4	österreich	urlaub	deutschland	land	nacht	italien	kroatien	armes	ausland	österreicher
5	super	idee	finde	leistung	toll	echt	find	gratuliere	tolle	aktion
6	nix	passiert	neues	sagen	bringt	anderes	sandy	verstehen	reden	hört
7	wien	hallo	stadt	wiener	quarantäne	rot	wahl	grün	sichtig	phil
8	eh	klar	jo	braucht	passt	scho	ned	egal	wurscht	net
9	bitte	geh	lesen	herr	lasst	bleiben	bleibt	erklären	sofort	vergessen
10	traurig	beileid	echt	wahr	einfach	familie	unfassbar	kraft	frieden	ruhe
11	leider	spät	wahr	passiert	geben	ändern	überall	meisten	gibts	dummheit
12	is	ma	ned	de	jo	net	scho	wos	wurscht	do
13	corona	virus	gestorben	grippe	sterben	stirbt	tote	krise	angst	influenza
14	stimmt	richtung	jo	maria	sabine	andreas	andrea	100	irgendwas	eva
15	geld	braucht	bekommen	eu	brauchen	staat	verdienen	ums	bekommt	kostet
16	bravo	herr	vollauf	griechenland	ungarn	respekt	dominik	gratuliere	strache	kickl
17	regierung	grünen	volk	schuld	gewählt	weg	fpö	maßnahmen	wirtschaft	experten
18	hoffentlich	bleibt	bekommt	weg	strafe	gesund	bekommen	finden	fällt	gefunden
19	einfach	leute	kinder	leben	bleiben	virus	lassen	hab	sagen	wissen



*unzensuriert.at*

0	1	2	3	4	5	6	7	8	9
0 regierung	gehört	volk	kotzen	sofort	sowas	fpö	grüne	danke	einfach
1 österreich	nacht	armes	schande	politik	gehört	österreicher	leben	land	traurig
2 eu	politiker	braucht	brauchen	europa	merkel	diktatur	land	verein	italien
3 sicher	einzelfall	sogar	grüner	stimmenland		traumatisiert	brauchen	bekommen	normal
4 genau	jo	humor	passt	stimmt	franz	meinung	traurig	weiß	schaut
5 grün	rot	wählen	schwarz	türkis	wähler	typisch	wählt	wien	danke
6 https	com	www	facebook	youtu	be	watch	youtube	php	at
7 deutschlandwählen		merkel	armes	genauso	passiert	schafft	bitte	gleiche	diktatur
8 grünen	erwarten	anderes	övp	erwartet	linken	roten	wählt	verbannen	wählen
9 mal	sagen	schauen	politiker	arbeiten	sehen	einfach	selber	bitte	schau
10 geld	bekommen	brauchenarbeiten		zahlen	braucht	bekommt	steuerzahlerwelt		kostet
11 gewählt	schuld	giefert	bekommenselber	volk	hab		österreicher	leute	jammern
12 super	sender	cool	einfach	finde	gratuliereerfüllung		idee	wille	wähler
13 wissen	österreicherglauben	nix		gerne	politiker	glaube	sagen	hauptsache	halt
14 klar	eh	is	anfang	braucht	nix	zahlen	passiert	egal	sowas
15 hoffentlich	geschichte	schnell	franz	wahl	renate	fällt	nächsten	satire	bleibt
16 jürgen	schaupp	kiki	langweilig	danke	linken	ahnung	antwort	schreiben	nix
17 leider	wahr	stimmt	traurig	leute	spät	peter	franz	helmut	fpö
18 bravo	herr	kickl	trump	ungarn	basti	polen	helden	jo	höchste
19 weg	gehört	sofort	gehören	gis	gebühren	diktatur	heimat	merkel	nix

*Felix Baumgartner*

0	1	2	3	4	5	6	7	8	9
0 birthday	happy	from	day	sir	to	best	keep	legend	today
1 felix	bravo	grüße	super	grande	2020	cool	top	sieht	österreicher
2 the	of	and	best	one	all	on	this	to	my
3 new	year	happy	2020	my	to	too	all	for	healthy
4 maske	tragen	masken	viren	virus	schützen	weg	schützt	astrid	sterzinger
5 geburtstag	liebe	gesundheit	bleib	schweden	läßt	nachträglich	wünsche	50	glückwunsch
6 you	are	thank	to	for	do	did	your	have	félix
7 danke	marina	ružić	worte	wünsche	toll	wagner	wunderschönfürs	mario	
8 baumgartnerherr	schweiz	felix	félix	mister		gewohnt	schade	peter	herrn
9 it	to	live	made	great	your	not	and	watch	can
10 beautiful	photo	enjoy	elegance	det	switzerland	mother	such	thank	looks
11 https	com	www	facebook	watch	youtube	de	videos	youtu	be
12 nice	day	have	very	thanks	pic	watch	one	landing	time
13 weiss	tobias	kommentar	sagen	wissen	benutzen	wichtig	nützen	manchmalhr	
14 genau	matthiaspasst	meinung	ronny	bestätigt	halten	virus	bravo	bauer	
15 michel	schrödermichalowskimasken	krank	unglaublichsehen		tragen	luft	tochter		
16 martin	kreiter	koisser	meinung	heuchler	witzig	abgeschlachtetauer	versteht	demonstriert	
17 rip	sterling	sir	remember	rennen	that	reden	anfang	sad	hear
18 mal	bitte	gerne	nachdenkenmitfliegenwort		österreich	lass	artikel	einfach	
19 is	that	good	mihaela	where	michael	to	wos	and	she



FPÖ

	0	1	2	3	4	5	6	7	8	9
0	kickl	hr	beste	herbert	herr	innenministerspitze	brauchenherrn	einfach		
1	genau	deutschlandpunkt		is	meinung	gebracht	wissen	monika	sehe	gehört
2	regierung	gehört	volk	grüne	kotzen	rücktritt	österreichergesamte	leute	jetzige	
3	super	rede	frau	tolle	punkt	gesprochen	gebracht	dagmar	wahrheit	einfach
4	österreichbraucht		deutschlandnacht		land	schande	armes	leben	verloren	bleiben
5	bravo	rede	punkt	frau	belakowitschvollauf	herbert	dagmar	gebracht	norbert	
6	danke	worte	wähler	bitte	einsatz	wahrheit	grünwähler	rede	mut	punkt
7	weg	gehört	masken	gehören	kreuz	maulkorb	grün	grüne	maske	gis
8	fpö	wählen	partei	wähler	liebe	övp	stimme	einzig	bitte	spö
9	sicher	impfen	lasse	lassen	lass	net	bekommt	impfstoffkogler	geben	
10	leider	spät	wahr	traurig	politiker	schreiben	österreichervolk	maria	glauben	
11	mal	bitte	einfach	sagen	wahrheit	corona	frau	leute	politiker	lassen
12	https	com	www	facebook	be	youtu	watch	youtube	php	at
13	stimmt	100	ingrid	andreas	monika	michael	josef	gruber	martin	vollkommen
14	herr	hofer	kogler	norbert	geburtstag	peter	kickel	wahrheit	kickl	bundeskanzler
15	grünen	övp	erwarten	kreuz	gehören	schwarzen	bleiben	türkisen	verbannentürkis	
16	grüße	liebe	wien	deutschlandhallo	schöne	tirol	graz	steiermarkgrüsse		
17	gewählt	grün	schwarz	türkis	schuld	rot	wähler	selber	volk	wählen
18	eu	geld	braucht	brauchen	austritt	türkei	land	europa	krise	zahlen
19	sofort	gehört	grenzen	rücktritt	gehören	abschieben	austritt	öxit	schließen	abgesetzt

Norbert Hofer

	0	1	2	3	4	5	6	7	8	9
0	herr	hofer	kickl	kogler	reden	stimme	geehrter	worte	spitze	vollkommen
1	regierung	gehört	rücktrittsreifgesamte	volk	grün	misstrauensantragjetzige	angst	opposition		
2	genau	andreas	deutschland	is	weiß	wissen	peter	passiert	jo	herbert
3	österreichdeutschlandarmes		nacht		verloren	heimat	leben	wien	schande	gehört
4	danke	wähler	schön	worte	grün	wünsche	gleichfalls	ebenfalls	liebe	info
5	leider	wahr	spät	traurig	wähler	herbert	wahrheit	österreicher	ändern	sagen
6	sicher	impfen	lasse	lassen	lass	stimme	net	bekommen	impfung	selber
7	fpö	wählen	wähler	partei	övp	kickl	spö	liebe	stimme	starke
8	weg	gehört	masken	gehören	maske	kogler	grünen	co	braucht	schnell
9	super	rede	geschrieben	gesprochen	kickl	gratuliereidee		finde	einfach	gratulation
10	mal	reden	schauen	sehen	denken	schau	gespannt	arbeiten	sag	ma
11	stimmt	100	sonja	monika	möstl	sagen	traurig	ingrid	richtung	hauptache
12	https	www	com	facebook	youtube	watch	at	youtu	be	php
13	meinung	gleichen	voll	sagen	selben	anderer	absolut	andreas	peter	geschrieben
14	eu	land	geld	corona	sagen	politiker	lassen	leute	braucheneinfach	
15	grünen	gewählt	grün	schwarz	türkis	wähler	wählen	hab	schuld	rot
16	hofer	hr	norbert	kickl	vollkommen	herrn	maria	beste	gerwald	rede
17	bitte	geh	lesen	bleiben	schreiben	schnell	helfen	nehmen	reden	aufklärung
18	sofort	abschieben	gehört	grenzen	heimat	gehören	rücktritt	abgeschoben	hause	schließen
19	bravo	norbert	kickl	genauso	frau	andy	vollauf	wahrheit	100	grün



## KPÖ

	0	1	2	3	4	5	6	7	8	9
0	https	www	com	youtube	watch	facebook	photo	at	twitter	status
1	kpö	gewusst	blümel	schreibt	drauf	graz	mannsberger	harri	kleintransporter	alten
2	mai	lebe	hört	genossinnen	grüße	liebe	netzwerken	beiträge	sozialen	twitter
3	flag	false	aktion	ne	empörung	sicher	arbeit	finde	genau	geldgier
4	einfach	kapitalismus	etc	politiker	ziehen	enteignen	gekündigt	haha	max	regierung
5	auto	österreichisches	fahren	mal	övp	kruckenkreuz	beschweren	malt	schreibt	drauf
6	be	youtu	https	zuw5zeqhabe	polizisten	leider	forderung	revolution	the	not
7	sachbeschädigung	sorry	anzumalen	lachen	verboten	türkis	randalierer	chaoten	toll	wirft
8	schwarz	barbara	ausgeschlossen	sickle	absatz	strafminderung	strafandrohung	verbalhornung	paragraphen	fakevernaderung
9	halt	wiener	reicher	ibiza	arschlach	andre	wahlkampfkosten	gesetzt	eleganter	massnahmen
10	iwanow	nikolai	ns	namen	proletarier	system	forderung	diskussion	lese	hysterie
11	augentrost	soundchoice	anarchie	verwechselt	anarchismus	simon	rabeder	anomie	meinst	blatt
12	links	ehler	draufgemalt	freizeitlichen	schätze	hält	rechts	bevölkerung	parlament	liegt
13	blümel	geil	mobil	aufregen	vergessen	gewusst	chinesische	vermute	schriftzeichen	wofür
14	herzog	klemens	kummerl	angrennt	versteh	elektroauto	billiger	liste	22	000
15	schön	hach	eigentlich	gefällig	fügt	design	verrechnen	vorher	oh	kosmonautinnen
16	weg	sieht	wirkt	platz	fehl	irgendwie	blümel	werkzeug	maler	hand
17	wien	echt	endsieg	lsufen	verleugnen	ideologie	türkis	hab	gesehen	jössas
18	jürgen	heigl	fragen	couleur	vergleicht	kp	sucht	linke	egal	überschreiten
19	danke	ausgebessert	hinweis	di	poster	voller	genossen	fürs	kommentar	film

## autonome antifa

	0	1	2	3	4	5	6	7	8	9
0	the	and	of	to	you	it	that	is	not	be
1	stobinski	oskar	schwurbelst	verstehen	redest	wovon	ahnung	geschnürt	laufschuhe	sorry
2	walter	maisinger	islamophob	versteckt	gespannt	pamphlet	komplex	verkürztes	schön	antisemitismus
3	linken	zionismus	ignorieren	getrost	karikaturen	eifernden	wahnhaft	kommentare	außerhalb	statement
4	parapluie	jonathan	leuten	rabbiner	eigentlich	richtige	umgang	rassistschutz	irgendeinem	gearte
5	okay	palästinensische	richtet	staat	halt	imperialismus	jüdischen	blees	klaus	link
6	hagir	hak	rahman	tschüss	regel	delegitimierung	fällt	laut	aussage	punkt
7	fabian	lehr	genossen	gratuliere	feine	guck	mal	troll	gemeinsam	gemeint
8	lustig	hendrik	protokolle	jakob	gähn	überraschend	antiimpottosfinden	echt	zion	
9	migrationshintergrundwurst	greift	antisemit	leute	antifa	beschmiert	gehirnschachclubbraun			blond
10	rodorff	maximilian	respektvolleumgang	laufschuhnen	spint	thema	halt	jungle		fuer
11	roque	josé	toimil	antonio	einfach	ablauf	checken	buch		gähn
12	mal	ribo	kader	ne	linke	linken	geschnürt	welt	laufschuhe	existiert
13	unger	armin	eigentlich	tbh	gerechnet	kampf	hab	plo	schreibst	raf
14	antisemitismus	verteidigen	islamophob	biokartoffel	weiß	klar	staat	mal	gleichsetzen	lebensende
15	großmaul	schule	hamas	geh	verpiss	antisemiteschtypisch	gleiche	dicke		schlägen
16	geil	antisemiteschutzversammelt	sofort	schande	linke	alter	rausgeeitert	jugendwiderstandschlägertruppe		
17	jens	feldmann	rassistischsolidarisch	selbstbestimmung	grassist	serbischen	sichs	la	kosovo	
18	fpö	scheiße	apologen	rassistischenkolonialen	bande	keineswegs	tragen	ideologien		links
19	verlag	tom	segev	ilan	pappe	schiller	propyläen	zweitausendeins	sand	gefält

## KJÖ – Kommunistische Jugend Österreichs

	0	1	2	3	4	5	6	7	8	9
0	rudi	golubich	gesucht	gerald	gerechtigkeit	gerhard	gericht	gerne	geschaffen	geschehen
1	kjö	jugend	kommunistische	österreichs	is	dinge	versetzen	nächste	katholische	gerne
2	www	https	com	youtube	watch	facebook	at	anbei	itjqxkur6mm	derstandard
3	eu	battlegroupsverteidigen	armee	defensiven	abzulehnen	kräfte	imperialismusneutralität	zahlreichen	at	
4	jakob	roth	folter	genauso	hätest	festen	überzeugung	verurteilen	unterdrückungmord	
5	bastian	monte	de	tatsachen	hätest	geschickt	form	interessieren	wären	sowjetunion
6	maurer	sigi	frau	stinkefingerklage		berichtet	peter	person	halt	gelten
7	danke	wien	moritz	gilt	steht	flores	fein	stefan	tatsächlich	menschenrechten
8	decrinispaul	kommentar	eh	brauchst	reicht	falsche	angriffspakt	komunisten	neides	
9	genozidverurteile	lager	verurteilt	derstandard	demografischeeinfach		frauenhäuserlangsam			at
10	thiel	cornelius	egal	ibiza	strache	star	wars	disney	twitter	vorbei
11	julian	schmidt	innen	nato	hab	generell	überfalle	tolle	udssr	genozids
12	grünen	links	partei	rechts	wahre	parlament	20	hält	meinst	sagen
13	genau	stehen	kommunisten	joseph	hartmann	sehe	dusek	dušek	demonstrant	demos
14	bauer	franz	zahlen	quelle	missverständenwerfen		hunger	kind	lügen	kapitalismus
15	the	of	tax	rich	politischen	group	to	worte	or	means
16	china	mal	klar	verteidigt	abgeht	kritisiert	kritisieren	junos	seite	indirekt
17	bitte	pau	li	verurteile	beweis	gehört	worten	verteidigung	junos	erwarteten
18	org	wiki	wikipedia	de	https	ad	scheinbar	deutsch	kooperation	geantwortet
19	andreas	auzinger	komm	völlig	anfang	schau	echt	xinjiang	html	abgeschlossen



## Greek

### Ελληνικά Hoaxes

	0	1	2	3	4	5	6	7	8	9
0	από	τις	μια	τη	ήταν	ένα	αλλά	έχουν	αυτά	όχι
1	facebook	com	https	php	story	www	posts	id	story_fbid	videos
2	hoaxes	ελληνικά	ellinika	ελληνικά	είστε	στα	λόγοι	εσείς	έχετε	αληθεία
3	είναι	όχι	αλλά	αλήθεια	πως	θέμα	έτσι	αυτά	μόνο	αντοί
4	σας	μπράβο	είστε	έχετε	εσείς	εσάς	κάνετε	ντροπή	τις	λέτε
5	youtu	be	https	lnunkp4lr6w	feature	twpjc1qzg84	nh1flotcop0	symg2fnyzws	r8uaec261lk	marg
6	τους	έχουν	τις	kánonv	autoí	στα	όλοι	αυτούς	όλους	λένε
7	σου	εσύ	έχεις	είσαι	λες	ξέρεις	πω	εσυ	θες	κάνεις
8	μας	πες	πείτε	έχουμε	στα	καλό	κρύβουν	αυτά	πως	5g
9	gr	www	https	ellinikahoaxes	ce	zouglia	ton	afistemenaziso	demosiografi	apantisi
10	ότι	λέει	είπε	υπάρχει	σημαίνει	λένε	ήταν	τώρα	κάτι	πρέπει
11	ρε	παιδά	φίλε	μπράβο	καλά	συ	φιλε	ουστ	άντε	αντε
12	μου	εγώ	έχω	φίλε	εμένα	είμαι	πες	κάνω	τη	εγω
13	watch	youtube	com	https	www	feature	share	43xzi6_abe	time_continue	3doejuvk8cm
14	αυτό	ακριβώς	γι	άρθρο	λέει	μόνο	εγώ	βίντεο	όχι	πρέπει
15	vai	αλλά	george	karydas	όχι	έχεις	νίκος	λέτε	πικος	έχουμε
16	της	τη	τις	μια	ως	αλήθειας	στα	εγει	σελιδάς	πως
17	γιατί	αφού	χαλάτε	κάτι	τώρα	στα	έχουν	λέτε	τέτοιο	μάσκα
18	έχει	κάνει	σχέση	σημασία	πει	πνυχίο	αυτός	μπορεί	γίνει	καμία
19	fake	news	demosiografi	zouglia	ton	ως	ήταν	ειδόηση	τόσο	ellinika

### Ομάδα της ιστοσελίδας ellinikahoaxes.gr

	0	1	2	3	4	5	6	7	8	9
0	μου	πως	εγώ	αλλά	έχω	όχι	γιατί	κάτι	ήταν	τις
1	com	https	facebook	www	watch	youtube	php	story	id	story_fbid
2	artsil	kopatz	wakandakis	βλέ	εγώ	ανταιλφηα	χαχαχαχχ	είμαι	σκαρλάτος	βλε
3	χρυσάνθη	κρ	έχω	μπλόκαρε	έβρισα	απαντάει	συνωμοσίας	σωστη	κλαίω	όμιλος
4	είναι	αλήθεια	όχι	θέμα	μπορεί	αλλά	αυτά	βίντεο	έτσι	αντή
5	vai	ισχύει	αλλά	george	είδα	όχι	οκ	απολύτα	προφανώς	είπε
6	αυτό	γι	ακριβώς	αλλά	λέω	συμβαίνει	λέτε	λέει	ισχύει	λέγεται
7	gr	www	https	ellinikahoaxes	gov	5g	gia	iefimerida	coronavirus	org
8	από	τη	πάνω	κάτω	τις	έχουν	όλα	μόνο	ένα	ιό
9	τους	τις	έχουν	έλληνες	κάνουν	αυτούς	απ	γιατί	σαν	αντοί
10	σου	έχεις	apazoglou	alexander	εσύ	είσαι	πω	μπορείς	μαζί	θες
11	marios	geront	xidous	μάριε	έχω	μετανάστες	νομίζετε	έλληνες	σκυλιά	στοιχεία
12	ότι	λέει	σημαίνει	υπάρχει	ισχύει	είπε	νομίζω	έχουν	δηλαδή	ήταν
13	maria	mindy	koumpanaki	melanie	lyda	ασφαλιστικες	μελέτη	σωστα	οχι	αρσβας
14	ce	cf	bf	el	wikipedia	org	b5	bd	b1	https
15	youtu	be	https	feature	watch	youtube	krondiras	βίντεο	nsyy5tqhvc	_gnh1as9aqy
16	της	ομάδας	τη	ήταν	ως	μια	ανάρτησης	αυτή	τις	στα
17	σας	ευχαριστώ	έχετε	καλή	πολό	κάνετε	είστε	εσείς	απάντηση	καλό
18	έχει	σημασία	σχέση	κάνει	καμία	μέσα	γίνει	αλλά	τις	καθένας
19	μας	στα	κρύβουν	τις	ζουν	ρε	εμείς	πες	πει	οχι



### ΠΟΤΕ ΞΑΝΑ ΑΡΙΣΤΕΡΑ

0	1	2	3	4	5	6	7	8	9
0 ουστ	βρε	κοπροσκυλακοπρόσκυλο	δω	μωρη	κοπροσκυλοκοπριτες	κοπριτη	κοπρόσκυλα		
1 είναι	ότι	αυτοί	αυτός	αυτή	έτσι	γιατί	όλοι	αυτά	μόνο
2 τους	όλοι	ολοι	αυτοί	έχουν	τις	γιατί	έλληνες	αυτούς	όλους
3 ρε	αντε	άντε	παιδιά	λες	παιδια	μαλακα	πας	ποιος	αι
4 μας	θεός	θεος	πει	λες	ελλάδα	πατρίδα	τις	γιατί	αρχιδια
5 σου	ντροπη	μαζί	ντροπή	σπίτι	εσυ	έχεις	εσύ	σπιτι	φτου
6 της	ελλάδας	αριστεράς	αριστερας	κοινωνίαςελλαδας	χώρας	κοινωνιαςτις			ελλάδος
7 χα	τέλειο	ραγικουσης	γέλια	χου	καλοοοοοοκυρια		μιλανε	ναι	γελάσαμε
8 από	ότι	ελλάδα	ένα	τις	ήταν	αλλά	άλλο	γιατί	μια
9 πολύ	καλό	σωστά	καλά	τόσο	πάρα	πιο	σωστό	ακόμα	ωραίο
10 σας	ντροπη	ντροπή	μαζί	όλοι	έχετε	είστε	εσείς	πάτε	μπραβο
11 μου	θεέ	παναγια	παναγία	φίλε	έλα	ντε	θεε	πάρετε	έκτακτη
12 βρε	λες	αντε	μωρη	άντε	παιδιά	παιδια	μωρή	διαολο	μαλακα
13 είσαι	εσύ	πόσο	καλά	βλάκας	ότι	πρέπει	έλληνας	ένα	μεγάλος
14 αυτό	γι	μόνο	ακριβώς	ήταν	κάνει	άλλο	λέει	γιατί	θέλει
15 στα	τσακιδιασκατα	τσακιδια	σπίτια	μούτρα	αρχιδια	αντε	παπαριασκουπίδια		
16 τη	χώρα	αυτή	λες	χωρα	μια	φατσα	κάνει	μονο	φάτσα
17 μπράβο	έτσι	ακριβώς	συγχαρητήριααξιος	φίλε	σωστά	χίλια	τέλειο	κυρανακη	
18 έχει	αυτός	κάνει	γιατί	ότι	λέει	καλά	αυτή	τώρα	ξέρει
19 μαλακαςπολυ	ποσο	πιο	πόσο	πεθαινεις	μεγάλος	μεγαλος	αυτός	πεθαίνεις	

### Καταγραφή Γεγονότων (Ομάδα)

0	1	2	3	4	5	6	7	8	9
0 είναι	αυτή	αυτοί	αλήθεια	αυτά	ότι	αυτός	θέμα	έλληνες	όμως
1 τους	στα	όλους	αυτοί	κάνουν	αυτούς	ας	έλληνες	θέλουν	τωρα
2 μας	στα	τώρα	εμείς	χώρα	ντροπή	ελλάδα	κάνει	τις	πως
3 ρε	λες	αντε	καραγκιόζηαι		παιδιά	τραβα	γαμησου	είσαι	οξω
4 πολλά	χρόνια	αληθώς	ανέστη	κύριος	νίκο	dimitriouεπίσης		milos	santorini
5 σας	μαζί	στα	γαμω	εσείς	σπίτια	είστε	παναγία	παναγιά	έχετε
6 σου	μαζί	εσύ	σπίτι	εσυ	στόμα	λες	είσαι	κάνεις	συμφωνώ
7 μπράβοσυγχαρητήριαπαλικάρι	μαζί	καλά		λεβέντη	λεβέντες	άνθρωπο	μεγάλο	παλληκάρια	
8 https	com	www	facebook	gr	watch	youtube	php	story	posts
9 kai	na	to	ta	8a	den	mas	gia	apo	me
10 από	γιατί	ότι	έχει	αλλά	όχι	μόνο	ελλάδα	καλά	πολύ
11 ουστ	κοπρόσκυλα	βρε	μωρή	κοπριτες	προδότες	maria	προβλήματαρεεεε		μπραβω
12 της	στα	ελλάδας	τις	κοινωνίας	μουνι	αρέσει	φατσα	δρομο	μανας
13 έτσι	ακριβώς	καλημέρααλλιώς		δυστυχώς	χαχαχαχαεγώ	λέτε	έχουν	χρηστος	
14 kostas	lamprinos	οκ	καλά	δούμε	στείλε	ειπες	πεστα	αναμάρτητος	θεος
15 μου	φίλε	θεέ	ντε	έλα	εγώ	εμένα	παιδιά	παναγία	είσαι
16 ναι	πολύ	φυσικά	ότι	πεις	όχι	βεβαίως	προίκα	ξεσηκωθουμεσυμφωνώ	
17 αντό	γι	κάνει	λέει	αλλά	ούτε	κάνουν	άλλο	ότι	γίνεται
18 όλοι	μαζί	πρέπει	προδότες	φύγουν	αυτοί	τσακιδια	όπως	στα	ιδιοι
19 άντε	βρε	γαμησου	αντε	καραγκιόζημαλακά	ái	σαλτα		μαλακες	διάολο



## Italian

### Matteo Salvini

0	1	2	3	4	5	6	7	8	9
0 matteo	ciao	mollare	caro	unico	bravissimo	abbraccio	buonanottebenedetto	papà	
1 capitano	mollare	ciao	unico	buonanotte	fianco	grandissimo	bravissimonotte	abbraccio	
2 salvini	lega	meloni	colpa	unico	viva	premier	mollare	bravissimocaro	
3 governo	incapaci	cadere	colpa	attuale	opposizione	incompetenti	incapace	andare	votato
4 buona	domenica	giornata	serata	notte	pasqua	famiglia	palme	ciao	abbraccio
5 auguri	tanti	tantissimi	papà	compleannocuore		campione	bellissima	piccola	mamma
6 italia	povera	viva	europa	rovina	giustizia	bisogno	via	germania	ormai
7 buongiornogiornata		domenica	sabato	settimana	inizio	caffè	abbraccio	mirta	felice
8 casa	andare	stare	devono	mandiamo	mandare	porti	mandarli	chiusi	porta
9 italiani	prodotti	vogliono	immigrati	clandestini	soldi	poveri	stranieri	cittadini	veri
10 https	com	www	facebook	groups	2547165505541349share		ref	it	posts
11 buon	lavoro	pomeriggio	ferragosto	compleanno	sabato	natale	senso	famiglia	appetito
12 conte	maria	soldi	de	pd	vuole	sinistra	andare	ce	lavoro
13 vergogna	schifo	italiana	giustizia	vera	veramente	dignità	assoluta	fate	povera
14 nn	mollare	ke	tt	so	xche	sn	xke	capito	capisco
15 parole	fatti	sante	belle	poche	giuste	tante	servono	vento	vogliono
16 vero	papa	problema	cuore	unico	assolutamente	dici	politico	amore	sai
17 ragione	perfettamente	pienamente	dedo	vendere	sindaco	fattene	porro	fatevene	giletti
18 accordo	pienamente	condivido	concordo	bravissimo	perfettamente	assolutamentepensiero	francesca	giuseppe	
19 italiano	popolo	orgoglio	politici	scrivere	sovraano	politico	piazza	impara	vergogno

### Alessandro Di Battista

0	1	2	3	4	5	6	7	8	9
0 auguri	tantissimi	tant	filippo	cuore	benvenuto	ragazzi	notizia	piccolo	papà
1 pd	lega	alleanza	renzi	partito	alleati	accordo	bibbiano	voto	votato
2 alessandro	caro	ciao	stima	gissi	giancarlo	bisogno	michela	bravissimo	bentornato
3 https	com	www	facebook	groups	php	it	posts	2547165505541349	watch
4 marco	giustizia	fatta	genitori	speriamo	spero	crescenzo	ragazzo	eremita	ramagini
5 ale	ciao	manchi	ritorna	bisogno	bentornato	stima	campo	tornato	tanti
6 battista	caro	accordo	politico	paragone	annunziata	parlamento	maio	lavorare	raggi
7 movimento	bisogno	paragone	anima	partito	tornare	morto	tradito	mano	rimasto
8 politica	nn	italiani	soldi	maio	capito	see	more	lavoro	cose
9 famiglia	bella	bellissima	complimenti	notizia	auguroni	augurissimi	abbraccio	filippo	meravigliosa
10 dibba	caro	aspettiamo	bravissimo	grandissimo	stimo	tornato	ciao	maio	idee
11 rossi	giuliano	spese	ignorante	parlamentare	coglione	italiani	cervello	sai	daniele
12 governo	cadere	renzi	conte	opposizione	berlusconi	maggioranza	soli	amici	andare
13 m5s	votato	maggioranza	tradito	partiti	legge	deluso	colpa	unico	conte
14 italia	europa	povera	bisogno	italiani	pensa	viva	fratelli	uscire	ue
15 stelle	votato	votare	riveder	voterò	voto	m5	viva	cdx	unica
16 salvini	meloni	berlusconi	renzi	lega	vota	caciotte	ciliege	mangiare	votare
17 andrea	romano	aquilino	sciallero	tosatto	cantini	francioni	giammarini	scanzi	losi
18 torna	bisogno	presto	guerrero	prego	campo	politica	manchi	aspettiamo	aspetta
19 fabio	tordo	salamida	massimi	de	pancrazio	gobbo	soldi	cognome	albanesi



### *Giorgia Meloni*

0	1	2	3	4	5	6	7	8	9
0 giorgia	cara	ciao	unica	grandissimamollare	donna	bellissima	continua	migliore	
1 buona	domenica	giornata	serata	palme	famiglia	ciao	bellissima	felice	guarigione
2 italia	fratelli	viva	povera	lega	rovina	bisogno	germania	giustizia	nazione
3 meloni	salvini	signora	onorevole	leader	on	mes	presidente	premier	firmato
4 https	com	www	facebook	groups	2547	share	ref	posts	it
5 governo	incapaci	opposizionecadere	attuale	colpa	salvini	popolo	votato	incapace	
6 buongiorno	domenica	cara	bellissima	carissima	onorevole donna	felice	settimana	splendida	
7 auguri	tanti	tantissimi	guarigione	compleannomamma	pronta	cuore	donna	bellissima	
8 conte	vero	popolo	soldi	sinistra	italiano	politica	destra	salvini	sa
9 complimentibellissima	bellissimo	donna	bravi	genitori	bravissimo	medici	bravissimibimba		
10 famiglia	condoglianzesente	pace	riposa	abbraccio	rip	altrettanto	pasqua	famiglie	
11 bravissima	bellissima	chiara	unica	condivido	preparata donna	leader	38w	continua	
12 casa	andare	stare	mandare	devono	mandiamomandarli	mandiamoliandate	chiusi		
13 italiani	prodotti	veri	soldi	immigrati	vogliono cittadini	politici	poveri	stranieri	
14 buon	lavoro	ferragosto	compleannosenso	settmana	pomeriggio giornata	viaggio	sabato		
15 onore	forze	ordine	arma	carabinieri	caduti	uomini	divisa	eroe	patria
16 parole	fatti	sante	belle	tante	servono	opposizionepoche	vogliono	chiacchiere	
17 europa	uscire	usciamo	dobbiamo	germania	euro	referendum serve	mes	via	
18 bella	donna	ciao	figura	notizia	persona	foto	bimba	bellissima mamma	
19 nn	ke	so	tt	sn	xche	capito	credo	ce	xke

### *Potere al Popolo*

0	1	2	3	4	5	6	7	8	9
0 popolo	potere	abuso	cubano	voto	post	sardine	vota	pagina	viva
1 lavoro	problema soldi	cose	post	giusto	lavorare	nn		lavoratoricredo	
2 https	www	com	it	facebook	posts	youtu	be	watch	youtube
3 vergognalega	italiana	parlare	viviamo	umano	sud	mo	giustizia	pietà	
4 merda	pezzo	pezzi	figura	fascista	uomini	bella	leghista	montagnafascio	
5 schifo	fate	persona	totale	assoluto	razzismo	veramente	terenzi	bello	lega
6 cuba	viva	embargo	fidel	gracias	cubano	cuore	cina	mille	que
7 gemma	roin	levis	gentile	эммануэла	esaltare	esaltato	esaltati	esaltarvi	esalta
8 sinistra	pd	destra	partito	voto	lega	votare	pap	sardine	vera
9 vero	problema verosimileesiste	almeno	credo	giorgio	pericolo	giusto	spero		
10 bravi	compagniragazzi	1y	complimenticuore	continuate	siciliani	iniziativa	culo		
11 italia	povera	viva	fratelli	italiani	europa	tasse	fca	succede	vuole
12 salvini	meloni	governo	matteo	renzi	conte	sud	post	voti	berlusconi
13 cazzo	testa	rotto	dici	teste	frega	entra	sti	capito	sai
14 grandi	compagni	ly	ragazzi	sibilla	patrimonipiccoli	bravissimievasori	compagne		
15 de	luca	filippo	bellis	martino	josi	angelis	magistris	campania	simone
16 andrea	martini	achille	gattucciorusso	barone	furlan	musso	zimbaro	aliperti	
17 marco	edoardo	beghi	felli	vitale	iacomelli colombo	stazi	biagi	franchi	
18 ordine	forze	violenza	polizia	armate	cittadini	abuso	video	divisa	regole
19 casa	stare	andare	pound	lavorare	bisogna	andate	uscire	rimanere	devono



## Dutch<sup>3</sup>

5G

0	1	2	3	4	5	6	7	8	9
0 nooit	gehaaldgriep	grieppri	gehad	lijf	30	gelukkig	jaar	vaccin	
1 precies	jaren	bestaat	da	verhaal	honden	gelogen	bedrogen	open	gewijzigd
2 denk	nou	beschikbaarhen	etc	succes	vaccin	krijgen	stel	griepniet	
3 dank	bron	betreur	achteraf	geworden	enige	eerder	wakker	echt	info
4 verstandig	heel	hoor	druipt	afschuwelijke	domheid	wonderen	geweldig	hopen	schrijft
5 story_fbid	story	php	id	943762986131354100014929840666facebook		com	leo	amp	
6 geld	betalen	eigen	begraveniskeuze		betaald	hoop	mensen	willen	zwakbegaafde
7 delen	bedankteven	ga	ervaringen	wakker	openbaar	mensen	opslaan	hopelijk	
8 jaar	gewooneerste	zoiets	vorig	kreeg	ziek	komt	grieppri	pak	
9 helemaal	enten	laat	mee	kunt	jou	5g	papiertje	gek	hebt
10youtu	be	watch	youtube	feature	com	5sesrr113ic0	wuhbyhe8mdg3124	tippel	
11 vrijwilligersplacebomenzen			konijnen	betaalbare	vergoedingen	geaccepteerd	blijken	bijwerkingsnechte	
12 gedaan	graag	lockstep	18	19	zitten	lezen	nederland	betaald	overstelp
13 risico	patiënt	patient	laten	boerstoel	janny	beurt	sla	vaccineren	komt
14 moeten	jullie	kijken	kopieer	deel	allen	samen	zorgen	baas	lijf
15 zegt	jaaaa	eh	genoeg	nergens	pech	verhalen	verkeerde	schadelijke	mocht
16 vooral	niks	hebt	echt	luisteren	lichaam	belangrijk	leren	getest	testen
17 sputen	anders	sput	raken	kwijt	echte	ziek	lieve	genomen	ingespoten
18 jij	staat	blijf	prikje	vooraan	uitdelen	slapen	krijgt	nederlandse	jou
19 griepprik	zelfde	schuitje	neem	zeker	riscogroep	exact	man	zit	weer

## fake news

0	1	2	3	4	5	6	7	8	9
0 fake	news	cnn	zoals	hopelijk	again	zeg	zegt	leugens	staat
1 vrt	news	fake	doet	vtm	nieuws	allemaal	stop	potje	nws
2 mensen	zien	maken	sommige	geloven	watertekort	laten	verder	media	gedeeld
3 opdracht	gemeente	afleid	hoorbaar	binnenkort	islamiseringaandacht		belgie	stad	belgische
4 volgens	trump	allemaal	getuigen	watertekort	cnn	houden	bezigt	bos	zakken
5 weer	potje	afgedaan	zeggen	verdoezelenhln		groen	verdronkenzoals		nee
6 the	media	at	work	to	and	are	that	you	my
7 desinformatiepoging		afkomt	mooie	informatie	links	feiten	wegens	facebook	helemaal
8 jullie	geloven	he	zoals	denken	vertellen	weten	beter	beschaamd	vernoemen
9 trump	zeggen	news	fake	gelijk	woorden	liegen	jaja	media	das
10 echt	misschien	laatste	staat	zever	nieuws	raadplegen	ineens	globale	schaal
11 verspreiden	kent	nee	vernietigd	wegens	best	zoals	rechts	rechtse	onschuldig
12 corona	sheep	maken	the	haha	zwijgen	virus	leggen	geld	ontploft
13 gaan	stop	gecontroleerd	geloven	zorgt	opzoeken	mee	etc	mag	mogen
14 hoop	vrees	bericht	probleem	drama	enorme	goedgelovigheidkaren	verheyen	schrijft	
15 jij	geloof	zoals	verspreidt	net	bunker	ga	zegt	wonen	niks
16 linkse	extrem	rechts	komt	links	media	zender	stop	strategie	leugenpers
17 zeker	landen	gezien	hoax	vermoed	wereld	missen	verspreid	volg	ingenomen
18 man	belachelijk	maakt	fantastischeafdoet	problemen	ofwel	ok	verdwijnenjezus		
19 gewoon	gecontroleerd	mee	rapporteren	watertekort	zien	ziekelijk	vaccin	angst	huis

<sup>3</sup> We aggregated some pages/groups in order to improve topic consistency.



### *zwarte piet*

	0	1	2	3	4	5	6	7	8	9
<b>0</b>	zwarte	piet	blijft	sinterklaaspieten	anders	leve	sint	mag	hoort	
<b>1</b>	mee	doe	bezig	bemoeien	klaar	mis	niks	idd	doet	helemaal
<b>2</b>	zwart	blijft	roet	wit	bruin	blijf	piet	grijs	schoorsteenhou	
<b>3</b>	ga	land	lekker	eigen	anders	terug	werken	nuttigs	laat	ergens
<b>4</b>	geweldig	hoort	man	nieuws	respect	super	houden	vind	horen	gedaan
<b>5</b>	blijven	moeten	traditie	natuurlijk	thuis	zwart	zwartepiet	kinderfeestzwarte		feest
<b>6</b>	helemaal	gelijk	klopt	top	gek	hebt	mee	gerard	nergens	slaat
<b>7</b>	gewoon	moorkop	hoort	laten	doorgaanzoals	vieren	hoor	negeren	boycotten	
<b>8</b>	jij	bent	hebt	jou	doe	weet	vind	jouw	denk	zegt
<b>9</b>	zeker	weten	weet	nooit	gehad	kind	zeer	zeggen	bang	moorkop
<b>10</b>	jullie	stilletje	maken	idioten	bezig	trots	racisten	vinden	schaam	willen
<b>11</b>	mensen	nederlandracisme	heel	maken	moeten	alleen	mag	niks		willen
<b>12</b>	inderdaaderg	mening	heel	dacht	stijl	durven	prima	nou		schreeuwen
<b>13</b>	precies	bedoel	hetzelfde	gezegd	anders	zegt	dacht	nou	genoeg	zeg
<b>14</b>	belachelijk	woorden	gedoe	zeg	maakt	gemaaktkinderfeestjeonzin		jezelf		maken
<b>15</b>	echt	nou	he	hoor	vind	erg	hè	triest	bent	zielig
<b>16</b>	onze	cultuur	traditie	af	tradities	blijf	moeten	afblijven	nederlandseeigen	
<b>17</b>	gaan	weer	laat	tijd	laten	nou	moeten	werken	anders	toe
<b>18</b>	nee	hoor	natuurlijkoh	idd	joh	mogen	lekker	gelukkig	jou	
<b>19</b>	weg	ermee	bol	com	kwijt	lui	juist	tuig	kopz	willen



## Related terms

A relatively new technique in Natural Language Processing is the use of word embeddings. Large volumes of documents are used to observe words in their context and count the number of co-occurrences. After dimensionality reduction, the result is a *word vector* for each word, a limited series of numbers that implicitly describes the syntactic and semantic properties of the words. Having words represented as numbers, allows us to perform “calculations” with words, most notably calculating their “similarity”. The word vectors allow us to automatically calculate the similarity of the words “cat”, “onion” and “dog” and observe that the word “cat” is more similar to “dog” than it is to “onion”. This effectively establishes a world knowledge model for a given data set in an unsupervised manner, as no further assumptions are made on the data.

We constructed such models for the languages in the COMMIT-dataset. This now allows us to get a glimpse into how the language areas semantically relate words to one another. In the example below we show the most similar words in the Dutch data for the word “neger” (archaic/offensive term to refer to a black person). It is clear that most of the most similar words are offensive (indicated in red).

source	Most similar
neger	negerin kleurling hoer negers klootzak snol kaaskop mongool kerel viespeuk blanke nikker marokaan dikzak bleekscheet trut mokkel marokkaan roetmop indiaan surinamer aziaat smeerlap mafkees baardaap halfbloedje antiliaan hufter spleetoog viezerik gozer

In essence this means that a racist word knowledge model has been established. This may be an uneasy thought: we effectively trained a machine to be racist, homophobic and misogynistic. But we can now use these models to understand the hive mind of the community under investigation. We repeat the exercise for the word “neger”, but this time in German:

source	Most similar
neger	pimmel hure zigeuner türke wixer macker lesbe teenis afrikaner euter flicker popo pisser tittchen türken arschlöcher araber möpse muselmann tusse polizist kerlen maenner rassisten blondinen luder drecksau rassist hurensohn ferkel huren dödel aersche köter

Furthermore, the technique allows us to expand the POW-lexicons for the languages of Greek and Italian, using the seed words lists compiled by the project partners, as illustrated below.

source	Most similar
frocio (IT: “faggot”)	puttaniere culattone cretino schifoso pedofilo guardone esibizionista furbetto lurido frocione schifo feticcio froccetto puttanone cinghialone segaiolo maschione disobbediente feticista procio ingoi culone pestaggio ingordo fetentone sporcaccione fustigatore
Πούστης (GR: “faggot”)	πούστη μισογύνης ρεζίλης μακάκας βλάκας ψυχάκιας χέστης κούλης μάπας άκας μπαμπέσης μπούλης παπάρας βλαμμένος χαζούλης άντρακλας κάνας φτυστός αυνάνας κοπρίτης συκοφάντης παλιόπουστας εργένης βρωμιάρης γιόκας αρχιψεύταρος



## Don't be afraid to ask

Another interesting avenue into the collective mind, is to understand what people are uncertain about. This can be done by searching for questions in the data. *why*-questions are of particularly interest in this exercise. We collected all questions from the data and clustered them thematically. Below you will find the most common type of questions for each language.

### Greek

- Γιατί δεν τους βοηθούν μέσα στην ίδιες τους τις χώρες?
- Γιατί οι ΜΚΟ δεν κάνουν το <<φιλανθρωπικό>> τους έργο μέσα στις χώρες των μεταναστών ώστε να μην χρειάζεται να φύγουν από τις πατρίδες τους
- Γιατί δε ρώτησαν αν θέλουμε μνημόνια?
- Γιατί η κάθε Γεροβασίλη δεν παίρνει κανένα από δάντους στο σπίτι της? Οι πολιτικάντηδες βουλευτές έχουν πολλές φοροαπαλλαγές. Εμείς οι υπόλοιποι τι φταίμε να πληρώνουμε φόρους και άντε πάλι φόρους? Πότε θα τελειώσει αυτό το γαϊτανάκι?
- Γιατί προσβάλλεται αυτόν τον αγνό και καλοπροαίρετο ήρωα, βαφτίζοντας με το όνομα, έναν αδίστακτο πολιτικό απαταιωνά?
- Γιατί δεν πάνε σε πλούσιες μουσουλμανικές χώρες?
- Γιατί είναι αλλαγμένη;;??;;Είσαι ανθέλληνας τσιράκι του Σόρος?
- Γιατί υποτιμούν τη νοημοσύνη μας?
- Γιατί δεν αφήνετε την πληροφορια να κινείται ελευθερα τρόμο χοαξ της νεοταξικης εποχής δηλαδή εσείς μπορείτε να διακρίνετε την αλήθεια και εμείς όχι?
- Γιατί έχουν αυτή την ανθεκτικότητα στο να αρνούνται οτιδήποτε mainstream αλλά δέχονται αμέσως οποιαδήποτε φωτοβολίδα αρκεί να είναι alternative?
- Γιατί η ελληνική κυβέρνηση Η κάθε κυβέρνηση δεν αρνείται το τσιπάκι?
- Γιατί οι αστυνομικοί οι πολιτικοί και άλλες ομάδες δεν είναι σε καραντίνα???

### Italian

- Perché voi di sinistra, che vi ritenete tanto intelligenti e istruiti, credete ai sondaggi tutti pilotati?
- Perché è normale che in un paese civile al governo ci siano persone non elette da nessuno?
- Perché non si vuol capire il modo in cui il virus agisce?
- perché non parla dei 600.000.000 spariti in una notte?
- Perché questi invasori non devono rispettare le leggi?
- Perché voi tutti, nessuno escluso, non vi dimezzate gli stipendi a vita e rinunciate ai vostri privilegi per sempre?
- Perché dobbiamo abbassare la testa ad un'Europa che non vede l'ora di metterci i piedi in testa perché i crucchi e i senza bidet accompagnati da quelli che si fumano pure le "palle" scusami il termine non arriveranno mai alla grandezza dell'Italia?
- Perché le forze dell'ordine intervengono solo dopo che avvengono certi episodi e non fanno prevenzione "liberando" queste zone dalle nefaste presenze che se ne sono impossessate?



- perché a maggio non andiamo in piazza a protestare pacificamente contro questo governo?
- Perché oltre che condividere SOLO post riguardanti gli immigrati non parli dei 4 miliardi di euro tagliati all'istruzione?
- Perché bisogno essere duri con questi sfollati?
- Perché i meccanici e gli artigiani cosa vanno a fare al lavoro se non c'è nessuno a prendere il virus?

#### German

- Warum sollen wir Flüchtlinge aufnehmen?
- Warum ist man rechtsradikal, wenn man an die Ermordung von Weißen erinnert?
- Warum wird die Merkel nicht angeklagt?
- Warum habt ihr bitte alle ein Problem damit das ihr beim einkaufen kurz a Maske Aufsätze müsst?
- Warum werden solche noch in unser Land gelassen??
- Warum redet niemand über den Rassismus der Schwarzen Afrikaner gegen die Weißen Afrikaner?
- warum kämpfen sie nicht so Zuhause gegen die Islamisten, vor die sie angeblich flüchten?
- Warum sollte ich mich für etwas impfen, was ich nicht möchte?
- Warum sollten wir uns immer nach denen Richten wenn es in ihren Ländern Null Toleranz andersgläubigen gegenüber gibt, wer so leben will kann ja dortbleiben wo er herkommt oder?
- Warum ist die Türkei noch bei der NATO?
- Warum wird immer genau bei \* TSCHETSCHENEN \* die NATIONALITÄT genannt?
- Warum sollte man einem ungeübten Inländer mehr als einem routinierten Ausländer bezahlen?
- WARUM werden Kriminelle nicht abgeschoben?

#### Dutch

- waarom wil dat tuig toch constant anderen voorschrijven wat hun mening moet zijn?
- Waarom moeten wij ons aanpassen en die rare gasten niet?
- Waarom mengen jullie je in Onze Nederlandse Binnenlandse aangelegenheden gooien jullie olie op het vuur van een discussie die in Nederland volledig gepolariseerd is?
- Waarom kiezen jullie partij voor dat hysterisch gillende BLM dat daarentegen racistisch is tegenover de Blanke Nederlanders?
- Waarom zwarte piet verbannen als je zelf liever als zwart wordt beschouwd dan bruin of donker?
- Waarom is het in Europa verboden om ook maar iets positief te zeggen over Trump?
- Waarom lijkt het dan of "alle" Europese media zeggen dat alle Amerikanen dom zijn, en alle Europeanen slim?
- Waarom worden moslims eigenlijk altijd gewelddadig en voelen zij zich gekrenkt als hun geloof, god of profeet wordt beledigt?



- Waarom kunnen wij nu eens niet met zijn allen 'gewoon' doen, minder aan onszelf denken en een beetje rekening proberen te houden met onze medemens?
- Waarom is een Antwerpenaar meer waard dan een andere Vlaming?
- Waarom speciale regels voor zwembaden, waarom hoofddoeken op officiële diensten, waarom levenslang leven op kosten van uitkeringen in je gastland, waarom onze meisjes uitschelden voor hoeren omdat ze zich westers kleden, ze desgevallend aan te randen waarom onze oudjes beroven op straat?
- Waarom mensen haten van een andere afkomst?
- Waarom wordt Israël nooit terecht gewezen voor de misdaden die het begaat?
- Waarom zo 'n drastische maatregelen als het niet erger is dan een griepje?



## **Islamic Radicalization**

Social media posts dealing with the theme of Islamic radicalization are becoming increasingly hard to source nowadays, as is apparent from the list of resources outlined above. Such content is almost by definition not shared on public pages anymore, and organizations such as Daesh are nowadays much less active on “open” platforms such as Twitter and Facebook in favor of recruitment on private communication channels such as Telegram and WhatsApp. Furthermore, social media sites have become a lot more proactive in removing such content, so that any sites that do pop up have an extreme short life span. We have only been able to capture some Arabic and English content so far. This unfortunately means there is very little current data to work with and many of the techniques used for the other communities are not applicable to this data set.

To understand the specific vocabulary of Islamic extremism, we used the Meerkat approach to compare in-domain data to more general datasets. This provides some insight into the specific vocabulary of the domain. Typical words such as “kuffar” are present in the data, but seemingly neutral words, such as coconut and lion are also overrepresented in the data, as they have a very specific meaning in the domain, as outlined below.

<b>KAFIR, KUFFAR, KUFR</b>	unbeliever(s)
<b>SHIRK</b>	idolatry = worshipping other gods but Allah
<b>MUJAHID</b>	jihad warrior
<b>MUHAJIR</b>	jihad warrior from a foreign country
<b>LION</b>	jihad warrior with bravery
<b>MURTAD</b>	a bad Muslim raised by good Muslims
<b>TAKFIR</b>	a bad Muslim
<b>UMMA</b>	<i>roughly:</i> the unity of all Muslims across states
<b>DAWLA</b>	<i>roughly:</i> a temporary state to promote Umma
<b>KHILAFAH</b>	the caliphate
<b>COCONUT</b>	takfir (i.e., brown outside but white inside)
<b>DOGS, PIGS</b>	US or European Christians or atheists
<b>KILL</b>	-

We will continue to attempt to source and harvest additional data for this use case and update the analyses accordingly.





# #COMMIT

COMMunication campaign against exTremism and radicalisation

 CENTRO SVILUPPO  
CREATIVO  
DANILO DOLCI

 kmnop  
KINDING A BETTER WORLD

 die Berater  
ZUKUNFT LERNEN

 La Benevolencia

 textgain

 UNIVERSITÀ  
DEGLI STUDI  
DI PALERMO



This publication was funded by the European Union's Internal Security Fund – Police under Grant Agreement No. 867019.

