
GAB.AI: HET PLATFORM WAAR HATE SPEECH EEN RECHT IS

ONDERZOEK NAAR DE DOELWITTEN EN IDEOLOGISCHE GRONDSLAG VAN HATE SPEECH
OP HET ONGECENSUREERDE SOCIALE MEDIAPLATFORM GAB

T.J. ten Heuvel
t.j.heuvel@students.uu.nl
Juni 2019

BA-eindwerkstuk, Communicatie- en informatiewetenschappen
Universiteit Utrecht
Begeleider: M.T. Schäfer
m.t.schaefer@uu.nl

ABSTRACT

Dit onderzoek tracht hate speech op het sociale mediaplatform Gab te identificeren, beschrijven en contextualiseren. Hiervoor is geanalyseerd wie de doelwitten van hate speech zijn, wat de ideologische posities van de sprekers zijn, en waar de sprekers deze posities op baseren. Met behulp van kwantitatieve tekstanalysemethoden is duidelijk geworden dat het discours op Gab overladen is met hate speech. Op het platform heerst een hoge mate van ideologische homogeniteit, waarbij gebruikers zelden worden blootgesteld aan content die ideologische grenzen doorbreekt, maar veel vaker gevoed worden met content die de politieke- en ideologische overtuigingen van gebruikers bevestigen. Het platform fungeert daarmee als een ‘echo chamber’ voor mensen met conservatieve, racistische, wit-nationalistische, antisemitische, en extreemrechtse opvattingen. Deze opvattingen worden bevestigd en versterkt door alternatieve en populistische media en complottheorieën. De meest opvallende doelwitten van hate speech op Gab zijn joden, moslims en zwarten. Bovendien zetten gebruikers zich fel af tegen politieke correctheid, mainstream-media en links-liberalen. Een opvallende paradox in het discours op Gab is dat gebruikers enerzijds sterk pleiten voor free speech en zich fel afzetten tegen censuur, terwijl zij tegelijkertijd mensen met linkse, liberale politieke opvattingen het zwijgen opleggen en aanvallen. Free speech lijkt hierdoor gebruikt te worden als rechtvaardiging om racistische en antisemitische uitlatingen te verspreiden. De manier waarop Gab-gebruikers free speech interpreteren heeft daarom in werkelijkheid weinig te maken met free speech, maar eerder met rechtvaardiging van extremisme, racisme, antisemitisme- en dus hate speech.

“Fuck YOU KAFFIR! Why are you here? To intentionally destroy and disrupt a white site with your fucking nigger bullshit!” Get AIDS and DIE!”

INHOUDSOPGAVE

ABSTRACT	1
INLEIDING	3
HATE SPEECH	5
FAR-RIGHT ONLINE SUBCULTUREN	5
ONLINE ECHO CHAMBERS	6
GERELATEERD ONDERZOEK	7
METHODEN	10
DATASET	11
ANALYSE	13
TERM FREQUENCIES	13
HASHTAGS	14
N-GRAMS	16
DOMEINEN VAN GEDEELDE URL'S	17
HATE SPEECH & TOXICITY	18
TERM FREQUENCIES IN TOXISCHE POSTS	19
SAMENHANG VAN WOORDEN MET TOXICITY SCORE	21
ANTISEMITISME	24
RACISME	25
CONSERVATISME	26
CONCLUSIE & DISCUSSIE	28
REFERENTIES	30
BIJLAGEN	33
BIJLAGE 1 - TOP 150 WOORDEN MET DE HOOGSTE FREQUENTIE OP GAB	33
BIJLAGE 2 – TOP 150 HASHTAGS MET DE HOOGSTE FREQUENTIE OP GAB	34
BIJLAGE 3 – TOP 100 N-GRAMS MET DE HOOGSTE FREQUENTIE OP GAB	35
BIJLAGE 4 – TOP 50 MEEST GEDEELDE DOMEINEN OP GAB	36
BIJLAGE 5 – TOP 50 POSTS MET DE HOOGSTE TOXICITY SCORE	37

INLEIDING

In juli 2016 kwam de rechts-conservatieve Breitbart-redacteur Milo Yiannopoulos in opspraak nadat hij permanent verbannen werd van Twitter wegens het schenden van de richtlijnen van het platform (Ohlheiser, 2016).¹ De ban volgde kort nadat Yiannopoulos zich meerdere malen racistisch uitgelaten zou hebben tegenover de zwarte *Ghostbusters*-actrice Leslie Jones. In een reactie aan Breitbart stelt Yiannopoulos dat zijn ban het bewijs zou zijn van Twitter als een ‘no-go zone’ voor conservatieven, en dat iedereen die waarde hecht aan vrijheid van meningsuiting niet welkom zou zijn op het platform (Hunt, 2016).

Kort na Yiannopoulos’ ban in augustus 2016, richtte de zelfbenoemde conservatief Andrew Torba *Gab* op: een nieuw sociaal mediaplatform dat in gebruik lijkt op Twitter en haar missie omschrijft als ‘het op één zetten van mensen en vrijheid van meningsuiting’ (“gab | Community Guidelines”, z.d.). Met de oprichting van Gab uit Torba kritiek op andere sociale media, die hij verwijt conservatieve ideeën structureel te censureren en te onderdrukken onder het mom van hate speech en intimidatie (Wendling, 2016). Hoewel Torba stelt dat Gab geen exclusieve ontmoetingsplaats is voor republikeinen, rechts-conservatieven of wit-nationalisten, kenmerkten nieuwsmedia het platform al als ‘Twitter for racists’ (VICE News, 2016) en een ‘social media alternative that attracts users banned from Twitter’ (Wilson, 2016). Ook werd de app verbannen uit de Google Play Store wegens *hate speech* (Price, 2017), waarvan de aanwezigheid al aangetoond werd door meerdere onderzoekers (Lima et al., 2018; Zannettou et al., 2018).

Een nieuw platform als Gab brengt vragen met zich mee omtrent de praktijken die erop plaatsvinden. In het bijzonder omtrent hate speech, gezien de gebleken controversen rondom de grens met vrijheid van meningsuiting. Dit onderzoek tracht daarom haatdragende interactiepatronen van Gab-gebruikers bloot te leggen. Zo is er onderzocht wie de doelwitten van hate speech zijn, welke ideologische posities sprekers innemen, en aan welke (nieuws)media zij refereren. Het betreft een exploratief onderzoek dat de (haatdragende) praktijken op Gab identificeert, beschrijft en contextualiseert. De resulterende hoofdvraag uit deze informatiebehoefte luidt daarom:

Hoe kunnen haatdragende interactiepatronen van gebruikers op Gab gekenmerkt worden?

Om deze algemene hoofdvraag te kunnen beantwoorden, is dit onderzoek onderverdeeld in deelvragen die specificeren hoe de haatdragende praktijken op Gab beschreven zullen worden:

- A) *Wie zijn de doelwitten van hate speech op Gab?*
- B) *Wat zijn de ideologische posities van Gab-gebruikers?*

¹ Aangezien Twitter geen uitspraken doet over de specifieke reden van verwijdering, laat het bedrijf in een reactie aan Yiannopoulos enkel weten dat de ban het gevolg is van het participeren in, of aanzetten tot “targeted abuse of individuals” (Ohlheiser, 2016).

C) *Aan welke (nieuws)media refereren Gab-gebruikers?*

Om de deelvragen te kunnen beantwoorden is er gebruik gemaakt van kwantitatieve tekstanalyse waarvan de resultaten kwalitatief geïnterpreteerd zijn. De drie deelvragen dienen om constructief met elkaar de haatdragende interactiepatronen op Gab te identificeren, beschrijven en contextualiseren. Belangrijk om op te merken is dat de deelvragen niet los van elkaar staan en in de analyse dus niet los van elkaar behandeld zullen worden. Zo hangen de ideologische overtuigingen van sprekers sterk samen met de doelwitten van hate speech en de nieuwsmedia waaraan zij refereren. De volgende sectie betreft een theoretisch kader waarin eerst hate speech gedefinieerd zal worden, gevolgd door een uiteenzetting van literatuur omtrent online communities zoals Gab. Tevens zullen kort de resultaten van eerder onderzoek naar Gab toegelicht worden.

HATE SPEECH

Hoewel er extensief onderzoek gedaan is naar de oorzaken, gevaren en verspreiding van hate speech, zijn er relatief weinig studies die pogen de term hate speech systematisch te definiëren (Sellars, 2016, p. 4). Volgens Faris, Ashar, Gasser & Joo (2016) wordt hate speech over het algemeen gekenmerkt als “speech which demeans or attacks a person or people as members of a group with shared characteristics such as race, gender, religion, sexual orientation, or disability” (p. 5). Het stereotyperen van een individu op basis van vaststaande persoonskenmerken is hiermee inherent aan hate speech (Cohen-Almagor, 2011). Echter variëren verschillende definities volgens Siegel (2018) in hoe specifiek ze zijn. De meest algemene definities omvatten een grote verscheidenheid aan uitingen gericht aan een individu of groep op basis van arbitraire of irrelevante persoonskenmerken (Parekh, 2006). Daarentegen kenmerken de meest specifieke definities uitingen enkel als hate speech wanneer ze expliciet aansporen tot geweld en fysieke schade toebrengen aan een groep (Benesch, 2012; Siegel, 2018). Deze meest expliciete uitingen noemt Benesch (2012) *dangerous speech*.

Belangrijk om op te merken is dat het doel van dit onderzoek niet is om Gab-posts te classificeren als ‘hate speech’ of ‘non-hate speech’. In plaats daarvan tracht dit onderzoek de gebruikerspraktijken op het platform bloot te leggen, om vervolgens te analyseren op welke manier er sprake is- of kan zijn- van hate speech, wie de doelwitten zijn en wat de ideologische motieven van de sprekers zijn.

FAR-RIGHT ONLINE SUBCULTUREN

Hoewel Gab-oprichter Torba stelt dat het platform geen exclusieve ontmoetingsplaats is voor rechts-conservatieven, blijkt uit eerder onderzoek (Lima et al., 2018; Zannettou et al., 2018) dat het discours op Gab gedomineerd wordt door rechts-conservatieven zoals Milo Yiannopoulos en andere ‘far-right’ online subculturen, soms samengevat als ‘alt-right’. Marwick & Lewis (2017, p. 3) specificeren dit als een verzameling van complotdenkers, techno-libertariërs, wit-nationalisten, verdedigers van mannenrechten, trollen, antifeministen en anti-immigratie-activisten. Hoewel deze subculturen divers zijn, laten zij zich over het algemeen kenmerken als anti-establishment in hun opvattingen over multiculturalisme en globalisme. Bovendien propageren zij vaak racistische, antifeministische en antisemitische ideologieën, die online tot uiting komen in de vorm van hate speech op platformen als 4chan en Gab. Deze praktijken hangen sterk samen met een extreme toewijding voor free speech en een sterk afkeurende houding ten aanzien van politieke correctheid (Marwick & Lewis, 2017).

ONLINE ECHO CHAMBERS

Lima et al. (2018) merken op dat de ideologische overtuigingen van Gab-gebruikers overwegend homogeen zijn. Daarom kenmerken zij Gab als een ‘right-leaning echo chamber’, waar lezers zelden blootgesteld worden aan content die ideologische grenzen doorbreekt, maar veel vaker gevoed worden met content die de politieke- en sociale overtuigingen van gebruikers bevestigen. Deze ideologische homogeniteit zorgt ervoor dat bestaande opvattingen elkaar versterken op een platform, waardoor er een community van gelijkgestemden ontstaat (Mihailidis & Viotty, 2017). Deze echo chambers kunnen ook binnen een platform ontstaan, bijvoorbeeld in de vorm van sterk gepolariseerde retweet-netwerken op Twitter. Zo laten Conover et al. (2011) zien dat er op Twitter retweet-netwerken met een sterk partijdige structuur ontstaan, met opvallend weinig connectie tussen links- en rechtsgeoriënteerde gebruikers. Dit is in overeenstemming met Garimella et al. (2018), die concluderen dat Twittergebruikers op grote schaal blootgesteld worden aan politieke opvattingen die overeenstemmen met hun eigen opvattingen.

Hoewel ideologische groeperingen in deze echo chambers dus weinig interactie met elkaar hebben, laten diverse studies een verschil in gedrag zien tussen conservatieven en liberalen. Opgemerkt moet worden dat liberaal hier niet verwijst naar de overkoepelende ideologische stroming met als uitgangspunt individuele vrijheid en autonomie, maar naar de progressief-liberale substroming hiervan die bestaat uit een mix van liberale en sociaaldemocratische uitgangspunten. Op deze manier is *liberal* in de Amerikaanse politiek vergelijkbaar met *links* in de Nederlandse politiek. Wanneer er in dit onderzoek gesproken wordt over liberaal, wordt er dus gerefereerd aan het progressief-liberalisme, oftewel het ‘politiek linkse’ in Amerika.

Adamic & Glance (2005) vonden verschillen in het gedrag van conservatieve en liberale blogs rondom de Amerikaanse presidentsverkiezingen in 2004, waarbij conservatieven onderling vaker naar elkaar verwezen dan liberalen. Aanvullend hierop concluderen Barberá et al. (2015) dat liberalen meer geneigd zijn zich te mengen in ‘cross-ideologische’ debatten op Twitter dan conservatieven. Mogelijk zijn conservatieve groeperingen online daarom gevoeliger voor het echo chamber-effect, aangezien zij minder geneigd zijn om deel te nemen aan liberale debatten (Barberá et al., 2015) en daarnaast onderling nauwer verbonden zijn (Adamic & Glance, 2005).

In contrast met de gesuggereerde ideologische homogeniteit binnen retweet-netwerken op Twitter, stellen Wieringa et al. (2018) dat de nieuwsmedia waaraan gebruikers binnen deze netwerken refereren overlappen binnen de Nederlandse Twittersferen. Gebruikers in zowel links- als rechtsgeoriënteerde clusters refereren voornamelijk aan mainstream nieuwsmedia, terwijl er aan alternatieve media een stuk minder wordt gerefereerd. Volgens Wieringa et al. (2018) selecteren Twittergebruikers bewust content die aansluit bij hun opvattingen en standpunten, en is de bron minder belangrijk bij dit proces. In plaats van een techno-deterministische filter bubble waarbij de nieuwsmedia bepaalde deelpublieken niet zouden bereiken, selecteert het (deel)publiek de content die het beste bij

hen aansluit en laten zij de rest buiten beschouwing. Hierdoor ontstaat er volgens Wieringa et al. (2018) een *willful echo chamber*.

Het is echter belangrijk om op te merken dat het onderzoek van Wieringa et al. (2018) zich beperkt tot de Nederlandse Twittersfeer. De conclusies zijn daarom niet generaliseerbaar voor andere landen of platformen. Dit wordt bevestigd door de resultaten van dit onderzoek, die laten zien dat op het gehele platform Gab een hoge mate van ideologische homogeniteit heerst. Bovendien wordt er op Gab zeer beperkt gerefereerd aan mainstream nieuwsmedia, terwijl er zeer extensief gerefereerd wordt aan alternatieve nieuwsmedia. De praktijken zoals die zich voordoen op de Nederlandse Twittersfeer lijken daarom eerder het gevolg van de Nederlandse mediapluriformiteit en/of platformspecifieke eigenschappen van Twitter, in plaats van algemene generaliseerbare gebruikerspraktijken op sociale media.

GERELATEERD ONDERZOEK

De resultaten van een demografische analyse van Gab-gebruikers door Lima et al. (2018) zijn weergegeven in tabel 1. Deze wijst uit dat meer dan de helft van de gebruikers een witte man is. Ter vergelijking: op Facebook is dit nog geen 30 procent (Lima et al., 2018). De dominantie van de witte man op het platform is in lijn met de stelling dat het discours op platforms zoals Gab sterk reageert op gebeurtenissen rondom wit-nationalisme (Zannettou et al., 2018), maar ook met Marwick & Lewis (2017), die far-right subculturen op Gab en 4chan's /pol/ samenvatten als een verzameling van onder andere wit-nationalisten, verdedigers van mannenrechten, antifeministen en anti-immigratie-activisten. Aanvullend hierop wijst een semantische analyse door Zhou, Dredze, Broniatowski & Adler (2018) uit dat het grootste deel van de gespreksonderwerpen op Gab betrekking heeft op politiek. Zie tabel 2 voor een overzicht van de geïdentificeerde topics en hun dominantie (Zhou et al., 2018).

Race	Gender		Total
	Male	Female	
Asian	3, 676 (10.4%)	1, 920 (5.4%)	5, 596 (15.8%)
Black	2, 106 (5.9%)	787 (2.2%)	2, 893 (8.2%)
White	18, 078 (50.9%)	8, 926 (25.1%)	27, 004 (76.1%)
Total	23, 860 (67.2%)	11, 633 (32.7%)	35, 493 (100%)

Tabel 1 Demografische verdeling van bijna 36 duizend Gab-gebruikers (Lima et al., 2018).

Politics Related		Other	
Ideology, religion and race	10.23%	Men and women	6.26%
Trump, Clinton and conspiracies	5.10%	Social media	4.83%
Right and left	5.08%	Profanity	4.63%
Miscellaneous politics	4.62%	Gab	3.29%
2016 Election and contemporary debates	3.42%	Pop culture	2.52%
Muslims and Europe	2.79%	Food	2.24%
MAGA	2.76%	Christmas and New Year	2.24%
Wars and international politics	2.39%	Literature and photos	2.02%
Las Vegas shooting and terrorism	2.29%	Economics	1.93%
Taxes and government spending	2.21%	Education and children	1.89%
Sex scandals	2.18%	Sports	1.65%
Guns	2.13%	Language	1.25%
Immigration	2.11%	German conversation	1.23%
State level issues	2.10%	Technology	1.13%
Globalism	1.95%	Health and nutrition	1.11%
Climate change	1.81%		
Legal	1.94%		
The media	1.25%		

Tabel 2 Geïdentificeerde topics in Gab-posts met bijbehorende dominantie, verdeeld in twee groepen: politiek en overig (Zhou et al., 2018).

Een topic-analyse zoals die van Zhou et al. (2018) brengt echter problemen met zich mee. Voor deze analyse werd een *latent dirichlet allocation* (LDA)-model getraind. Dit model schrijft geen eenduidig topic toe aan een document (Gab-post), maar representeert het document als een kansverdeling van verschillende topics. Deze verschillende topics bestaan op hun beurt weer uit kansverdelingen van alle woorden in het corpus. Wanneer een woord zeer specifiek betrekking heeft op een bepaald topic, zal dit woord dus een relatief hoge kans binnen de kansverdeling van de woorden in dit topic hebben. De manier hoe Zhou et al. (2018) dit model gebruiken voor analyse is echter problematisch omdat zij een document (Gab-post) toewijzen aan het topic met de hoogste kans binnen de kansverdeling van topics in dat document, terwijl dat document mogelijk bestaat uit meerdere onderliggende topics. Bovendien werden er handmatig labels aan topics toegewezen door te kijken naar de kansverdelingen van de

woorden voor de topics. Dit maakt de classificatie subjectief en onzeker, mede doordat aan 17 van de 50 topics geen betekenisvol label toegewezen kon worden. De keuzes die hier zijn gemaakt worden bovendien niet gerapporteerd. Zo zijn de kansverdelingen van woorden per topic- op basis waarvan labels aan topics toegekend werden- niet weergegeven en geven zij geen voorbeelden van posts die toegewezen zijn aan een bepaald topic om het functioneren van de methode te bewijzen.

Met betrekking tot hate speech zijn er weinig studies die systematisch inzichtelijk maken wie de doelwitten van hate speech zijn en welke ideologische posities sprekers innemen. Wat betreft de detectie van hate speech doen Zannettou et al. (2018) een poging door met een dictionary-based methode vast te stellen dat 5,4 procent van de Gab-posts ‘hate words’ bevatten. Hiervoor maken zij gebruik van Hatebase, een database met haatdragende woorden.² Het probleem van dictionary-based methoden als deze is echter dat hate speech niet gedetecteerd kan worden op basis van woordgebruik. Het is namelijk mogelijk dat een haatdragende uiting bestaat uit louter woorden die afzonderlijk van elkaar niet haatdragend zijn en daarom niet in de database voorkomen. Bovendien is de context waarin een woord wordt gebruikt van belang. Wanneer een uiting bijvoorbeeld het gebruik van een haatdragend woord afkeurt hoeft er geen sprake te zijn van hate speech, maar bevat de uiting wel het haatdragende woord.

In tegenstelling tot Zannettou et al. (2018), pogen Lima et al. (2018) hate speech op Gab te detecteren aan de hand van een model-based benadering. Hiervoor maken zij gebruik van Perspective. Dit is een initiatief van Google dat getrainde modellen via een API ter beschikking stelt om de *toxicity score* van posts te bepalen, waarbij toxicity gedefinieerd wordt als *rude, disrespectful, or unreasonable comment that is likely to make people leave a discussion*” (Perspective, 2018). Belangrijk om op te merken is dus dat toxicity niet gelijk is aan haatdragendheid. Volgens de definitie van Perspective is toxicity een algemenere maat voor ‘grof’ taalgebruik waarbij er niet inherent sprake hoeft te zijn van stereotypering, zoals dit bij hate speech wel het geval is. Dit maakt hate speech een subcategorie van toxic speech.

Lima et al. (2018) gebruiken dit Perspective-model enkel om een indicatie te verschaffen van de mate waarin hate speech zich op Gab voordoet. Zo merken zij op dat 8,6 procent van de posts op Gab beoordeeld wordt met een toxicity score van hoger dan 0.7, en daarom als ‘toxic speech’ gekenmerkt wordt. Deze methode is echter problematisch in onderzoek naar hate speech omdat een hoge toxicity score niet per se hoeft te duiden op hate speech. Daarnaast wordt de output niet kwalitatief geëvalueerd en geïnterpreteerd, waardoor de analyse volledig afhankelijk is van een model waarvan onbekend is hoe het bepaalde keuzes maakt. Bovendien maakt het ook niet structureel inzichtelijk wie de doelwitten van hate speech op Gab zijn en welke ideologische posities de sprekers motiveren. Dit onderzoek focust zich daarom niet alleen op het detecteren van hate speech op Gab, maar ook op het contextualiseren en interpreteren ervan. Hiervoor wordt er gebruik gemaakt van een methode waarbij de Perspective API als hulpmiddel dient. Deze zal in de volgende sectie worden toegelicht.

² <https://hatebase.org/>

METHODEN

Om haatdragende interactiepatronen op Gab bloot te leggen is er gebruik gemaakt van kwantitatieve tekstanalyse waarvan de uitkomsten kwalitatief geïnterpreteerd zijn. Belangrijk om op te merken is dat het doel van dit onderzoek niet is om Gab-posts te classificeren als ‘hate speech’ of ‘non hate speech’. In plaats daarvan tracht dit onderzoek de gebruikerspraktijken op het platform bloot te leggen, om vervolgens te analyseren in hoeverre en op welke manier er mogelijk sprake is- of kan zijn- van hate speech.

Om een eerste indicatie te verschaffen van het discours op Gab is er een exploratieve analyse uitgevoerd die zich in eerste instantie niet specifiek richt op het detecteren van hate speech. De eerste, meest algemene en exploratieve analyse betreft een analyse van de term frequenties.³ Hierbij wordt er gekeken wat de meest voorkomende woorden in Gab-posts zijn. Hetzelfde is gedaan voor de n-grams.⁴ Stopwoorden zijn gefilterd met behulp van een Engelse stopwoordenlijst die geïmplementeerd is in Scikit-learn.⁵

Vervolgens is er gekeken welke hashtags het meest gebruikt worden in Gab-posts. Dit is relevant aangezien hashtags vaak gebruikt worden om een onderwerp of standpunt van een bericht duidelijk te maken. Om deze reden is het mogelijk dat hashtags informatie blootleggen over de ideologische opvattingen of posities van de sprekers (deelvraag B), en daarmee mogelijk ook de doelwitten van eventueel hate speech (deelvraag A). Zo kan een spreker zich bijvoorbeeld met de hashtag #AltRight of #BanIslam met alt-right of anti-islam associëren. Hoewel dit niet direct hate speech hoeft te betreffen, zeggen dergelijke hashtags wel degelijk iets over de ideologische positionering van de sprekers en hun houding ten aanzien van bepaalde groepen.

Aanvullend hierop is er geanalyseerd aan welke (nieuws)media Gab-gebruikers refereren (deelvraag C). Hiervoor is er gekeken welke domeinen het meest voorkomen in URL's die op Gab gedeeld worden. Hoewel uit analyse van Wieringa et al. (2018) blijkt dat er in de Nederlandse Twittersfeer weinig verschillen zijn in de nieuwsmedia waaraan links- en rechtsgeoriënteerde gebruikers refereren, laten de referenties op Gab echter een zeer kenmerkende ideologische positionering zien. Merk hierbij op dat op deze manier deelvraag C bijdraagt aan de deelvragen A en B, en dat deze drie deelvragen- zoals eerder aangegeven- dus niet los van elkaar staan.

Na deze algemene exploratieve analyses is er een meer diepgaande analyse uitgevoerd die zich specifiek richt op het blootleggen van hate speech op Gab. Hiervoor is er gebruik gemaakt van de Perspective API, die ook door Lima et al. (2018) gebruikt wordt. Zoals eerder vermeld is Perspective een initiatief van Google dat getrainde machine learning-modellen via een API ter beschikking stelt om de *toxicity score* van posts te bepalen, waarbij toxicity gedefinieerd wordt als “rude, disrespectful, or

³ Term frequency: De frequentie waarmee een term of woord voorkomt in een corpus.

⁴ N-gram: Een sequentie van woorden in een lopende tekst, waarbij n staat voor de lengte van de sequentie. Hoogfrequente n-grams kunnen duiden op onderlinge afhankelijkheid van de woorden.

⁵ https://github.com/scikit-learn/scikit-learn/blob/master/sklearn/feature_extraction/stop_words.py

unreasonable comment that is likely to make people leave a discussion” (Perspective, 2018). Perspective zelf geeft op hun website aan dat de modellen het makkelijker en efficiënter kunnen maken om online conversaties te modereren. Toch is het belangrijk om op te merken dat de modellen fouten kunnen maken, en dat Perspective sterk adviseert om een menselijke beoordeling in het moderatieproces op te nemen.

De Perspective-modellen bieden echter meer mogelijkheden dan het modereren van online conversaties, bijvoorbeeld bij het blootleggen van hate speech. Het is hier zeer belangrijk om op te merken dat de modellen enkel als hulpmiddel kunnen dienen, en dat kwalitatieve interpretatie daarom te allen tijde leidend moet zijn bij het trekken van conclusies omtrent hate speech.

De Perspective API is bij dit onderzoek op verschillende manieren als hulpmiddel gebruikt. Ten eerste is het gebruikt om een subset van posts met een hoge toxiciteit te selecteren. Vervolgens zijn de term frequenties in posts met een hoge toxiciteit vergeleken met de term frequenties in het gehele corpus, om een beeld te krijgen van waarop toxische posts betrekking hebben.

Hoewel er op deze manier een indicatie wordt verkregen van de woorden die duiden op een hoge toxiciteit, wordt met deze methode nog niet de samenhang tussen woordgebruik en toxiciteit blootgelegd. Dit komt doordat er ook woorden zijn die vaak gebruikt worden in zowel het gehele corpus als in het toxische corpus. Om te analyseren welke woorden daadwerkelijk samenhangen met een hoge toxicity score is er gebruik gemaakt van een lineaire regressieanalyse. De coëfficiënten van de resulterende regressievergelijking zijn daarbij een maat voor de relatie tussen een woord en de toxiciteit van het discours waarin dat betreffende woord gebruikt wordt. Als een woord vaak voorkomt in een toxische context, dan zal dat woord een hogere coëfficiënt in de regressievergelijking krijgen. Subjecten van toxische uitingen- waaronder mogelijk hate speech- kunnen daarmee blootgelegd worden.

Om de resultaten uit de bovenstaande analyses te contextualiseren is de Perspective API gebruikt om toxische posts te selecteren die een specifiek woord bevatten. Deze posts zijn vervolgens kwalitatief geanalyseerd om het toxische discours rondom dat woord te contextualiseren. De woorden waarvoor dit is gedaan zijn bepaald aan de hand van de voorgaande analyses.

DATASET

De dataset die gebruikt is voor dit onderzoek is afkomstig van een database-dump⁶ van Gab. Omdat de originele dataset meer dan 25 gigabyte aan Gab-posts bevat is ervoor gekozen om een subset samen te stellen met alle posts vanaf 01-08-2018 tot en met de datum van de database-dump, 29-08-2018. De resulterende dataset bevat alle 2.276.804 posts die er in deze periode op Gab geplaatst werden. Door te kiezen voor deze manier van filteren heeft dit onderzoek betrekking op de huidige staat van het platform. Het doel van dit onderzoek is dan ook niet om te kijken naar hoe het discours op Gab in de loop der tijd zich zou hebben ontwikkeld.

⁶ Pushshift, 2018. <https://files.pushshift.io/misc/>

De originele dataset heeft een 'Newline delimited JSON'-structuur. Hierbij bevat iedere regel een JSON-object of waarde, in dit geval één post. Het voordeel hiervan ten opzichte van een normale JSON, is dat de gehele datastructuur niet ontleed hoeft te worden wanneer je maar een deel van de dataset nodig hebt. Hierdoor is het niet nodig om de originele dataset eerst volledig in het geheugen te laden voordat er gefilterd kan worden.

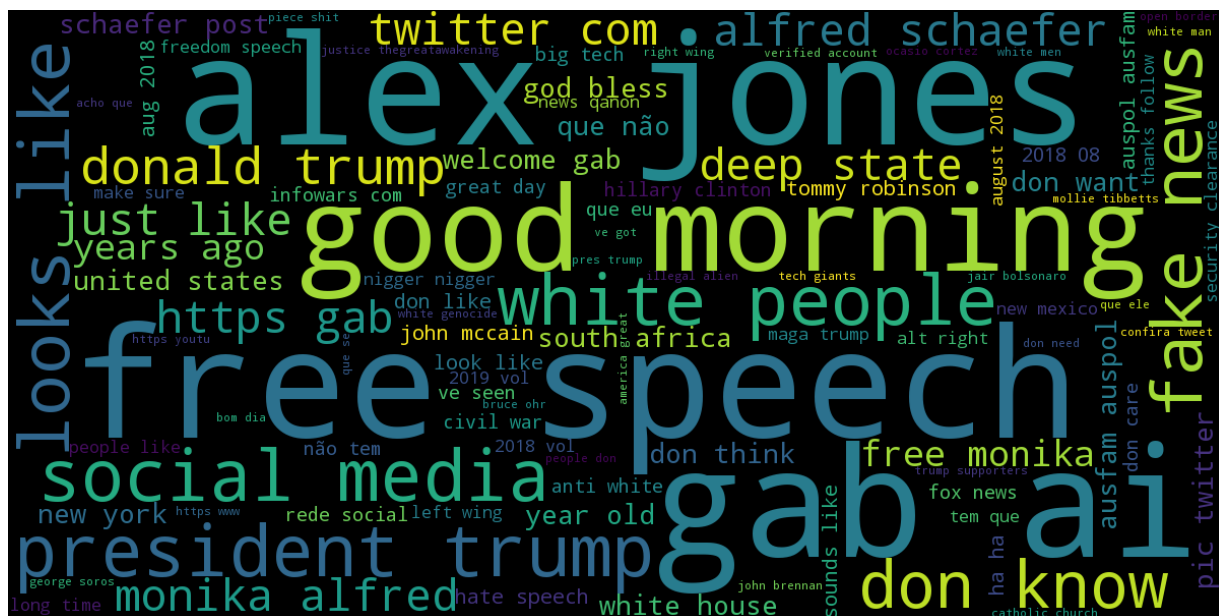
zouden zijn en als een ‘deep state’ samen zouden zweren tegen president Trump en zijn aanhangers (Martineau, 2017).

Ook lijkt het erop dat Gab-gebruikers sterk pleiten voor free speech en zich afzetten tegen censuur met de hashtags #speakfreely, #freespeech en #censorship. Dit bevestigt de stelling van Marwick & Lewis (2017) dat far-right online subculturen sterk pleiten voor free speech en een afkeurende houding aannemen ten aanzien van censuur en politieke correctheid. Tegelijkertijd wordt er verwezen met de hashtags #Infowars en #Altmedia naar alternatieve nieuwsmedia, waaronder Infowars. Hoe deze alternatieve media een rol spelen op Gab en hoe zij zich verhouden tot complottheorieën zal later in de analyse van de gedeelde nieuwsbronnen besproken worden. Gezien de hashtags die op het platform gebruikt worden is het aannemelijk dat er onder gebruikers aanhangers van complottheorieën zijn die stellen dat alternatieve media (ten onrechte) onder druk van de ‘deep state’ worden gecensureerd en beschuldigd van het verspreiden van *fake news*. Volgens hen bieden alternatieve media, zoals Infowars de échte waarheid die ons zou doen ontwaken (#TheGreatAwakening, #RedPill). #RedPill is hier een verwijzing naar de film *The Matrix* (1999) waarin het nemen van de rode pil een persoon zou doen ontwaken uit een- door een heersende macht- gecreëerde werkelijkheid waarna de echte werkelijkheid zichtbaar wordt.

Hoewel context van belang is om te bepalen of een uiting gekenmerkt kan worden als hate speech, schetsen hashtags als #BanIslam, #ExpelAllMuslims, #MuslimsAreTheEnemy en #BuildTheWall de afkeer van de Gab-community tegenover de Islam en buitenlanders. Overigens is bij #BanIslam, #ExpelAllMuslims en #MuslimsAreTheEnemy context niet nodig. Deze hashtags zijn inherent hate speech, omdat zij een groep (moslims) aanvallen op basis van hun religie of afkomst, iets wat volgens de definitie van Faris et al. (2016) hate speech is. Bovendien refereert de hashtag #WhiteGenocide aan de gelijknamige neo-Nazi, wit-nationalistische complottheorie die stelt dat niet-witte rassen zouden samenzweren om door middel van onder andere immigratie en integratie het witte ras te onderdrukken of zelfs op lange termijn te doen verdwijnen (Wilson, 2018).

N-GRAMS

Er is tevens gekeken naar de meest voorkomende n-grams in Gab-posts. Een overzicht van alle n-grams die 10 keer of meer voorkomen in het corpus is te vinden in 'NGRAMS.csv'.¹⁰ In de WordCloud in afbeelding 3 zijn de 100 meest frequente n-grams weergegeven. De bijbehorende frequenties zijn weergegeven in de tabel in bijlage 3.



Afbeelding 3: WordCloud van de 100 meest voorkomende n-grams op Gab.

Kenmerkend voor de praktijken op Gab en in overeenstemming met de hashtag-analyse is dat ‘Alex Jones’ en ‘free speech’ de meest voorkomende woordcombinaties op het platform zijn. Alex Jones is een bekend complottheoreticus en oprichter van Infowars: een far-right ‘nieuws’website voor complottheorieën. Infowars werd al meerdere malen beschuldigd van het verspreiden van fake news en hate speech (BBC, 2018). Ook werd Jones in februari 2018 beschuldigd van antisemitisme, racisme en seksuele intimidatie (Parry, 2018).

Tevens opvallend is de mate waarin er gesproken lijkt te worden over Monika en Alfred Schaefer: Een Duits-Canadees koppel dat werd veroordeeld voor opruiing middels het ontkennen en bagatelliseren van de holocaust (Schneider, 2018). Echter blijkt uit verdere analyse dat van de 5170 posts waarin er gerefereerd wordt aan de Schaefers, 5096 posts hetzelfde zijn en geplaatst werden door één gebruiker: ‘FREE Monika and Alfred Schaefer !!!!!!!!! Please POST this EVERYWHERE !!!!!!!!!’. Omdat deze gebruiker naar grote waarschijnlijkheid een bot is zullen posts van deze gebruiker uitgesloten worden van verdere analyse. Toch moet opgemerkt worden dat praktijken als deze bewust het publieke debat op Gab trachten te manipuleren. Ook dit is een kenmerk van far-right online subculturen (Marwick & Lewis, 2017).

¹⁰ <http://bit.ly/GabResearchOutput>

DOMEINEN VAN GEDEELDE URL'S

Met betrekking tot deelvraag C is er nagegaan aan welke (nieuws)media Gab-gebruikers refereren. Om deze deelvraag te kunnen beantwoorden is er gekeken welke domeinen het meest voorkomen in URL's die op Gab gedeeld worden. Een volledig overzicht van de domeinen en het aantal keer dat zij gedeeld werden is te vinden in 'SHARED_DOMAINS.csv'.¹¹ In de WordCloud hieronder in afbeelding 4 zijn de 100 meest gedeelde domeinen weergegeven. De bijbehorende frequenties zijn weergegeven in bijlage 4.



Afbeelding 4: WordCloud van de 100 meest gedeelde domeinen op Gab.

Uit analyse blijkt dat ruim 22 procent van de posts één of meerdere URL's bevatten. In bovenstaande WordCloud en in de tabel in bijlage 1 is te zien dat er op Gab vrijwel alleen maar wordt verwezen naar alternatieve media in plaats van mainstream media. Dit is erg kenmerkend voor de praktijken op Gab, aangezien uit de eerdere analyse reeds is gebleken dat Gab-gebruikers zich afzetten tegen mainstream media. Volgens hen bieden alternatieve media de waarheden die mainstream media verzwijgen. De meeste van deze nieuwsbronnen zijn echter niet onomstreden en kwamen regelmatig in opspraak wegens het verkondigen van complottheorieën en foutieve verhalen.

Opmerkelijk is dat deze bevindingen in groot contrast staan met de bevindingen in de studie door Wieringa et al. (2018). Hierin werd namelijk opgemerkt dat zowel links- als rechtsgeoriënteerde gebruikers in de Nederlandse Twittersfeer refereren aan- voor een deel overlappende- mainstream nieuwsmedia. Gab-gebruikers refereren daarentegen alleen maar aan alternatieve media. Waar de nieuwsbron in de Nederlandse Twittersfeer minder van belang is, zijn de nieuwsbronnen die op Gab gedeeld worden sterk ideologisch homogeen. Het lijkt er dus op dat mensen die kiezen voor alternatieve

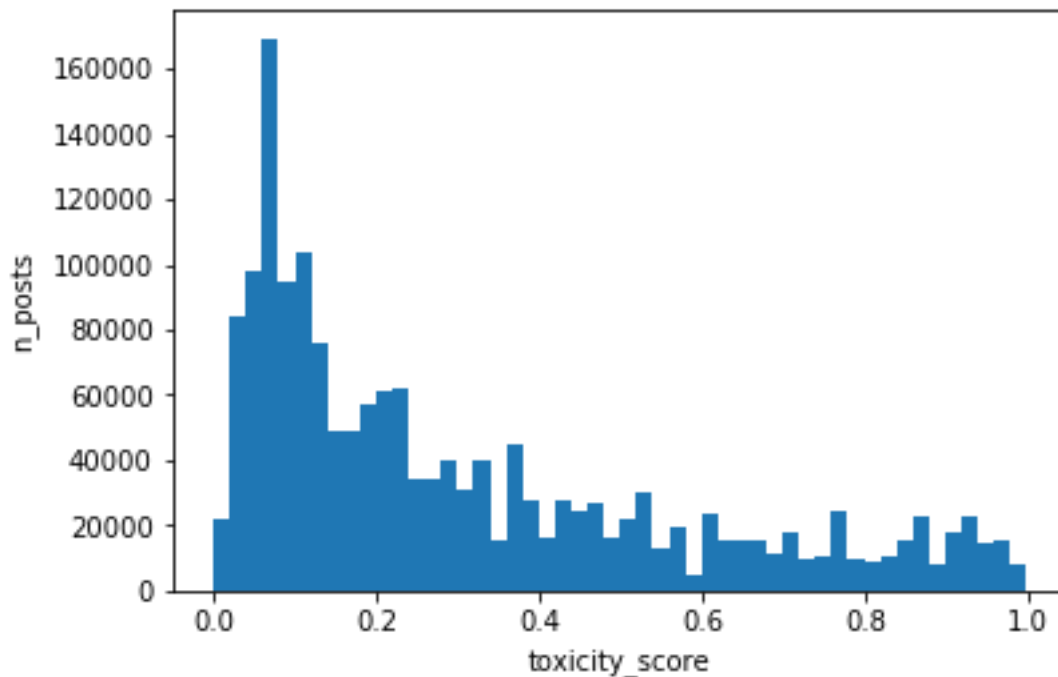
¹¹ <http://bit.ly/GabResearchOutput>

nieuwsmedia in plaats van mainstream nieuwsmedia tevens alternatieve sociale media (Gab) verkiezen boven mainstream sociale media (Twitter).

HATE SPEECH & TOXICITY

Zoals eerder vermeld is is het doel van dit onderzoek niet om Gab-posts te classificeren als ‘hate speech’ of ‘non-hate speech’. In plaats daarvan tracht dit onderzoek de gebruikerspraktijken op het platform bloot te leggen, om vervolgens te analyseren in hoeverre en op welke manier er mogelijk sprake is- of kan zijn- van hate speech. Een hulpmiddel hierbij is Google’s Perspective API, welke toegelicht is in de methodesectie.

Een overzicht van alle posts met de bijbehorende toxicity scores is te vinden in ‘TOXICITY.csv’.¹² In afbeelding 5 is een histogram weergegeven van de verdeling van de toxicity scores van de posts.



Afbeelding 5: Histogram van de toxicity scores van Gab-posts.

Afbeelding 5 laat zien dat hoewel het grootste deel van de posts een lage toxicity score heeft, een aanzienlijk deel toch hoog scoort. Zo blijkt uit de analyse dat de gemiddelde toxicity 0.31 bedraagt en de mediaan 0.21. Dit bevestigt de right-skewed verdeling die afbeelding 5 laat zien.

Om een eerste indruk te krijgen van wat de toxicity score inhoudt, zijn de posts gesorteerd op basis van toxicity score. In de tabel in bijlage 5 is een overzicht weergegeven van de 50 posts met de hoogste toxicity score. Hierin is te zien dat de post met de hoogste toxicity score luidt:

¹² <http://bit.ly/GabResearchOutput>

“Fuck all filthy Muslims! Between the Muslims and filthy abnormal homosexuals taking over society, we literally have a war on our hands!”

Hoewel definities van hate speech uiteenlopen, mag het duidelijk zijn dat bovenstaande post- ongeacht welke definitie gehanteerd wordt- een geval van hate speech betreft. Volgens de definitie van Faris et al. (2016), die hate speech kenmerken als “speech which demeans or attacks a person or people as members of a group with shared characteristics such as race, gender, religion, sexual orientation, or disability” (p. 5), zijn zowel moslims als homoseksuelen hier het slachtoffer van hate speech.

Toch is het belangrijk om nogmaals op te merken dat een hoge toxicity score niet per definitie duidt op hate speech. Volgens de eerder besproken definitie omvat toxicity een breder scala aan grove, respectloze of onredelijke uitingen. Zo kan zonder aanvullende context niet gezegd worden of de uiting *“Fuck off you STUPID fucktard!!!!”* een geval van hate speech betreft aangezien het niet duidelijk is of de aanval gericht is op een individu of groep op basis van arbitraire of irrelevante persoonskenmerken- iets wat ondanks de uiteenlopende definities inherent lijkt aan hate speech (Cohen-Almagor, 2011).

TERM FREQUENCIES IN TOXISCHE POSTS

Om na te gaan waarop posts met een hoge toxiciteit veelal betrekking hebben, zijn de term frequenties geanalyseerd van posts met een toxicity score van hoger dan 0.8. Door deze term frequenties te vergelijken met de term frequenties in het gehele corpus kan een indicatie worden verkregen van de woorden die duiden op een hoge toxicity. In afbeelding 6 is een WordCloud weergegeven van de 200 woorden met de hoogste frequentie in het gefilterde corpus.



Afbeelding 6: WordCloud van de 200 meest gebruikte woorden in posts met toxicity score > 0.8.

Opgemerkt moet worden is dat niet alle woorden in afbeelding 6 op zichzelf toxisch zijn. Woorden als bijvoorbeeld *jew*, *muslim*, *liberal* en *women* zijn op zichzelf geen toxische woorden. Het gegeven dat deze woorden zo extensief voorkomen in een corpus met een hoge toxiciteit duidt erop dat deze woorden vaak in een toxische context worden gebruikt. Om deze reden pleit dit onderzoek ervoor om onderscheid te maken tussen *intrinsieke toxiciteit* en *extrinsieke toxiciteit*. Intrinsieke toxiciteit refereert aan de mate waarin de **innerlijke semantiek** van een woord toxisch is. Zo worden woorden als *stupid*, *idiot*, *fucking* en *bastards* door de Perspective API beoordeeld met een toxicity score van respectievelijk .95, .97, .98 en .96. Merk hierbij op dat bij woorden met een hoge intrinsieke toxiciteit de afkeuring, belediging of aanval in de semantiek van het woord zelf zit. Hierdoor is de intrinsieke toxiciteit niet afhankelijk van de context waarin het woord gebruikt wordt, en dus niet corpuspecifiek.

In tegenstelling tot intrinsieke toxiciteit, refereert *extrinsieke toxiciteit* aan de mate waarin de **context** waarin een woord wordt gebruikt als toxisch wordt gekenmerkt. Zo worden woorden als *jew*, *muslim*, *liberal* en *women* beoordeeld met een toxicity score van respectievelijk .32, .21, .04 en .07. Gegeven de gemiddelde toxicity score in het gehele corpus van .31 kan daarom worden gesteld dat deze woorden op zichzelf niet toxisch zijn. Wanneer het dus blijkt dat deze woorden relatief vaker voorkomen in een corpus met een hoge toxiciteit, is het de context waarin deze woorden gebruikt worden hetgeen dat deze hoge toxiciteit veroorzaakt. Hierdoor bestaat de mogelijkheid dat dat betreffende woord het subject is van de toxische uiting, en dus het doelwit of slachtoffer is van hate speech. Het probleem is echter dat extrinsieke toxiciteit niet te bepalen is door term frequenties te analyseren in een corpus met een hoge toxiciteit, zoals in afbeelding 6 gedaan is. Woorden als *people*, *really* en *just* zijn namelijk woorden die over het algemeen vaak gebruikt worden en daarom niet per se zorgen voor de hoge toxicity score van de post.

extrinsieke toxiciteit mogelijk betrekking op de subjecten of doelwitten van hate speech. Het feit dat woorden als *jews*, *jewish*, *muslims*, *black* en *islam* allen een hoge extrinsieke toxiciteit hebben, laat zien dat dit mogelijke doelwitten van hate speech op Gab zijn (deelvraag A). Bovendien draagt deze analyse bij aan deelvraag B, die betrekking heeft op de ideologische positionering van Gab-gebruikers. Zo hebben *liberal*, *obama*, *democrat*, *hillary* en *left* allen een hoge extrinsieke toxiciteit, terwijl *conservative* en *right* juist een lage extrinsieke toxiciteit hebben. Het discours op Gab lijkt daarmee gedomineerd te worden door gebruikers met rechts-conservatieve, anti-links-liberale ideologische posities.

Om de bevindingen rondom de doelwitten van hate speech en de ideologische posities van Gab-gebruikers te contextualiseren worden er in de volgende sectie een aantal posts uit de data uitgelicht. Waar de analyse in de vorige secties de interactiepatronen van gebruikers op Gab in grote lijnen heeft beschreven, dient de volgende sectie om een beeld te krijgen van hoe deze patronen zich uiteindelijk manifesteren op het platform.

ANTISEMITISME

Analyse van de regressiecoëfficiënten en toxicity scores wijst tevens uit dat op er op Gab op grote schaal antisemitische uitingen worden verspreid. Zo is de gemiddelde toxicity score van posts met het woord *jew* erin .59, tegenover een totaal gemiddelde van .31. Bovendien bewijst de term frequency-analyse dat dit op grote schaal gebeurt, en niet slechts in een klein deel van de posts. Zo is in bijlage 1 te zien dat het woord *jews* één van de meest gebruikte woorden in het gehele corpus is. Het werd zelfs frequenter gebruikt dan bijvoorbeeld de woorden *America* en *president*. Tevens hebben de woorden *jew* en *jews* een sterk positieve regressiecoëfficiënt (resp. .091 en .090) en daarmee een hoge extrinsieke toxiciteit. Ter illustratie zijn in tabel 3 een aantal zeer haatdragende, antisemitische posts weergegeven.

Text	Toxicity Score
I fucking know about their fucking noses. Fuck off with you fake "history" knowledge, you jewish scum!!!!	0.995675
ALL jews go to hell, you false prophet faggot. kill yourself. i hope you die of cancer.	0.994948
fuck the jews. KILL'EM ALL	0.994689

Tabel 3: Antisemitische posts met hoogste toxicity score.

In tabel 3 is te zien dat antisemitisme zeer extreme vormen aanneemt, waarbij er zelfs expliciet opgeroepen wordt tot actie of aanval. Dergelijke uitingen worden door Benesch (2012) als *dangerous speech* gekenmerkt. Deze uitingen zijn niet alleen gevaarlijk omdat ze expliciet oproepen tot geweld, maar ook omdat het publiek op Gab ontvankelijk is voor dergelijke uitingen omdat zij veelal dezelfde ideologische opvattingen hebben. Hierdoor kunnen de grieven en angsten ten aanzien van joden die heersen onder het publiek door de spreker gemakkelijk gecultiveerd worden (Benesch, 2012).

RACISME

Omdat zowel de woorden *black* als *white* beiden positief samenhangen met toxicity score is er in de data gekeken naar de posts die één van deze twee woorden bevatten. Zoals verwacht blijkt dat in de toxische posts met *black* erin, zwarten veelal het doelwit zijn van een haatdragende en racistische uiting. Ter illustratie zijn in tabel 4 de drie posts weergegeven met de hoogste toxicity score, waarbij gefilterd is op het woord *black* in de post.

Text	Toxicity Score
Miserable black bitches	0.993910
black tide stinks like shit.. black shit..like you stupid nigger! LOL	0.992345
FUCK YOU AND YOUR DAD TO. YOU ARE BOTH BLACK PIECES OF SHIT.	0.991714

Tabel 4: Top 3 meest toxische posts met het woord 'black'.

Ter vergelijking zijn in tabel 5 de drie meest toxische posts met *white* erin weergegeven.

Text	Toxicity Score
Fuck YOU KAFFIR! Why are you here? To intentionally destroy and disrupt a white site with your fucking nigger bullshit! Get AIDS and DIE!	0.997251
You're a fucking race traitor and a cuck! STFU you fucking anti-white loser! You deserve nothing but vomit on your stupid head!	0.993730
so fucking stelaing land from white people? stupid fucks	0.993058

Tabel 5 : Top 3 meest toxische posts met het woord 'white'.

Tabel 5 laat zien dat in posts met het woord *white*, blanken niet het doelwit van hate speech zijn, terwijl dit voor *black* wel het geval is. Opvallend is dat deze uitingen vaak betrekking hebben op- een door de spreker gesuggereerde- onderdrukking of bedreiging van blanken, door zwarten. Dit is in overeenstemming met resultaten uit eerdere analyses, waaronder de dominantie van alternatieve, rechtse media als Infowars en TheGatewayPundit, en complottheorieën als #WhiteGenocide.

Het is tevens interessant om op te merken dat de bovenste post in tabel 5 stelt dat Gab een 'white site' zou zijn, en dat iemand die dat zou willen ontworpen moet verdwijnen. Hetzelfde geldt voor de tweede post in tabel 5, waarin expliciet iemand het zwijgen opgelegd wordt omdat het een 'anti-white loser' zou zijn. Hoewel Gab-gebruikers dus pleiten voor free speech en een sterk afkeurende houding aannemen ten aanzien van censuur, vallen zij zelf mensen met andere ideologische opvattingen aan en leggen zij hen het zwijgen op. De interpretatie van free speech door de Gab-community is daardoor een

stuk minder ‘free’ dan zij zelf beweren. Free speech lijkt dus gebruikt te worden door sprekers om zich achter te verschuilen om zo hate speech te rechtvaardigen, en niet omdat zij daadwerkelijk van mening zijn dat iedereen op Gab moet kunnen zeggen wat hij of zij wil.

CONSERVATISME

Uit analyse van de regressiecoëfficiënten bleek dat de praktijken op Gab zich laten kenmerken door een positief discours rondom conservatisme (negatieve coëfficiënt) in vergeleken met een negatief discours rondom (progressief) liberalisme (positieve coëfficiënt). Om deze resultaten te contextualiseren zijn hieronder de posts met de hoogste toxicity score weergegeven, waarbij gefilterd is op respectievelijk de woorden ‘conservative’ en ‘liberal’.

Text	Toxicity Score
F*** OFF with your boomer shit. Conservative boomers care about this country, you little f***ing fairy gay faggot millennial.	0.990670
Hey Facebook Twitter GO FUCK YOURSELF!!!! We are conservatives and we will fight for our future	0.990670
Today: FUCK YOU ZUCKFUCK...FUCK YOU FACEBOOK... GET FUCKED ALL OF YOU SCUMBAG POS FACEBOOK EMPLOYEES FOR LETTING THIS SLIDE...#conservative #Censorship	0.990670

Tabel 6: Top 3 meest toxische posts met het woord 'conservative'.

Tekst	Toxicity Score
Fuck Islam Fuck Muslims Fuck Liberals and all the fucked up wacky people! So sick and tired of this out of hand bullshit!	0.993910
Stupid ugly liberal bitch.	0.992075
Hey the fucking #Jew fucks and neocons / #RINOS can't play dirty ass war games with out false flag operations from the media from fucking Israel from the fucking chicoms from whatever dirty ass bastards!! And the stupid liberal puppets eat the shit up, even the so called right wingers are #brain...	0.991548

Tabel 7: Top 3 meest toxische posts met het woord 'liberal'.

Identiek aan de vergelijking van tabel 4 en tabel 5, blijkt uit tabel 6 en tabel 7 dat in toxische posts met *conservative*, conservatieven niet het doelwit van de uiting zijn, terwijl dit bij *liberal* wel het geval is. In tabel 7 is te zien dat gebruikers zich haatdragend uitlaten ten aanzien van liberalen, moslims en joden. Ook is in de tweede en derde post in tabel 6 te zien dat de conservatieve spreker zich fel afzet tegenover Twitter en Facebook, vermoedelijk vanwege het censureren van rechts-conservatieve nieuwsmedia of

hate speech (#Censorship). Gab-gebruikers lijken van mening te zijn dat mainstream media en –social mediaplatformen structureel rechts-conservatieve ideeën censureren, net zoals Milo Yiannopoulos Twitter als een ‘no-go zone’ voor conservatieven bestempelde (Hunt, 2016).

CONCLUSIE & DISCUSSIE

Waar studies naar hate speech op Gab zich tot op heden beperkten tot het tellen van haatdragende woorden of het bepalen van een gemiddelde toxicity score, is met dit onderzoek getracht hate speech op Gab structureel en betekenisvol in kaart te brengen. De gehanteerde methodologie combineert diverse methoden die tezamen een betrouwbaar beeld schetsen van de praktijken op Gab. Met dit onderzoek is hate speech op Gab niet alleen gedetecteerd, maar tevens betekenisvol in kaart gebracht.

De hoofdvraag die in de inleiding van dit onderzoek gesteld werd luidt: “*Hoe kunnen haatdragende interactiepatronen van gebruikers op Gab gekenmerkt worden?*” Tijdens de analyse zijn drie deelvragen als uitgangspunt genomen die samen een complete karakterisering van de haatdragende interactiepatronen op Gab vormen. De drie deelvragen staan niet los van elkaar, maar dienen constructief om de haatdragende interactiepatronen te begrijpen en beschrijven. De beantwoording van hoofdvraag gebeurt dus niet aan de hand van een losstaand antwoord, maar gebeurt in plaats daarvan aan de hand van de constructieve deelvragen.

Zo is met deelvraag A duidelijk geworden dat hate speech op Gab veelal betrekking heeft op joden, moslims en zwarten. Uit de regressieanalyse blijkt dat het discours rondom woorden als *jews*, *muslims*, *Israël*, *islam* en *black* sterk toxisch is. Kwalitatieve analyse van de meest toxische posts met deze woorden bevestigen dat het in veel gevallen daadwerkelijk om hate speech gaat. Bovendien roepen de haatdragende uitingen in sommige gevallen expliciet op tot geweld of onderdrukking van de doelwitten, iets wat in bestaande studies naar hate speech op Gab (Lima et al., 2018; Zannettou et al., 2018; Zhou et al., 2018) nog niet specifiek werd blootgelegd.

De ideologische en politieke posities die ten grondslag liggen aan deze haatdragende praktijken (deelvraag B) zijn sterk wit-nationalistisch, rechts-conservatief en racistisch van aard. Zo blijkt uit analyse van de samenhang tussen woordgebruik en toxicity score dat discours met woorden als *conservative* en *right* veelal positief sentiment bevat, terwijl discours rond *liberal*, *Obama*, *democrat*, *Hillary* en *left* veelal negatief, afkeurend en aanvallend is. Mede hierom kan geconcludeerd worden dat er op Gab een hoge mate van ideologische homogeniteit heerst, aangezien er bij ideologische diversiteit geen significant verschil zou zijn gevonden in de samenhang van de eerdergenoemde woorden met toxicity score.

Een andere bevestiging van deze ideologische homogeniteit komt voort uit deelvraag C, waarbij blootgelegd is aan welke nieuwsmedia Gab-gebruikers refereren. Deze referenties worden zeer sterk gedomineerd door alternatieve, rechtse (alt-right) media, terwijl er nauwelijks gerefereerd wordt aan mainstream media, waartegen gebruikers zich bovendien fel afzetten. Zo blijkt uit de regressieanalyse dat het discours rondom *msm* (mainstream-media) en *CNN* sterk toxisch is door de positieve samenhang van de woorden met toxicity score. Tegelijkertijd hangen de alternatieve media *Infowars* en *RT* (Russia Today) juist negatief samen met toxicity score, wat duidt op een positieve houding ten aanzien van deze media.

Uit de analyse blijkt dat Gab-gebruikers sterk pleiten voor free speech en zich fel afzetten tegen censuur en politieke correctheid. Paradoxaal genoeg worden tegelijkertijd mensen met andere ideologische overtuigingen, zoals links-liberalen, het zwijgen opgelegd en aangevallen. Free speech lijkt hierdoor gebruikt te worden als rechtvaardiging om racistische en antisemitische opvattingen te verspreiden. De manier waarop Gab-gebruikers free speech interpreteren heeft daarom in werkelijkheid weinig te maken met free speech, maar eerder met rechtvaardiging van extremisme, racisme, antisemitisme- en dus hate speech.

Hoewel Gab zich dus profileert als een platform voor free speech en individuele vrijheid, manifesteert het zich als een echo chamber waar hate speech een recht is. In deze echo chamber worden de opvattingen van gebruikers bevestigd en versterkt met eenzijdige berichtgeving door alternatieve media en medegebruikers. Waar Gab aangeeft dat iedereen welkom is op het platform, blijkt dit in werkelijkheid dus een stuk minder het geval.

REFERENTIES

- Adamic, L. A., & Glance, N. (2005). The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. In *Proceedings of the 3rd International Workshop on Link Discovery* (pp. 36–43). New York, NY, USA: ACM. <https://doi.org/10.1145/1134271.1134277>
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science*, 26(10), 1531–1542. <https://doi.org/10.1177/0956797615594620>
- Cohen-Almagor, R. (2011). Fighting Hate and Bigotry on the Internet. *Policy & Internet*, 3(3), 89–114. <https://doi.org/10.2202/1944-2866.1059>
- Conover, M. D., Ratkiewicz, J., Francisco, M., Goncalves, B., Menczer, F., & Flammini, A. (2011). Political Polarization on Twitter. In *Fifth International AAAI Conference on Weblogs and Social Media*. Geraadpleegd van <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2847>
- Faris, R., Ashar, A., Gasser, U., & Joo, D. (2016). *Understanding Harmful Speech Online* (Onderzoekspublicatie No. 2016–21). Cambridge, MA: Berkman Klein Center for Internet & Society. Geraadpleegd van <http://www.ssrn.com/abstract=2882824>
- Gab | Community Guidelines. (z.d.). Geraadpleegd 25 september 2018, van <https://gab.ai/about/guidelines>
- Garimella, K., De Francisci Morales, G., Gionis, A., & Mathioudakis, M. (2018). Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship. In *Proceedings of the 2018 World Wide Web Conference* (pp. 913–922). Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee. <https://doi.org/10.1145/3178876.3186139>
- Hunt, E. (2016, juli 20). Milo Yiannopoulos, rightwing writer, permanently banned from Twitter. *The Guardian*. Geraadpleegd van <https://www.theguardian.com/technology/2016/jul/20/milo-yiannopoulos-nero-permanently-banned-twitter>
- Lima, L., Reis, J. C. S., Melo, P., Murai, F., Araújo, L., Vikatos, P., & Benevenuto, F. (2018). Inside the Right-Leaning Echo Chambers: Characterizing Gab, an Unmoderated Social System. In *Proceedings*

of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 515–522). Geraadpleegd van <http://arxiv.org/abs/1807.03688>

Marwick, A., & Lewis, R. (2017). *Media Manipulation and Disinformation Online* (p. 106). New York: Data & Society Research Institute.

Mihailidis, P., & Viotty, S. (2017). Spreadable Spectacle in Digital Culture: Civic Expression, Fake News, and the Role of Media Literacies in “Post-Fact” Society. *American Behavioral Scientist*, 61(4), 441–454. <https://doi.org/10.1177/0002764217701217>

Ohlheiser, A. (2016, juli 21). Just how offensive did Milo Yiannopoulos have to be to get banned from Twitter? *Washington Post*. Geraadpleegd van <https://www.washingtonpost.com/news/the-intersect/wp/2016/07/21/what-it-takes-to-get-banned-from-twitter/>

Parekh, B. (2006). Hate Speech. *Public Policy Research*, 12(4), 213–223. <https://doi.org/10.1111/j.1070-3535.2005.00405.x>

Parry, R. (2018, februari 28). Former Infowars employees claim Alex Jones harassed them. *Mail Online*. Geraadpleegd van <http://www.dailymail.co.uk/news/article-5419371/Former-Infowars-employees-claim-Alex-Jones-harassed-them.html>

Perspective API reference. (2018). Jigsaw/Google. Geraadpleegd van <https://github.com/conversationai/perspectiveapi> (Original work published 2017)

Price, R. (2017, augustus 18). Google’s app store has banned Gab — a social network popular with the far-right — for “hate speech”. Geraadpleegd 25 september 2018, van <http://uk.businessinsider.com/google-app-store-gab-ban-hate-speech-2017-8>

Schneider, A. (2018, mei 4). 62-Jähriger wegen Holocaust-Leugnens verurteilt. *Sächsische Zeitung*. Geraadpleegd van <https://www.saechsische.de/62-jaehriger-wegen-holocaust-leugnens-verurteilt-3931021.html>

Sellars, A. (2016). *Defining Hate Speech* (Onderzoekspublicatie No. 2016–20). Rochester, NY: Berkman Klein Center. Geraadpleegd van <https://papers.ssrn.com/abstract=2882244>

Siegel, A. A. (2019). Online Hate Speech. In *Social Media and Democracy: The State of the Field*. Stanford, CA. Geraadpleegd van https://alexandra-siegel.com/wp-content/uploads/2018/09/Siegel_Online_Hate_Speech.pdf

- VICE News. (2016). *This Is Twitter For The Internet's Biggest Racists (HBO)*. Geraadpleegd van <https://www.youtube.com/watch?v=QaMZ5mDIBTo>
- Wendling, M. (2016, december 14). Gab: Free speech haven or alt-right safe space? Geraadpleegd 5 oktober 2018, van <https://www.bbc.com/news/blogs-trending-38305402>
- Wieringa, M. A., van Geenen, D., Schaefer, M. T., & Gorzeman, L. (2018, mei 15). Political topic-communities and their framing practices in the Dutch Twittersphere. <https://doi.org/10.14763/2018.2.793>
- Wilson, J. (2016, november 17). Gab: alt-right's social media alternative attracts users banned from Twitter. *The Guardian*. Geraadpleegd van <https://www.theguardian.com/media/2016/nov/17/gab-alt-right-social-media-twitter>
- YouTube removes "hate speech" videos from InfoWars. (2018, juli 26). Geraadpleegd 18 februari 2019, van <https://www.bbc.com/news/technology-44965160>
- Zannettou, S., Bradlyn, B., De Cristofaro, E., Kwak, H., Sirivianos, M., Stringhini, G., & Blackburn, J. (2018). What is Gab? A Bastion of Free Speech or an Alt-Right Echo Chamber? In *WWW '18 Companion Proceedings of the The Web Conference 2018* (pp. 1007–1014). Lyon. <https://doi.org/10.1145/3184558.3191531>
- Zhou, Y., Dredze, M., Broniatowski, D. A., & Adler, W. D. (2018). *Gab: The Alt-Right Social Media Platform* (p. 6). Baltimore, MD: Center for Language and Speech Processing. Geraadpleegd van http://www.cs.jhu.edu/~mdredze/publications/2018_sbprims_gab.pdf

BIJLAGEN

BIJLAGE 1 - TOP 150 WOORDEN MET DE HOOGSTE FREQUENTIE OP GAB

Term	Frequency	Term	Frequency	Term	Frequency
like	109337	eu	26964	facebook	18345
just	108754	president	26890	money	18278
people	104866	ll	26250	hope	18204
trump	94906	does	26178	things	18147
que	90158	para	26090	long	18082
don	78025	shit	25733	read	17919
gab	77349	look	25580	evil	17882
good	67335	state	24767	por	17714
know	61478	maga	24301	true	17693
time	58534	war	24279	didn	17668
white	57925	die	24215	fucking	17608
right	52419	country	24076	oh	17502
com	51405	use	23428	children	17304
think	51303	video	23364	anti	17270
não	47022	rt	23298	women	17188
new	46327	thing	23185	help	17048
twitter	44594	jew	22889	year	17013
want	43750	speech	22808	tem	17003
news	41375	qanon	22790	mais	16964
free	38946	hate	21931	fake	16915
need	38765	yes	21827	law	16864
make	37971	uma	21448	obama	16842
did	37807	work	21300	follow	16807
jews	36393	watch	21265	bad	16746
going	36240	big	21229	muslim	16583
god	35287	government	21058	pra	16477
media	35131	black	21045	doing	16378
say	34438	fuck	20965	morning	16261
world	34363	believe	20897	men	16259
day	34116	life	20605	best	16226
2018	34068	american	20428	used	16225
way	33848	na	20310	aqui	16138
man	33732	better	20160	tell	16133
lol	32428	thanks	20146	support	16125
great	32372	youtube	20140	alex	16103
post	32337	thank	20044	actually	15827
left	31661	sure	20032	won	15783
da	31321	today	19976	jones	15751
um	30866	come	19931	vote	15715
really	29909	truth	19825	trying	15672
em	29712	bolsonaro	19791	democrats	15371
america	29548	welcome	19772	islam	15178
years	28640	old	19686	mas	15122
ve	28550	little	19437	isn	15010
love	27789	os	19344	jewish	14844
said	27487	live	19325	control	14838
let	27456	stop	18961	getting	14747
got	27275	social	18783	https	14693
real	27202	doesn	18737	away	14670
se	27095	says	18470	called	14491

BIJLAGE 2 – TOP 150 HASHTAGS MET DE HOOGSTE FREQUENTIE OP GAB

Hashtag	Freq	Hashtag	Freq	Hashtag	Freq
#qanon	13915	#linux	1056	#art	602
#maga	13781	#brexit	1022	#christians	596
#trump	5886	#winning	1020	#uk	586
#wwglwga	5430	#texasfirst	1002	#pizzagate	585
#news	5195	#afd	997	#cnn	583
#gabfam	5141	#buildthewall	996	#bolsonarojornalnaciona l	582
#ausfam	5019	#1776	949	#africansongab	575
#q	4962	#bolsonaro2018	944	#redwave	572
#thegreatawakening	4939	#spygate	903	#welcome	561
#auspol	4927	#msm	868	#qanuck	555
#walkaway	3775	#facebook	866	#russia	554
#politics	3426	#pnn	864	#google	550
#potus	3380	#antifa	851	#beloudbeheard	540
#kag	3192	#oathkeepers	845	#bitchute	537
#greatawakening	3161	#threepercenters	840	#bolsonaropresidente	536
#kenyansongab	2871	#bolsonarogab	837	#rn ccn	524
#direitaunida	2810	#liberty	832	#canada	522
#speakfreely	2548	#metoo	830	#southafrica	522
#usa	2465	#bansharia	823	#bolsonaropresidente17	518
#justice	2289	#1	799	#eu	516
#nsfw	2223	#meseguenogab	782	#freedomofspeech	514
#thestorm	2217	#sweden	763	#obama	512
#freespeech	2122	#bolsonaro	753	#jews	510
#australia	1916	#conservative	752	#america	502
#fakenews	1804	#americafirst	723	#us	502
#gab	1789	#whitegenocide	722	#libertarian	497
#deepstate	1773	#democrats	721	#gamergate	494
#censorship	1748	#gabbrasil	721	#tommyrobinson	494
#freedom	1685	#deutschertrump	715	#sharefreely	493
#2a	1587	#altright	711	#midterms	493
#direitaseguedireita	1574	#bitcoin	708	#republicans	485
#banislam	1572	#qarmy	695	#prolife	484
#infowars	1568	#patriots	688	#putin	471
#draintheswamp	1533	#gaming	686	#military	469
#redpill	1507	#foxnews	685	#keepamericagreat	465
#altmedia	1431	#muslimsaretheenemy	683	#fbi	463
#patriotsfight	1425	#gnu	682	#europe	463
#britfam	1423	#gop	670	#youtube	463
#hrcratline	1399	#theyfearthegreatawakeningt hemost	664	#crime	463
#bolsonaro17	1384	#china	653	#qanon8chan	455
#wethepeople	1379	#cia	651	#pjnet	454
#alexjones	1327	#muslims	642	#proudboys	450
#	1281	#word	639	#dnc	446
#tcot	1217	#banshariathe	630	#history	446
#twitter	1192	#superelite	627	#nrx	437
#trump2020	1152	#translations	624	#education	435
#nra	1148	#immigration	623	#makeearthgreatagain	434
#pedogate	1100	#expelallmuslims	611	#truth	433
#islam	1071	#israel	610	#merkel	430
#1a	1059	#muslim	608	#clinton	428

BIJLAGE 3 – TOP 100 N-GRAMS MET DE HOOGSTE FREQUENTIE OP GAB

N-gram	Freq	N-gram	Freq
alex jones	13505	fox news	2736
free speech	12630	que eu	2659
gab ai	10016	infowars com	2656
good morning	9716	ve seen	2620
president trump	9314	great day	2593
social media	9274	freedom speech	2572
white people	7654	maga trump	2506
looks like	7540	security clearance	2474
fake news	7224	2018 vol	2427
don know	7063	left wing	2422
donald trump	6797	tem que	2401
deep state	5890	2019 vol	2389
just like	5552	new mexico	2363
https gab	5267	nigger nigger	2355
monika alfred	5112	alt right	2328
alfred schaefer	5105	people like	2326
twitter com	5097	make sure	2325
free monika	5096	long time	2307
years ago	5094	aug 2018	2294
schaefer post	5090	news qanon	2267
don want	5055	thanks follow	2262
ausfam auspol	4729	justice thegreatawakening	2209
welcome gab	4649	john brennan	2169
south africa	4639	httpsyoutu	2164
year old	4586	ve got	2094
new york	4303	acho que	2083
united states	4254	don need	2082
god bless	4127	right wing	2070
que não	4047	trump supporters	2059
don think	3958	que se	2010
pic twitter	3928	httpswww	2000
white house	3784	tech giants	1998
tommy robinson	3755	que ele	1979
john mccain	3694	ocasio cortez	1973
don like	3507	illegal alien	1904
2018 08	3424	jair bolsonaro	1894
auspol ausfam	3207	pres trump	1877
look like	3206	white men	1874
hate speech	3077	george soros	1850
sounds like	3043	bom dia	1850
ha ha	3022	confira tweet	1846
hillary clinton	2988	piece shit	1843
big tech	2984	bruce ohr	1842
não tem	2982	mollie tibbetts	1826
2018 08	2961	people don	1812
anti white	2855	verified account	1812
civil war	2844	white genocide	1800
don care	2804	catholic church	1785
rede social	2785	open borders	1774
fox news	2736	america great	1769
que eu	2659	white man	1765

BIJLAGE 4 – TOP 50 MEEST GEDEELDE DOMEINEN OP GAB

Domein	Frequentie	Domein	Frequentie
Nan (Geen URL in post)	1776213	voiceofeurope.com	1910
kek.gg	51337	www.youtube.com	1883
twitter.com	29151	www.journalistenwatch.com	1821
www.breitbart.com	18732	www.naturalnews.com	1798
www.infowars.com	14605	www.americanthinker.com	1771
www.thegatewaypundit.com	12279	thehill.com	1718
t.co	10144	www.jihadwatch.org	1688
www.bitchute.com	8469	thedailycoin.org	1647
gab.ai	7425	www.twitlonger.com	1594
www.foxnews.com	7371	theconservativetreehouse.com	1576
bit.ly	7084	www.newswars.com	1571
www.zerohedge.com	6314	hooktube.com	1544
dailycaller.com	5438	pic.twitter.com	1496
i.redd.it	5412	pbs.twimg.com	1496
www.dailymail.co.uk	4967	www.blazingcatfur.ca	1488
www.facebook.com	4603	www.minds.com	1445
truepundit.com	3722	www.yahoo.com	1414
www.westernjournal.com	3139	redd.it	1368
www.rt.com	2996	help.gab.ai	1353
steemit.com	2636	thefederalist.com	1266
tinyurl.com	2341	www.reddit.com	1242
www.real.video	2233	pjmedia.com	1241
threadreaderapp.com	2066	en.wikipedia.org	1228
www.express.co.uk	1956	www.washingtonexaminer.com	1216
www.dailywire.com	1938	townhall.com	1197

BIJLAGE 5 – TOP 50 POSTS MET DE HOOGSTE TOXICITY SCORE

text	Toxicity_score
Fuck all filthy Muslims! Between the Muslims and filthy abnormal homosexuals taking over society, we literally have a war on our hands!	0,997263
Fuck YOU KAFFIR! Why are you here? To intentionally destroy and disrupt a white site with your fucking nigger bullshit! Get AIDS and DIE!	0,997251
You piece of shit PEDOPHILE MOTHERFUCKER!!!	0,997054
Fuck off you STUPID fucktard!!!!	0,996726
fucking muslim pigs!!	0,996718
Fuck you, you piece of shit.	0,996571
fuck off, you jewish piece of shit	0,996438
All fucking muslims should be shot fucking dead.	0,996131
Go fuck yourself you traitorous piece of shit.	0,996040
You're a fucking arrogant prick, who you think you are, to talk shit about my country. Go to hell!	0,996024
FUCK YOU YOU RACIST BASTARD	0,995939
fuck you, moslem-sucks !! eat dogshit !	0,995899
Fuck that nigger Communist. All shit-skins must die	0,995828
Fuck you you racist bastard.	0,995804
Fuck you, Erdogan. Å Eat shit Å	0,995797
Fuck off you stupid fucking libtard!!!	0,995752
You are a fucking dumb bitch!	0,995699
I fucking know about their fucking noses. Fuck off with you fake "history" knowledge, you jewish scum!!!!	0,995675
FUCKERS CUNTS SHIT ASSHOLES MUSLIMS PIECE OF SHIT	0,995674
adam you're a fucking moron and need to study history. sorry. fuck off you patronizing piece of shit	0,995665
Fucking niggers are DISGUSTING!	0,995657
YOU FUCKIN BITCH NIGGER FAGGOTS LIKE YOU END UP HOMELESS YOU FUCKING JEW ASS LICKER	0,995651
FUCKING NIGGERS and their Bible bullshit	0,995650
fuckin donkey ass feminist bitch	0,995628
You fucking niggers are stupid	0,995605
WHOâ€™S THE FUCKING NIGGER THAT DOWNVOTED THIS POST?????	0,995508
FUCK YOU, NIGGER	
Fuck that Muslim piece of shit.	0,995429
Youâ€™re a fuckin idiot.	0,995428
fuck that punk motherfucker.	0,995283
Fucking IDIOTS!!	0,995272
Fuck islam and all their goat fucking followers.	0,995255
fuck you motherfucker!!!!!!!!!!!!!!!!!!!!!!!!!!!!@charly caruso	0,995218
Your a fuckin idiot..	0,995119
go fuck yourself you fucking faggot	0,995119
You are a fucking idiot.	0,995047
listen you are fuckin garbage have a shitty steak bitch	0,995046

Here's an idea. Go fuck yourselves. Idiots.	0,995033
You're a fuckn GOMER PILE, bitch	0,995028
fuck you you muslim cocksucker!!!	0,995026
stupid fucking sub-human nigger piece of shit!	0,994982
ALL jews go to hell, you false prophet faggot. kill yourself. i hope you die of cancer	0,994948
Mark Zuckerberg is a cum dumpster fucking bastard	0,994920
fuck you lyin communist fuckin pigs, nigger lovers	0,994891
Fuck Islam, Fuck Muhammad and fuck all Raghead Muslim cocksucking Demons from Hell.	0,994880
fuck off you piece of shit	0,994879
Jesus you are a dumb fucking cunt. You may be the most wretchedly pathetic moron I've seen on this platform yet. You are PAINFULLY stupid. Fuck.	0,994867
fuck this motherfucker, muslim asshole..#metoo	0,994850
Fuck you idiot!	0,994847
Fuck that dirty motherfucker	0,994838
YOU ARE A FUCKING COCKSUCKER... MUTE...	0,994824