

COLLECTIE

Jeroen Fernandez-Alonso
archivaris, Amsab-ISG

waardering van hashtags

‘Waardering’ is een van de kerntaken van een archivaris: horen de bronnen thuis in onze collectie? Een belangrijk kenmerk van waarden is dat het een tijds- en plaatsgebonden proces is. Het is een momentopname van wat we op dat moment als waardevol beschouwen voor een gemeenschap.¹

FARO, het Vlaams steunpunt voor cultureel erfgoed, definieert waarden als: ‘Het proces van onderzoeken, begrijpen en vastleggen van verschillende erfgoedwaarden op een rationele en gestructureerde wijze [...]. Een waardering van cultureel erfgoed is een participatief proces en vormt de basis voor een duurzame cultureel-erfgoedwerking. Het wordt ingezet met als doel op een rationele manier te

herkennen en verzamelen, behouden en borgen, onderzoeken, presenteren en toeleiden.’²

Begin 2020 is Amsab-Instituut voor Sociale Geschiedenis gestart met het archiveren van hashtags, voornamelijk op Twitter, maar ook op Instagram. In eerste instantie was dat nog heel ad hoc, naar aanleiding van grootschalige activistische gebeurtenissen zoals #BlackLivesMatter en #FridaysForFuture. In die beginfase onderzochten we nog de technische mogelijkheden om met software op een geautomatiseerde wijze sociale media te downloaden (*scrapen*). Het belangrijkste criterium om tot archivering over te gaan was dat de hashtag vermeld werd in de traditionele media (kranten, radio en televisie).

Geleidelijk aan zijn we digitaal activisme systematischer gaan waarderen, maar we botsten op heel wat hindernissen bij het opstellen van een rationeel en gestructureerd waarderingskader voor de beoordeling van hashtags. In deze tekst bied ik een overzicht van typisch gebruikte criteria binnen waarderingskaders in archieven. Aan de hand van onze trial-and-erroraanpak toon ik aan waarom ze niet altijd even bruikbaar zijn voor hashtags.

'Slacktivisme' en de historische waarde

Sinds de Arabische Lente is er een grote toename van onderzoek naar digitaal activisme en meer bepaald naar de rol die socialemediaplatformen spelen bij sociopolitieke verandering.³ Het gebruik van hashtags werd een focuspunt binnen de academische literatuur over digitaal activisme en kreeg de term 'hashtagactivisme', weliswaar zonder een eenduidige definitie.⁴



Al gauw kwamen ook de termen 'slacktivism' of 'clicktivism' in gebruik. In eerste instantie verwees de term 'slacktivism' naar bottom-upinitiatieven van jongeren, met een beperkte impact op de samenleving.⁵ Al snel werd de term denigrerend gebruikt om een feelgood-, performatieve en vooral ineffektieve vorm van activisme te benoemen.⁶ Een typisch voorbeeld is de #Kony2012-campagne tegen guerrillaleider Joseph Kony en het gebruik van kindsoldaten. Die bewustwordingscampagne was succesvol in de westerse wereld, maar veel meer kan een online actie natuurlijk niet bereiken tegen een krijgsheer die actief is in Oeganda.⁷

Hashtagactivisme wordt ook vaak afgedaan als slacktivism en gebruikt als drogreden om niet naar de boodschap te hoeven luisteren. Toch kan slacktivism er juist voor zorgen dat een boodschap wijd verspreid wordt. De #MeToo-beweging is in 2017 pas echt viraal gegaan door een tweet van de Amerikaanse actrice Alyssa Milano, waarin ze opriep om die hashtag te gebruiken. Het succes van de beweging kwam er in grote mate door de activiteiten van miljoenen slacktivisten, zodat de aantijgingen niet meer genegeerd konden worden.

De academische waardering en de vraag of een hashtag al dan niet slacktivism is, mag geen invloed hebben op de archivalische beslissing om een hashtag op te nemen in een archief. Omwille van technische redenen kun-

nen we niet afwachten tot het academische debat tot een consensus heeft geleid over de rol van hashtags bij het beïnvloeden of veroorzaken van bepaalde sociopolitieke veranderingen. Idealiter beginnen we zelfs al met archiveren in de eerste zeven dagen na het ontstaan van de hashtag (cf. infra). Dat leidt tot een vervelende situatie: grote hoeveelheden data die we opnemen in ons digitaal archief, blijken later misschien waardeloos voor toekomstig gebruik of onderzoek. Hashtags realtime archiveren is een vorm van hedendaagse documentatie, waarvan we de historische waarde op het moment zelf nog niet kunnen inschatten.

In de praktijk richten veel academische studies over hashtags zich nog op de live online data, en niet op statische datasets in archieven. Dat vormt een probleem voor bronnenkritiek. De studies kunnen niet worden gereproduceerd, omdat data op het internet snel verdwijnen of worden aangepast. Deels komt dat doordat de meeste archieven nog geen datasets met gearhiveerde hashtags kunnen aanbieden. Deels doordat studies over hashtags uitgevoerd worden in een grote verscheidenheid aan academische domeinen, waarin velen zich niet bewust zijn van de rol die archieven kunnen spelen. In de praktijk zijn geschiedkundigen het voornaamste doelpubliek van archieven, waardoor de (kunst)historische waarde een gebruikelijk criterium in waarderingskaders is. Bij hashtags

is het echter lastig oordelen over dat criterium, omdat we een waardeoordeel moeten geven aan gebeurtenissen terwijl ze nog gaande zijn. Bijgevolg is ook het typische waarderingscriterium 'te verwachten vraag' moeilijk te beantwoorden. Van sommige collecties hashtags, zoals #BLM, krijgen we al vragen als ze nog bezig zijn, maar van kleinere hashtags is niet in te schatten of ze ooit opgevraagd zullen worden. Daarnaast moeten we ook ruimer kijken dan alleen naar de waarde voor historici. We moeten onze collecties van sociale media en hashtags ook kenbaar maken bij andere academische domeinen, zoals de mediastudies, de datawetenschappen, de digital humanities, de politieke en sociale wetenschappen.

Participatie voor volledigheid

Bij de waardering is het aanbevolen om een participatieve aanpak te gebruiken, zodat het oordeel niet afhangt van één persoon, maar van een brede groep belanghebbenden. In het geval van hashtags moeten we er zoveel mogelijk mensen bij betrekken. Er is zoveel gaande op het internet, en thema's worden plots actueel, maar de historische waarde van de online activiteit is op het moment zelf nog niet te schatten. Personen uit de desbetreffende gemeenschappen moeten meehelpen bij de waardering en ons zaken aanreiken waarvan zij vinden dat ze bewaard moeten worden. Daarnaast hebben we hun hulp nodig om-

wille van technische redenen: sommige sociale media kunnen we niet zomaar *scrapen*. De gemeenschappen moeten hun onlinedata in dat geval exporteren en aan ons bezorgen. Een uitdaging daarbij is dat de bewustwording van de noodzaak om lokale kopieën te maken van onlinedata nog vaak ontbreekt. In gesprekken met sommige (ondergerepresenteerde) groepen horen we al wel dat ze hun Facebook- of Instagrampagina beschouwen als hun historisch archief. Het historische besef is er dus dikwijls al, maar de kwestie van het eigenaarschap van onlinedata ligt moeilijker. De socialemediaplatformen hebben volledige controle over hun onlinedata. Zeker activistische groepen lopen een groot risico geband te worden.

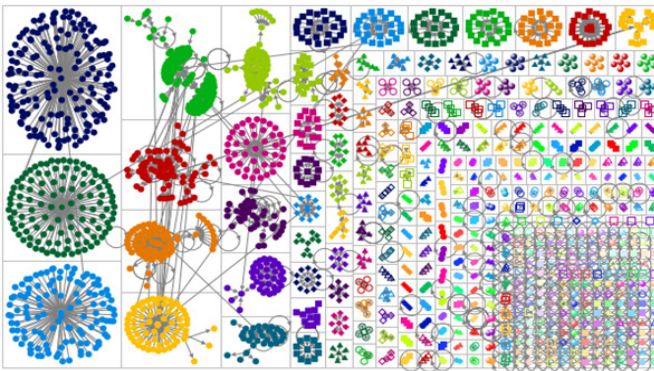
Voordat we participatief kunnen waarderen, moeten we eerst aan bewustwording werken. Dat vraagt veel tijd en moeite. We moeten de doelgroepen heel gericht benaderen, omdat elke groep een ander type online aanwezigheid heeft. Zo zijn de traditionele vredesbewegingen en de Noord-Zuid-beweging in Vlaanderen relatief weinig aanwezig op Twitter. 11.11.11 heeft wel meer dan 8000 volgers op Twitter, maar hun tweets hebben amper 'engagement' en hashtag-campagnes zoals #changemakers en #murals4change worden amper gebruikt.

De relevante gemeenschappen laten participeren is bovendien onvoldoende. Zeker bij controversiële gebeurtenissen is het belangrijk dat een archiefin-

stelling er op een neutrale manier mee omgaat en ook eventueel een kritische stem of tegenstem bewaart.

Ook volledigheid is een belangrijk waarderingscriterium. Datavisualisatie in de vorm van hashtag-*mapping* en netwerkanalyses is absoluut noodzakelijk om alle activiteiten van groot-schalige events in beeld te brengen. Zo hebben we een deel van de tweets over de Leopold II-standbeelden tijdens de Black Lives Matter-protesten met Nodexl gevisualiseerd. Het grote aantal clusters en het relatief beperkte aantal contacten tussen de clusters tonen duidelijk aan dat het om een gedecentraliseerde sociale beweging gaat. De dataset is niet gearchiveerd op basis van een hashtag, maar wel met de trefwoorden 'LeopoldII' en 'Leopold II'. In totaal hebben we 55.532 tweets gearchiveerd over de standbeelden, waarvan ik er 2000 heb gevisualiseerd op

de afbeelding. Een verdere analyse van de dataset met enkele Pythonscripts toont aan dat maar 46,39 procent van de tweets (25.759) een hashtag bevat en dat er 683 verschillende hashtags de ronde deden over dat ene onderwerp. Een groot deel van de hashtags verwees naar de plaatsnamen waar de standbeelden staan, zoals #Ekeren, #Brussel en #Bruxelles, of naar nieuws-media, zoals #Vrtnws en #Atvnieuws. Dergelijke hashtags zijn voor ons niet interessant om te *scrapen*, omdat we zo ook heel wat informatie die niets met het onderwerp te maken heeft, mee krijgen. Daarom archiveren we liever met trefwoorden. Wel bleek uit de analyse dat we sommige (minder gebruikte) hashtags over het hoofd gezien hadden. Zo hadden we #BLMBelgium en #BlackLivesMatterBelgium wel gearchiveerd, maar #BLM_Belgium niet.



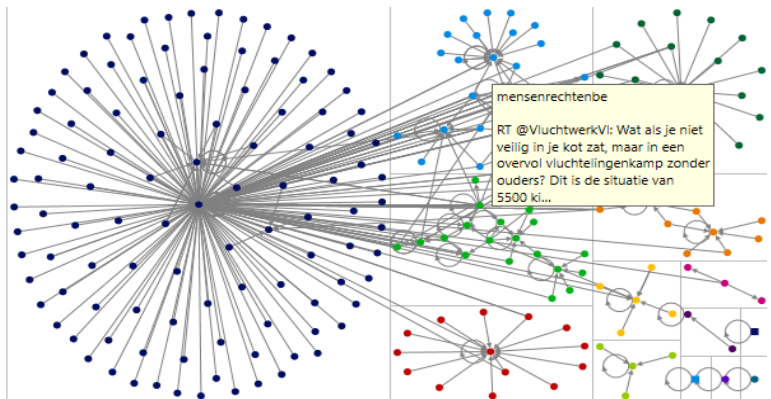
Nodexl-visualisatie van 2000 tweets over de Leopold II-standbeelden tijdens de Black Lives Matter-protesten van 2020.

Een tweede voorbeeld is de visualisatie van een centraal gestuurde hashtag-campagne van Vluchtelingenwerk Vlaanderen. Vanaf april 2020 verspreidde dat #ikbensolidair om lokale overheden tijdens de coronapandemie te overtuigen solidair te zijn met vluchtelingen. Het is duidelijk dat weinig personen de hashtag los van de officiële deelnemende kanalen gebruikten: aan de verbindinglijnen is te zien dat nagenoeg alle clusters in relatie staan tot Vluchtelingenwerk Vlaanderen (de paarse cluster). De lichtblauwe cluster is bijvoorbeeld opgesteld rondom een retweet van de Liga voor Mensenrechten. De visualisatie maakt onmiddellijk duidelijk dat de campagne centraal gestuurd is en dat wellicht geen verdere Pythondata-analyses nodig zijn om naar andere hashtags te speuren.

Praktische gebruikswaarde en sociaal-maatschappelijk belang

Naast het gebruik in academische studies hebben datasets van gearcheiverde hashtags vooral een groot potentieel voor sociaal-maatschappelijk en educatief werk. Het zijn immers gebruikelijke historische bronnen. Zo gaan de Amerikaanse Twitterarchieven The COVID-19 Hate Speech Twitter Archive (CHSTA) en The Counter-anti-Asian Hate Twitter Archive (CAAHTA) veel verder dan louter documenteren. Tijdens de covidperiode wilden ze haatberichten tegen Amerikanen met Aziatische achtergrond in kaart brengen. Van het overvloedige en heterogene digitale materiaal dat realtime werd gemaakt, stelden ze een pakket op dat onmiddellijk werkbaar was, ook terwijl de gebeurtenissen nog gaande waren.

Nodexl-visualisatie van tweets met #ikbensolidair in de periode 9-16 april 2020.



Dat soort analyses werd ingeschakeld voor lobbywerk en gebruikt door politici, journalisten en onderwijzers.⁸

Het sociaal-maatschappelijke belang is een waarderingscriterium dat in het geval van Amsab-ISG overlapt met de vraag of de bronnen binnen ons collectieprofiel passen. We richten ons specifiek op groepen die in de traditionele archieven minder voorkomen. Sociale media bewaren kan helpen om die groepen een plaats te geven in het archief. Sommige gemeenschappen, zoals Black Twitter⁹, zijn namelijk heel actief op bepaalde platformen.

Ook het educatieve potentieel valt niet te onderschatten. Wanneer leerkrachten en leerboeken actuele thema's aanhalen, kunnen ze niet meer om de rol van sociale media heen, zie bijvoorbeeld het belang van #Fridaysforfuture voor de organisatie van de klimaatmarsen en de rol van #klimaathysterie in de verspreiding van fake news.

Gevolgen van technische moeilijkheden voor de collectiewaarde

In waarderingskaders voor archieven zijn verschillende criteria te vinden over de waarde voor de collectie. Is het materiaal uniek? Is de collectie volledig? Kunnen we het toegankelijk maken? De technische uitdagingen van *data scrapen* bemoeilijken de objectieve invulling van de waarderingscriteria. Door de limieten van de – voor ons toe-

gankelijke – Twitter-API kunnen we maar zeven dagen terug in de tijd gaan om hashtags te bewaren. We moeten daarom elke week opnieuw beslissen wat we willen opnemen. Als we iets missen, te laat zijn of een fout maken, is die in het geval van hashtags heel moeilijk te herstellen, zelfs als de data nog online staan. Toen het nepfestival La Boum op 1 april 2021 plaatsvond, heb ik te lang getwijfeld of we dat moesten opnemen. Toen ik wat later dan toch besliste om tot archivering over te gaan, ging juist een Koreaanse popgroep met de naam Laboum viraal onder dezelfde hashtag. Filteren op taal of coördinaten was geen optie, omdat de muziekgroep ook in België heel wat fans had. Doordat een groot percentage van de tweets over K-pop gaan en niet zomaar weg te filteren zijn, is onze dataset over La Boum dus volledig onbruikbaar voor onderzoek.

Het volledigheidscriterium zal altijd een probleem zijn. Het voorbeeld van de standbeelden van Leopold II (cf. supra) toonde al aan dat een groot percentage van de gebruikers niet met hashtags werkt. Om toch te streven naar volledigheid doe ik dikwijls ook een bijkomende *harvest* op trefwoord. Door de anti-abortusgebeurtenissen in de VS in juni 2022 laaide ook het debat in andere landen, waaronder België, hoog op. Enkele activistische hashtags gingen viraal, waaronder #abortionrightsarehumanrights en #mybodymychoice. Maar door alleen

die hashtags te bewaren, kwam het hele debat niet in beeld. Om zoveel mogelijk naar volledigheid te streven, heb ik elke tweet met het Nederlandstalige trefwoord 'abortus' gearchiveerd. Dat is geen perfecte oplossing. Zo kregen we ook de tweets uit Nederland mee. Daarnaast waren er ongetwijfeld Belgische tweets over het onderwerp waarin het woord abortus niet voorkomt. De comments leveren een bijkomend volledigheidprobleem. Door de technische moeilijkheden bij het *scrapen* nemen we ze dikwijls niet op.

De interpretatie van het volledigheidscriterium is ook onduidelijk als het gaat over sociale media: telt alleen de inhoud of moeten we ook de look-and-feel (lay-out) mee bewaren om volledig te zijn? In de praktijk is het dikwijls simpeler om alleen de inhoudelijke data en metadata te *scrapen* dan om de look-and-feel te bewaren. Dat brengt ons tot de problemen met het waarderingscriterium 'toegankelijkheid'. Hashtags bewaren we in JSONL-bestanden. Dat zijn tekstbestanden met de inhoud van de tweets en de bijbehorende metadata. De afbeeldingen en video's van de tweets moeten we apart archiveren. Met JSONL zijn data-analyses te maken. Er is wel kennis nodig van de *command line interface* en programmeertalen zoals Python om bestaande scripts te gebruiken of er zelf te schrijven en dat vormt een grote hindernis voor de toegankelijkheid.

Conclusie

De gebruikelijke waarderingscriteria voor archieven zijn moeilijk toepasbaar op hashtags. Het academische belang voor de geschiedkunde kunnen we nog niet inschatten, omdat we door de technische beperkingen al in de eerste zeven dagen na een gebeurtenis de beslissing moeten nemen of we de hashtags willen bewaren. Binnen andere academische domeinen zijn er al meer studies naar de rol van hashtags, maar die vakgebieden vinden dikwijls de weg niet naar archiefinstellingen.

Participatie is cruciaal. Niet alleen om ethische redenen, maar ook omdat we simpelweg geen andere mogelijkheid hebben om in deze periode van *information overload* te weten wat er allemaal gaande is. Trial-and-error bij het wekelijks archiveren van hashtags heeft aangetoond dat zelfs de meest vanzelfsprekende waarderingscriteria een probleem kunnen geven door technische hindernissen. Positief is dat er heel wat opportuniteiten zijn op sociaal-maatschappelijk vlak en voor onderwijs, maar dat moeten we nog verder in de praktijk brengen.

- 1 Eindrapport van Cemper. Naar een generieke methode cross-collections waarderen, pilootproject waarderen. 01/10/2018-30/09/2019.
- 2 <https://faro.be/erfgoed-waarderen>; <https://archiefwiki.pleio.nl/wiki/Waardering>. Laatst geraadpleegd op 29/09/2022.
- 3 Zie in het bijzonder: Ghonim WAEL, *Revolution 2.0: The Power of the People is Greater Than the People in Power: a Memoir*, 2012, 308 p.; Zachary STEINERT-THRELKELD, Delia MOCANU, Alessandro VESPIGNANI, James FOWLER, Online social networks and offline protest. In: *EPJ Data Science*, 4(2015)19, pp. 1-9.
- 4 Derek MOSCATO, Media Portrayals of Hashtag Activism: A Framing Analysis of Canada's #Idlenomore Movement. In: *Media and communication*, 4(2016)2, p. 6; Aim SINPENG, Hashtag activism: social media and the # FreeYouth protests in Thailand. In: *Critical Asian Studies*, 53(2021)1, p. 4
- 5 Henrik Serup CHRISTENSEN, Political activities on the Internet: Slacktivism or political participation by other means? In: *First Monday*, (2011)
- 6 Evgeny MOROZOV, *The net delusion. How not to liberate the world*, Penguin Books Ltd, 2011, 408 p.
- 7 Anthony Wing KOSNER, The Solar Storm of KONY 2012, It's Not That Simple. In: *Forbes Magazine*, www.forbes.com/sites/anthonykosner/2012/03/11/the-solar-storm-of-kony-2012-its-not-that-simple/#7fe255e31a05. Geraadpleegd op 29/09/2022.
- 8 Powerpoint van de presentatie Archival Data Thinking and Practices for Social Media Collections, (2021, UCLA IS240 Lecture). In: *OSF*. <https://osf.io/2tg7v/>. Geraadpleegd op 29/09/2022.
- 9 https://en.wikipedia.org/wiki/Black_Twitter. Geraadpleegd op 29/09/2022.