



# **Archivering van Nederlandstalige tweets: de website [twiqs.nl](http://twiqs.nl)**

Erik Tjong Kim Sang

Hilversum, 11 oktober 2018

Sinds eind 2010 verzamelen wij Nederlandstalige tweets voor wetenschappelijk onderzoek

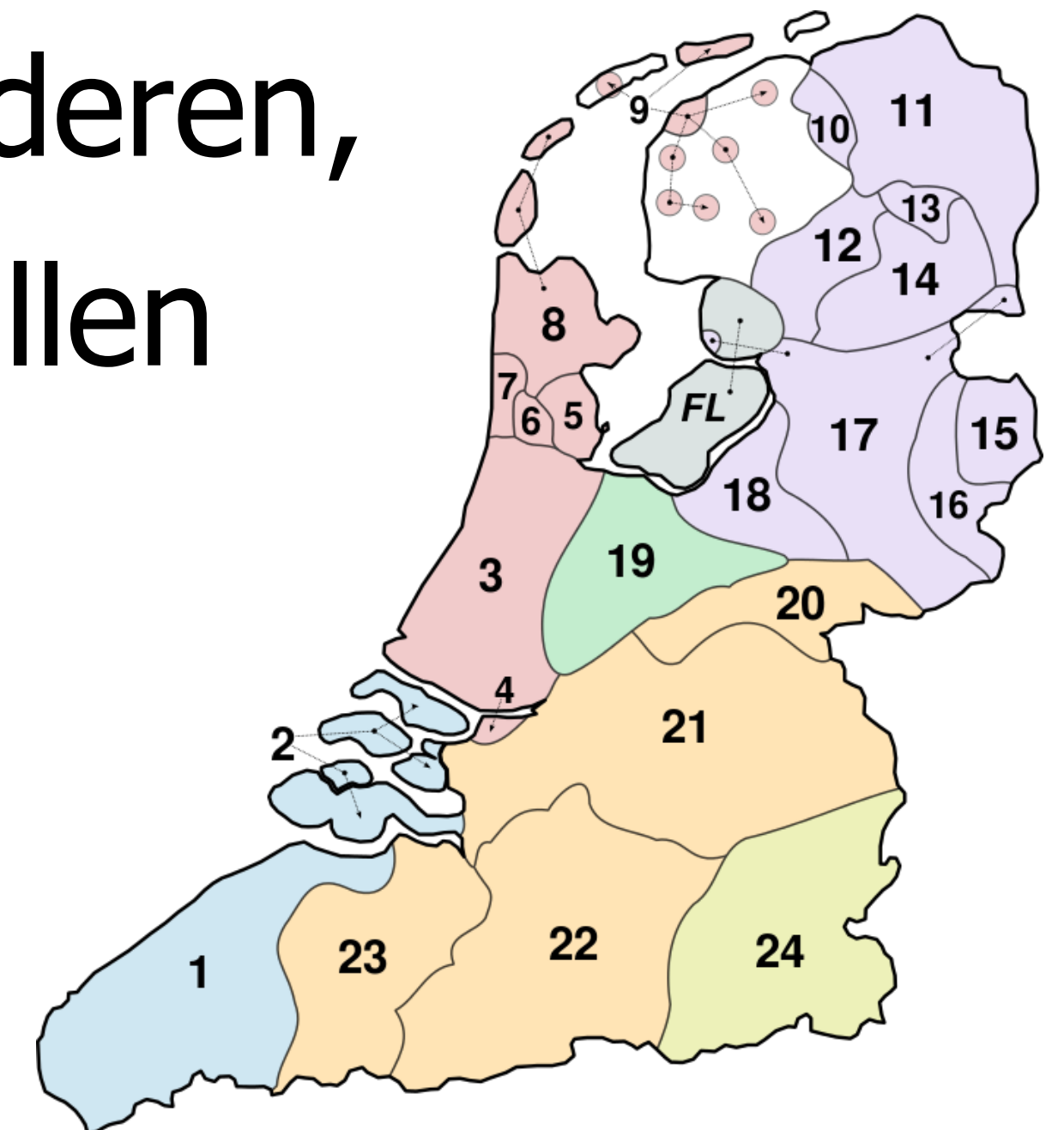
We zoeken in alle nieuwe tweets naar woorden die uniek zijn voor het Nederlands en slaan de tweets met deze woorden op



Vanaf 2013 zijn de tweets voor iedereen doorzoekbaar op de website [twiqs.nl](http://twiqs.nl)

Tweets worden gekozen op vier manieren:

1. via een zoekwoordenlijst van 229 Nederlandse woorden
2. met een lijst van 5000 veelschrijvers op Twitter
3. door geografische selectie: Nederland, Vlaanderen, Suriname en de voormalige Nederlandse Antillen
4. via een lijst met 49 dialectwoorden



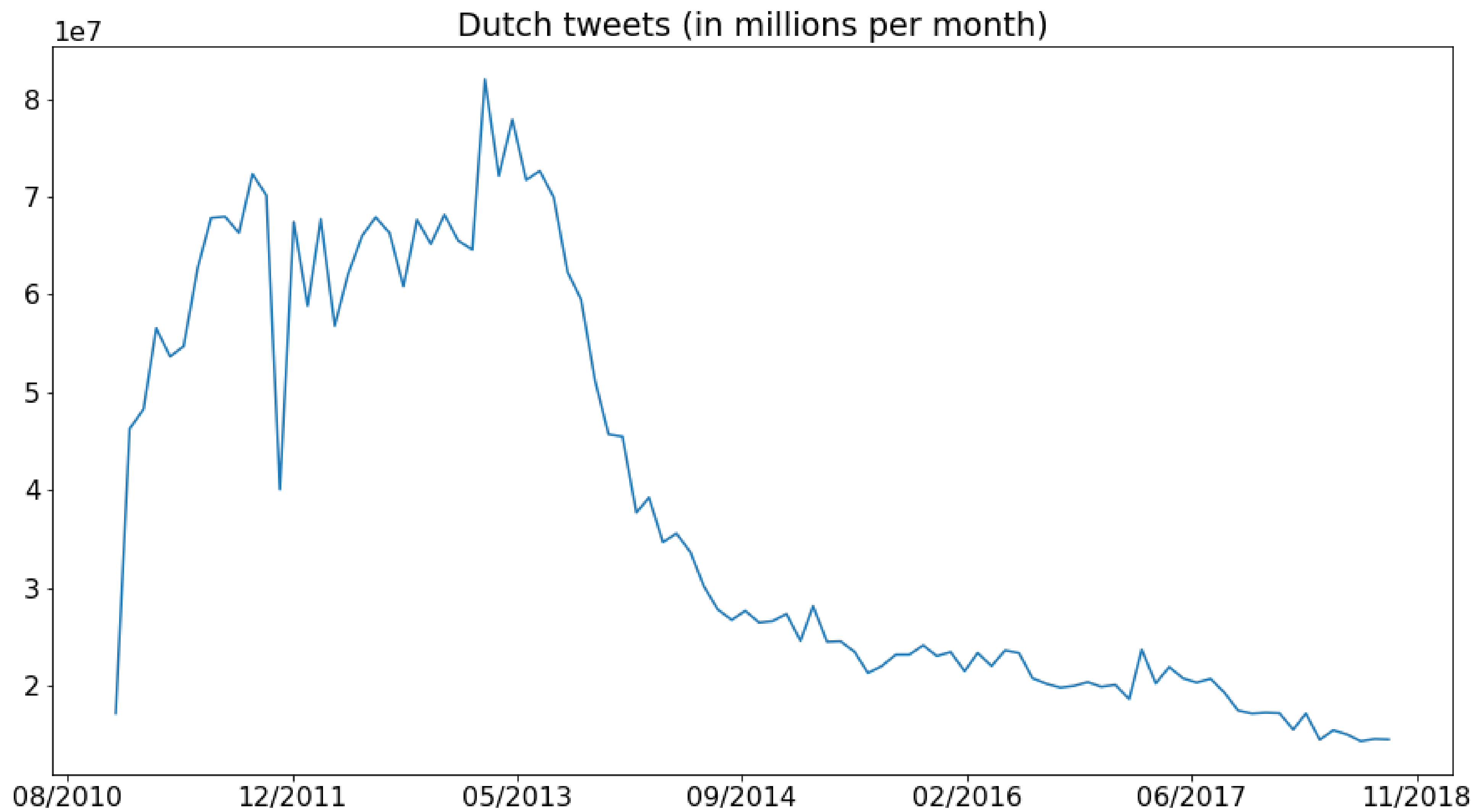
Sinds eind 2010 hebben we 3,6 miljard Nederlandstalige tweets verzameld

Er zijn tientallen wetenschappelijke publicaties die naar de tweetcollectie verwijzen

De website wordt per dag door ongeveer 50 unieke ip-adressen bezocht



# Aantallen verzamelde tweets per maand



Naam:

Wachtwoord:

Inloggen

[Registreren](#)

## Zoek in tweets:

Zoeken

11  oktober  2018  -

Op twiqs.nl kan je in Nederlandstalige tweets vanaf december 2010 zoeken. De gevonden tweets kan je dan op Twitter bekijken. De zoekdatabase is niet compleet. Contactpersoon voor deze website is Erik Tjong Kim Sang <erikt(at)xs4all.nl>

### Voorbeelden:

- ["mh17" op donderdag 17 juli 2014](#) (grafiek)
- ["hedde" in juli-september 2014](#) (kaart voor dialectwoord)
- ["nagels" in september 2017](#) (gebruikersinformatie)
- ["ajax,pec" op zondag 20 april 2014](#) (grafiek)
- ["piet" in november-december 2016](#) (woordenwolk)
- [alle tweets van woensdag 29 november 2017](#)

Created at: 10 October 2018 11:37:23. [erikt\(at\)xs4all.nl](#) ([clusterstatus](#))

Via de website kunnen 7 aspecten van tweets over een bepaald onderwerp worden bekeken:

1. Grafiek percentages tweets gedurende een bepaalde tijd
2. Kaart verzendpositie tweets
3. Frequente woorden in tweets
4. Frequente gebruikers
5. Kenmerken van de gebruikers: geslacht en leeftijd
6. Gemiddeld geschat sentiment van de tweets
7. Ids van de tweets

Twitter staat niet toe dat verzamelde tweets worden gedeeld met derden

Wel mogen de ids van tweets worden verspreid

Op twiqs.nl zijn de ids van gevonden tweets en de bijbehorende gebruiker te vinden

Hiermee kan op Twitter de tekst van de tweet worden gevonden

twiqs.nl [FAQ](#) [Blog](#) [Ngrams](#)    Naam:     Wachtwoord:     [Inloggen](#)    [Registreren](#)

**Zoek in tweets:** hedde


1 juli 2014 00:00 - 30 september 2014 23:00

**GRAFIEK**    **KAART**    **WOORDEN**    **GEBRUIKERS**    **TWEETS**    **HOME**

Gezocht in: 84.767.923 tweets. Gevonden: 1.968 tweets met het trefwoord "hedde" (0.002%).

@1723525310: 483743803460050944  
@142039870: 483797972430618625  
@488835358: 483853918804647936  
@632124927: 483855138525020160  
@2328728249: 483855148582961152  
@2429860816: 483856334103019520  
@2298909323: 483866040141905920  
@423636647: 483899040632156160  
@1095648626: 483926356708065280  
@373889332: 483972387516723200  
@408730292: 483975171750256640  
@1699587283: 484001116532011008  
@216355052: 484020912455684096  
@54201947: 484027132629495808  
@620478822: 484037020994768897  
@906675122: 484039053726457856  
@108990209: 484045116358000642  
@108990209: 484046017827176448  
@1328458860: 484047135328186372  
@47885314: 484049245495123968  
@263718987: 484056419604656128  
@2191425664: 484059981634109440  
@81853255: 484072237851742208  
@20839777: 484078835424829440  
@171143578: 484079897519083523  
@272045941: 48408452517379076  
@148846858: 484108830599835648  
@323833337: 484119264463699968  
@400076878: 484119348563693568  
@36638809: 484221359724457984

Klik op de getallen om de tweets te bekijken.  
[\(download ids\)](#)

Created at: 10 October 2018 12:07:59. [erikt\(at\)xs4all.nl \(clusterstatus\)](#)    



Ton van der Wouden onderzoekt het gebruik van partikels: kleine inhoudsloze woorden

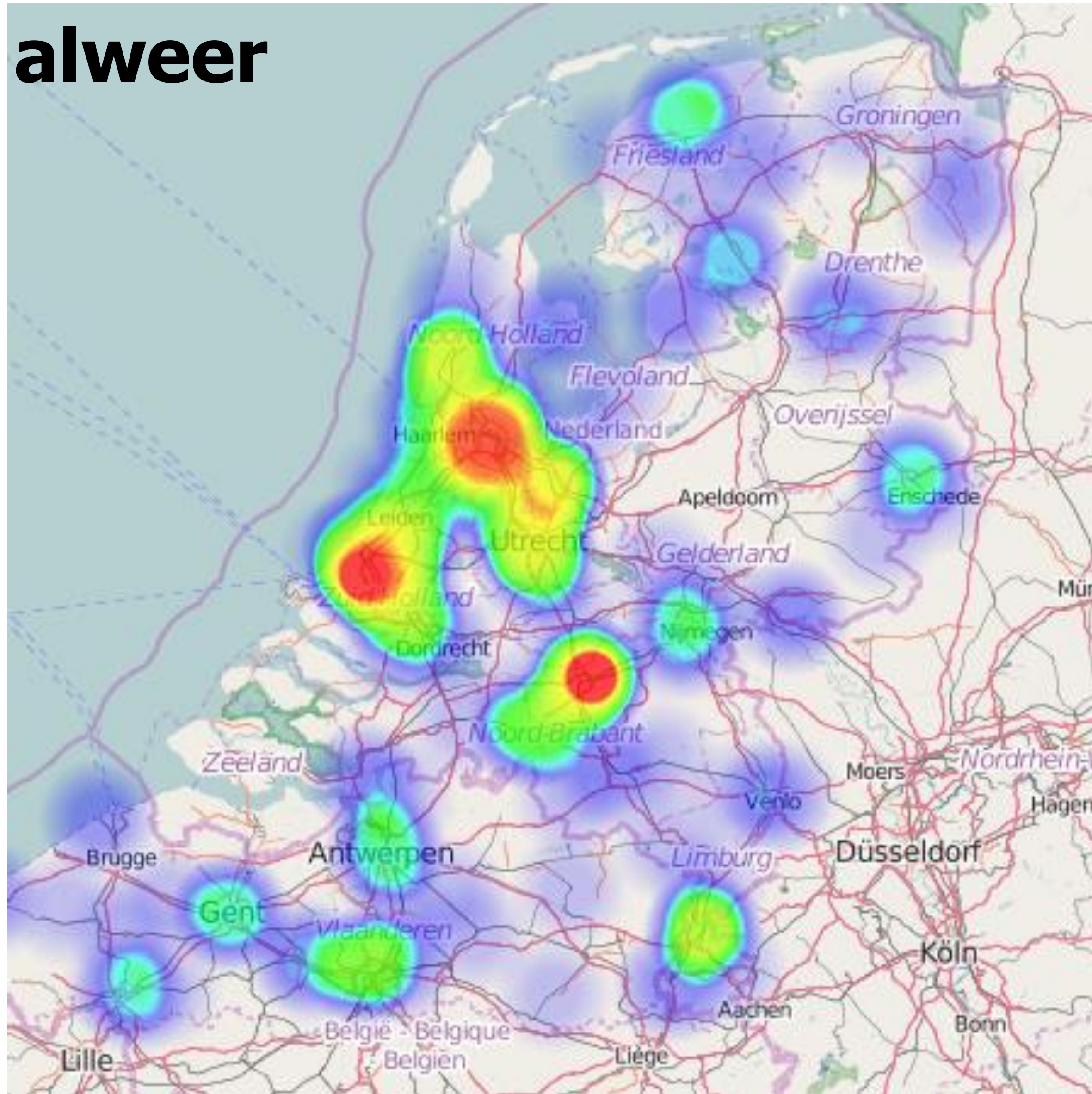
Voorbeeldzin: Je wil zeker wel iets te drinken, hè?

Er is regionale variatie in het gebruik van partikels

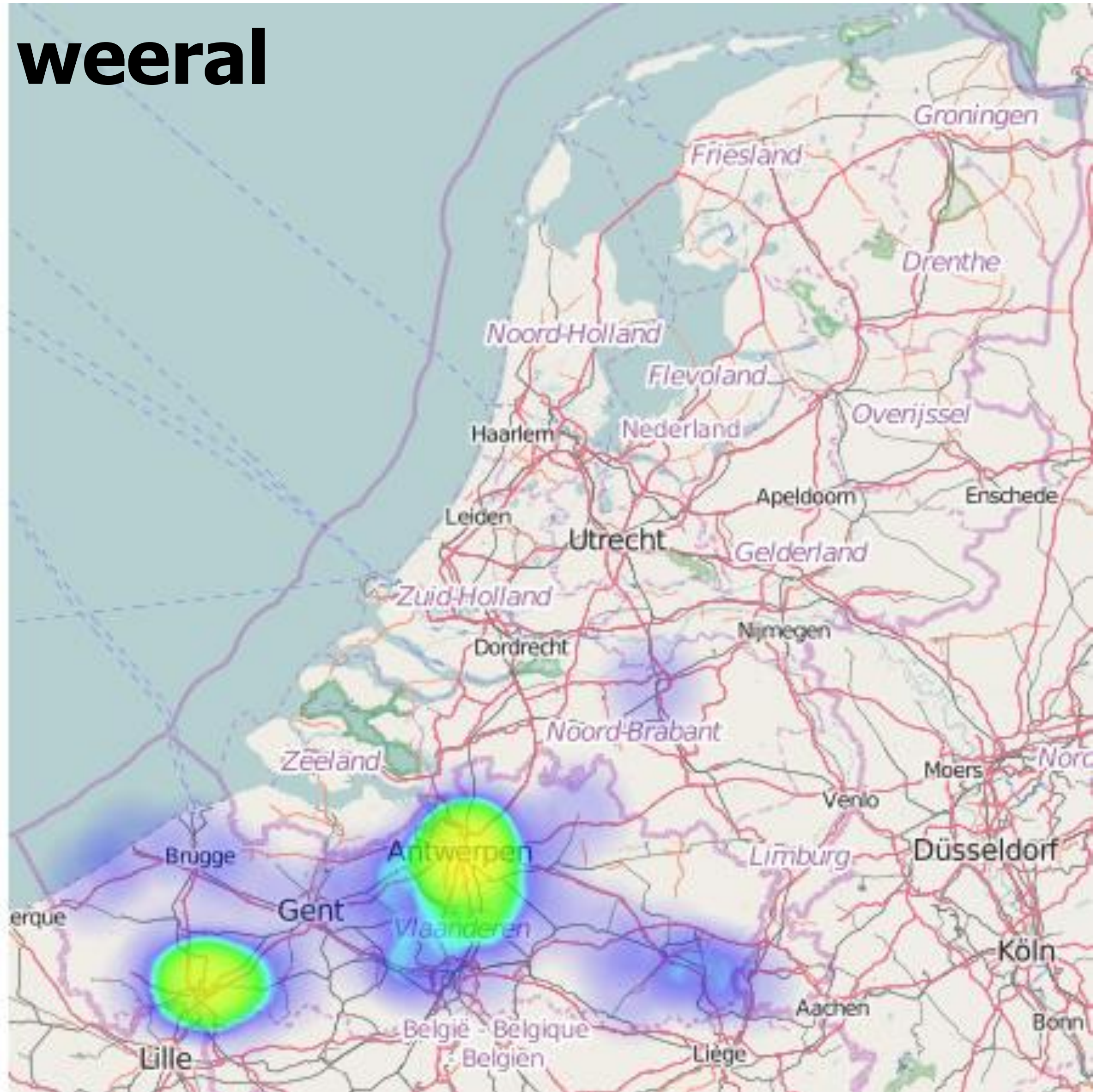
Bijvoorbeeld: alweer versus weeral

# Onderzoeksvoorbeeld: Nederlandse taal

## alweer



## weeral



In 2011 voorspelden Erik Tjong Kim Sang en Johan Bos de Provinciale Statenverkiezingen aan de hand van 1 week tweets

Die week werden 15 miljoen Nederlandstalige tweets verstuurd

Dat waren er te weinig om resultaten per provincie te voorspellen

Daarom hebben ze de resultaten van de verwante landelijke

Eerste Kamerverkiezingen voorspeld

## Onderzoeksvoorbeeld: Provinciale Statenverkiezingen (Nederland)

Partij	Uitslag	Politieke Barometer	Maurice de Hond	Twitter
VVD	16	14	16	14
PvdA	14	12	11	16
CDA	11	9	9	8
PVV	10	11	12	10
SP	8	9	9	6
D66	5	7	5	8
GL	5	4	4	3
CU	2	3	3	3
50+	1	2	2	2
SGP	1	2	2	2
PvdD	1	1	2	2
OSF	1	1	0	1
<b>Afwijking:</b>		14	14	18



# Conclusies

We presenteerden de website [twiqs.nl](http://twiqs.nl)

De website ontsluit een verzameling historische Nederlandstalige tweets

De tweets zijn gebruikt in divers wetenschappelijk onderzoek