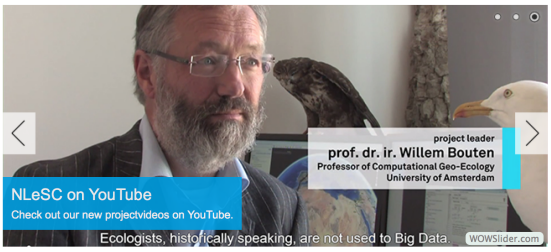
Using Tweets for Assigning Sentiments to Regions

Erik Tjong Kim Sang
Meertens Institute
Amsterdam, The Netherlands

27 May 2014

netherlands eScience center

Home | About us | News | Agenda | Projects | Current Calls | Media | Platform | Contact



Our mission:
The Netherlands eScience Center reinforces and accelerates multi-disciplinary and data-intensive research in the Netherlands by developing and applying eScience and by combining forces.

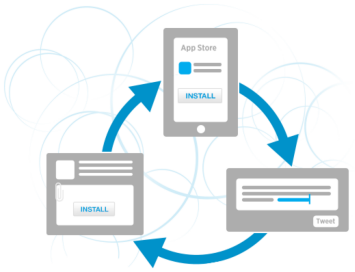
Developers API Health Blog Discussions Documentation Search Sign in

More downloads for your app

with Twitter Cards

Twitter Cards offer a fast and easy way to grow your user base for mobile apps. Simply add some new markup to your pages: when users tweet links to your domain, Cards will let other users viewing those Tweets to download and launch your app across a number of mobile platforms.

Learn More



Twitter Cards | Embedded Timelines | Embedded Tweets | Tweet Button | Follow Button

The project TwiNL

Enable researchers and students to collect statistics about tweets written in Dutch without needing technical knowledge:

- Collect tweets written in Dutch (we have more than 2 billion)
- Make the tweets searchable
- Offer different views of the search results

eScience center SURF SARA FAQ Blog Naam: Wachtwoord: Inloggen Registreren

Zoek in tweets: bier Zoeken
21 april 2014 - 27 april 2014

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>

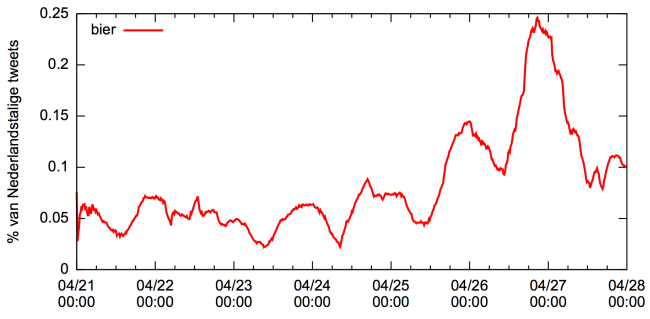
Voorbeelden:

- "griep" in januari/maart 2013 (grafiek)
- "carnaval" in februari 2013 (kaart)
- "ajax,feyenoord" op zondag 28 oktober 2012 (grafiek)
- "vaatstra" op maandag 19 november 2012 (gebruikersinformatie)
- "nait" in november 2012 (kaart voor dialectwoord)
- "fail" op maandag 3 december 2012, 08:00-08:59 (woordenwolk)

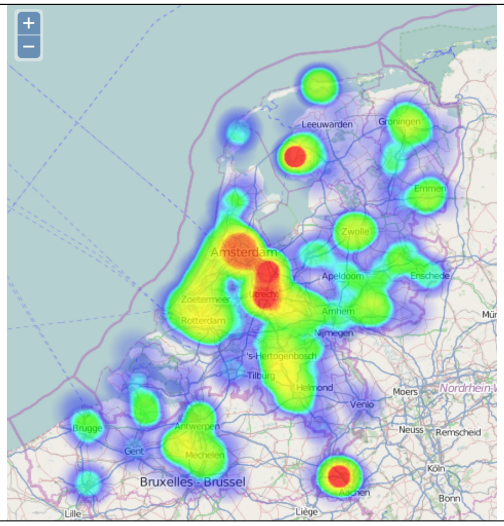
twiqs.nl

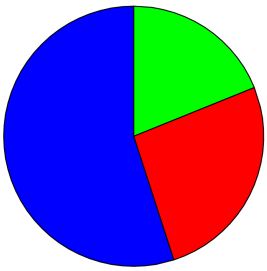
Created at: 26 May 2014 22:12:41. erikt(at)xs4all.nl (clusterstatus)

Gezocht in: 7.774.081 tweets. Gevonden: 6.805 tweets met het trefwoord "bier" (0.08%).

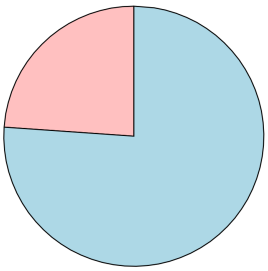


(download data van grafiek)





Leeftijd
< 18 jaar: 19%
18-25 jaar: 26%
> 25 jaar: 55%



Geslacht
man: 76%
vrouw: 24%

Sentiment: +11

What does the sentiment score mean?

The sentiment score is a number between -100 (minimum: negative) and +100 (maximum: positive)

The sentiment score is equal to the percentage of positive tweets minus the percentage of negative tweets

A single tweet can be marked as positive, negative or neutral

We use lists of positive words, negative words and emoticons to assign sentiment values to tweets

Methods used for selecting the sentiment lexicons

Start with assigning sentiment to tweets based on their emoticons

1. Automatic strategy: the top n ranked sentiment words from a ranked list were selected
2. Manual strategy: manually select sentiment words from frequently used words in these tweets

We use the t test for automatically creating ranked lists

We compared different values of n and different minimum word frequencies

Evaluation results on manually annotated data

Name	Lexicon size	Accuracy
baseline: 4 emoticons	4	87.6%
n-best t-scores, threshold 0	75,000	48.2%
n-best t-scores, threshold 5	4,400	53.2%
n-best t-scores, threshold 10	2,700	53.6%
incremental selection	644	84.2%
incr. sel. with 4 emoticons	100	93.2%
manual selection	338	82.1%

The emoticons have a low recall score, n-best does not perform well and incremental selection does not generalize

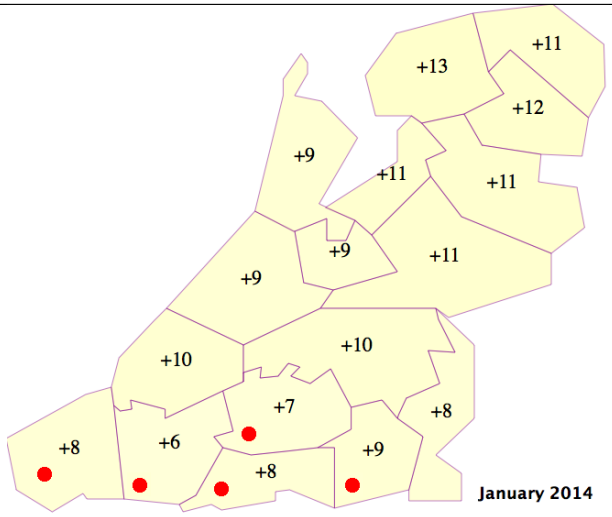
Can we obtain sentiment scores per region?

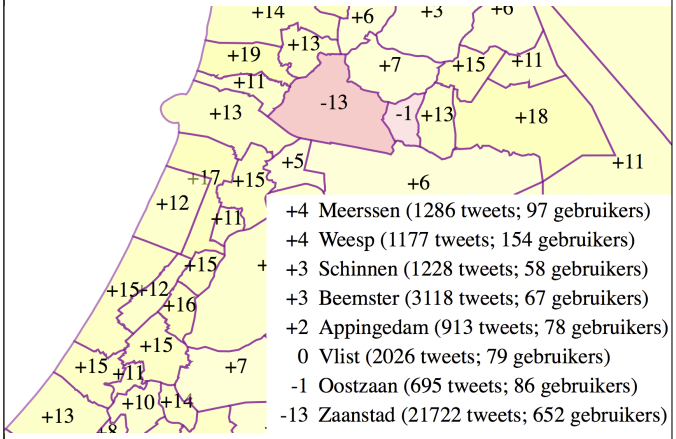
Between 1 and 2% of the tweets contain geolocation

For another 10% we can estimate a common location based on profile information of the user

We can combine this information with region boundaries and automatic sentiment analysis to average sentiments per region

We made map of the Dutch-speaking provinces and obtained municipality borders from the government





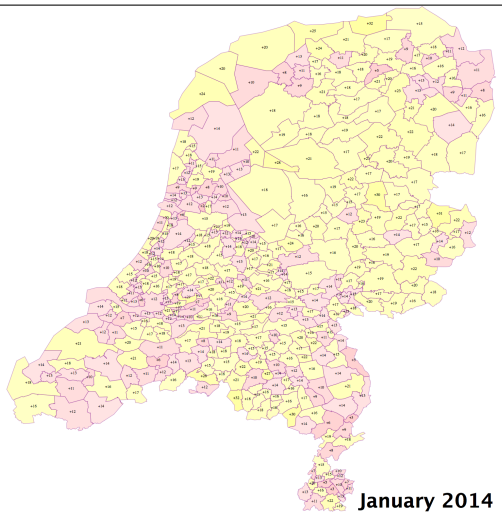
MeteoWestzaan.nl
@MeteoWestzaan

Follow

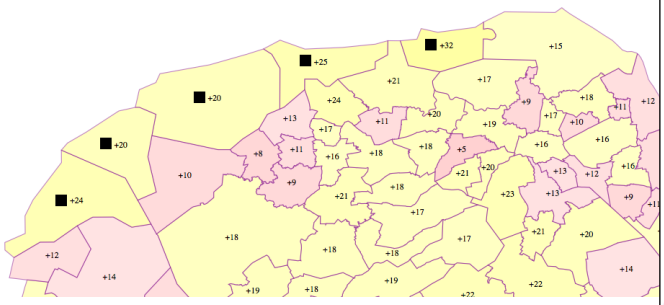
Baro 1025,0mb-Stijgt langzaam. Temp 16,8c (-0,4). Hum 44%. Rain last 24h 0,0mm. Wind 0,2kph WNW / gust 2,2kph. UV 0. #hetweerinwestzaan

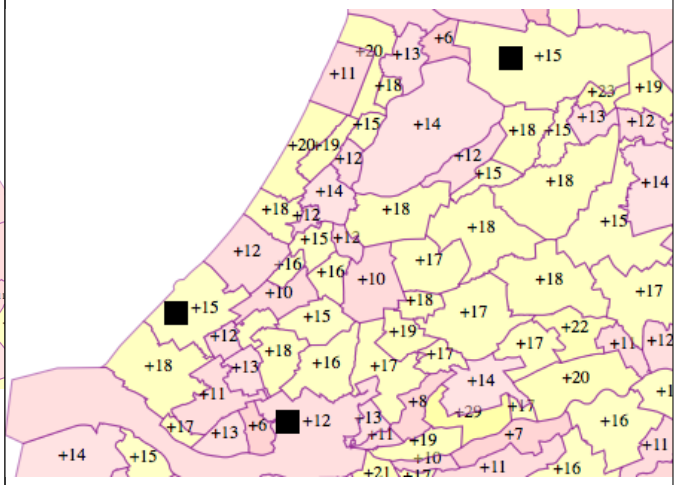
Reply Retweet Favorite More

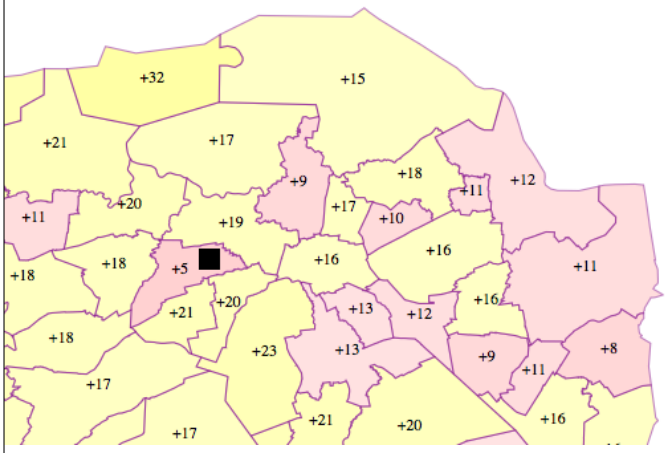
8:00 AM - 6 Oct 13 from Zaanstad, North Holland



January 2014







Concluding remarks

We compared several methods for deriving sentiment lexicons and concluded that manual construction worked best

We computed average sentiment scores and found differences between Dutch and Flemish regions, and between busy and quiet regions

It would be nice if these strategies provided useful information for policy makers

But remember: Twitter user demographics are different from population demographics!

THE END