

A fresh approach to web scraping of news sites for in-depth studies of crashes.

Rob Decae, Bart van den Broek,

Frits Bijleveld

27 September 2022

Presentation overview

- Background of the project
- Aims
- Challenges
- Methods, techniques and results
- Discussion

Project background: registration of crashes in The Netherlands

- Registration by police, road inspectors, salvage companies, etc.
- Self-reporting of crashes (app) for insurance companies
 - Not yet in official registration
- Missing crashes:
 - Motor vehicles crashes without (serious) injuries
 - crashes with bikes or mopeds without motor vehicles involved

Project background: in-depth studies of crashes

- information about all aspects of a crash:
 - focus on a specific type of road crash
 - the traffic situation,
 - the immediate surroundings,
 - the road users involved,
 - their vehicles,
 - and the injuries of the people involved.

Information gathering in depth

- Inspection of crash sites
- Interviews
- Police information
- Analyse information from news sources
- Medical information about victims

Home > Regio > Leiden

Twee gewonden bij botsing tussen scooter en fietser op kruispunt in Leiden



NET BINNEN

Het eind van de droogte komt in zicht: waterschappen draaien de noodkraan langzaam dicht

Blikjes, flesjes, zonnebrillen en zelfs een riool. Duikers halen honderden afvalresten van bodem Zegerplas naar boven: 'Helaas een recordopbrengst'

Premium De Nederlandse afgang in 1990; gaat de geschiedenis zich herhalen? Zou de oude Van der Poel het er met zijn zoon over gehad hebben? | column

Premium Bij brand neemt Peter Prinssen het trouwboekje van zijn overgrootouders meteen mee

Premium Britten staan eindelijk in de rij voor hun koningin

Premium Volgens een paranormaal medium: inkerroeg 'heavy stuff' en waarnemen. Wat moet ik doen

stelt zich na tik in derby is het dat we nu gewoor en Katwijk staan'

ari legt de mensheid ui

laag gewond na ongeval met vrachtwagen monstersestraat/

1-10 oeding on fire | rapportage 2022-01-01 | Middelb - SWOV | GDPR And EU Data L... | https://www.2hangs... | Reguler | JavaScript R... | (D-) Networ Wafflen... | Zelf kee maken | Solog... |



Fietser zwaargewond na ongeval met vrachtwagen Monstersestraat

Vrijdag 26 augustus 2022 11:22

Den Haag - Vanochtend is een fietser zwaargewond geraakt na een ongeval met een vrachtwagen op de Monstersestraat.

Het ongeval gebeurde om even voor half twaalf ter nabij de kruising met de Loosduinsekade.

Meerdere eenheden van de politie, de brandweer, ambulancedienst en de traumahelicopter met het Mobiel Medisch Team uit Rotterdam werden met spoed opgeroepen. De inzet van het MMT werd even later geannuleerd.

Westlanders.nu

Nieuws Veiligheid Agenda Kerk & geloof Dieren Politiek Zakelijk Vacaturebank Vastgoed Oekraners

75 € 6.166,95 € 8.170 € 2.396,95 € 9.565,50 € 3.370 € 2.599

13% 14% 32% 11%

1,95 € 2.399 € 2.399 € 1.350 € 15.637 € 1.599 € 7.990

POPAL
Bekijk de actiemodellen

GRATIS BACK TO SCHOOL PAKKET
t.w.v. € 70

€ 1.499 € 1.499

Westlanders.nu zakelijk

Westlanders.nu tijljin
Bel 06 - 300 90 416

Je bent geliefd

HEMA

Fietser gewond na botsing met scootmobiel

Geplaatst door Westlanders.nu op 24-08-2022 18:00 · Gewijzigd op 24-08-2022 19:24

Honselersdijk 24.08.2022 - Aan de Molenlaan in Honselersdijk raakte woensdagmiddag even voor half zes een fietser gewond na een ongeluk op het fietspad.



Nieuws op Beeld

Home Nieuws Regio Nieuws Insturen Over Ons Contact

Motorrijder overleden bij ongeval op Moordrechtboog (N457) in Waddinxveen (video)

Gouda E.O. | Waddinxveen

Door: Redactie Date: 9 augustus 2022



Project background:

Why web scraping and text analyses?

- Keeping up with 'new' techniques/improvements
- Obtain knowledge allowing us to assess products made by third parties
- Generic solution: Cost and time efficiency for research
- Use as a exploration environment for text analyses in restricted data sources (e.g. police)

Project targets

- Use Universal search terms – finds relevant URL's
- Convert webpages to a uniform, easy to scrape, format
- Select only relevant content on these webpages
- Extract the main text
- Practise text analyses
- Present the information to researchers

advertising

Upper banner

Search button

linking

Title

Side banner, links to articles

questionnaire

Photos

Main text



Home > Regio > Duin- en Bollenstreek

Beschonken fietser met telefoon in hand botst al slingerend op scooterrijder in Voorhout



© Foto's: Flashlight Fotografie

Internetredactie

Zaterdag 27 augustus 2022 om 08:41

VOORHOUT

Een beschonken fietser is in de nacht van vrijdag op zaterdag frontaal in botsing gekomen met een scooter op het fietspad langs de Leidsevaart (N444) in Voorhout. De man verklaarde zelf te hebben gedronken en slingerend te

NET BINNEN

- Premium** 'Links bolwerk' heeft plaatsgemaakt voor een vol terras met volle borden | recensie Over de tong
- Premium** Leiderdorp wil honderd asielzoekers opvangen: 'veiligheid moet gewaarborgd worden'
- Premium** Alphen is straks niet meer met excuusgebied waar de Regiotand 'geoorloofd' te laat kan komen, hoopt de gemeente
- Premium** Stichting Diogenes verbouwt monument Hogewoerd 132 in Leiden tot twee appartementen
- Nederlandse (27) gesneuveld aan het oorlogsfront in Oekraïne
- Premium** Eefjes Singelpark, kinderboek over het langste stadspark van Nederland - Wie vriend wil worden van een boom, omhelst hem gewoon
- Premium** Ook Alphenaren met een grotere portemonnee kunnen per 1 januari weer gesubsidieerde huishoudelijke hulp krijgen: 'Beleid was in strijd met de wet'
- Brand in schuur bij woning in Leiden, gezin met baby beland door brand op straat
- Premium** Dorpscentrum Oegstgeest wil nieuwe burelen op termijn 'uitfaseren'

Geef uw mening!

Leidsch Dagblad is benieuwd of het toevoegen van bepaalde extra's een abonnement voor u aantrekkelijk maken. Wilt u meedoen aan een onderzoek?

Deelnemen

The challenges

- Centralise and automate the search facility
- Avoid changing HTML-structures and dynamic pages
- Avoid pollution from linked pages, banners etc.
- Analyse text and extract relevant information
- Group articles about same incident
- Recognise relevant /discard non-relevant articles

SWOV

Solutions, Techniques and Results

Solutions and techniques:

Centralise and automate search facility

➤ Challenges:

- Many different websites with different search facilities
- Scraping is difficult / changing web structure, dynamic HTML, multiple pages in search
- Solution: use Twitter as search facility and harvest linked URL's

➤ Techniques:

- Academic Twitter account
- Nice API's and interfaces for R and Python
- Advantage: No scraping overload of websites

Techniques:

Centralise and automate search facility 1

➤ Platform: R package

➤ Selection string

AND operator = " "

OR operator = ,

- ```
query1 <- build_query(c("aanrijding fiets","botsing fiets","ongeval fiets",
 "aanrijding ebike","botsing ebike","ongeval ebike",
 "fietser","fietster","rijwiel"),
 is_retweet = FALSE, has_links=TRUE)
```

# Techniques:

## Centralise and automate search facility 2

### ➤ Additional search parameters

- `get_all_tweets(`
- `start_tweets = "2018-01-01T10:00:00Z",`
- `end_tweets = "2022-05-26T10:00:00Z",`
- `users=c("regio15","RedactieD8","WestlandersNu","112HM","wosmedia",  
              "mediaterplaatse","nieuwsopbeeld", "TelstarOnline","leidschdagblad",  
              "KatwijkActueel","asmedianl","rtvkatwijk","112Maasmedia")`

# Results: Raw tweets imported in R

Tweet ID

List of  
URL's

List of  
hashtags

Author ID

| conversation_id     | author_id | id                  | pu |
|---------------------|-----------|---------------------|----|
| 1570878544414216199 | 22567659  | 1570878544414216199 |    |
| 1570388525136359424 | 28526900  | 1570388525136359424 |    |
| 1570326089691713537 | 23572475  | 1570326089691713537 |    |
| 1570044743114948608 | 28526900  | 1570044743114948608 |    |
| 1569734523402723330 | 28526900  | 1569734523402723330 |    |
| 1569693687537606656 | 23572475  | 1569693687537606656 |    |
| 1569673539975876612 | 78959883  | 1569673539975876612 |    |
| 1569655175366721541 | 78959883  | 1569655175366721541 |    |
| 1569615559041359874 | 735371256 | 1569615559041359874 |    |
| 1569343036521316353 | 78959883  | 1569343036521316353 |    |
| 1569316329714290689 | 78959883  | 1569316329714290689 |    |

| entities.media_keys                                    | entities.urls | entities.hashtags |  |
|--------------------------------------------------------|---------------|-------------------|--|
| "3_1570878459009695745", "3_1570878458963664897", "... | 6 variables   | 3 variables       |  |
| 8521906585603                                          | 17 variables  | 3 variables       |  |
| 6086151700481                                          | 20 variables  | NULL              |  |
| 1044737771524098", "3_1570044740963340288")            | 6 variables   | 3 variables       |  |
| 1734519133020160", "3_1569734521385361409")            | 6 variables   | 3 variables       |  |
| 13683951489025                                         | 20 variables  | NULL              |  |
|                                                        | 16 variables  | NULL              |  |
| 5172736983040                                          | 6 variables   | NULL              |  |
|                                                        | 17 variables  | NULL              |  |
|                                                        | 16 variables  | NULL              |  |

ID's of  
linked  
media

# Solutions and techniques: Obtain URL's for weblinks

## ➤ Challenges:

- Extract links and hashtags from Tweet

## ➤ Solution: convert Tweet to table format and extract separate URL's and hashtags

## ➤ Preference for external URLs over Twitter URL's

## ➤ Techniques:

- Conventional R conversion tools, mainly dplyr



# Twitter information in table format

| id                  | datum_tijd       | source          | url                                                                                                                                                                                                                               | hashtags                      |
|---------------------|------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------|
| 1563417929819230000 | 27/08/2022 08:47 | dlvr.it         | <a href="http://dlvr.it/SXK9JD">http://dlvr.it/SXK9JD</a>                                                                                                                                                                         | leeg                          |
| 1563126605588810000 | 26/08/2022 13:29 | DBgallery       | <a href="https://regio15.nl/nieuws/ongevallen/34838/fietser-zwaargewond-na-ongeval-met-vrachtwagen-monstersestraat">https://regio15.nl/nieuws/ongevallen/34838/fietser-zwaargewond-na-ongeval-met-vrachtwagen-monstersestraat</a> | Monstersestraat,DenHaag       |
| 1563104235046450000 | 26/08/2022 12:01 | dlvr.it         | <a href="http://dlvr.it/SXGjF0">http://dlvr.it/SXGjF0</a>                                                                                                                                                                         | leeg                          |
| 1563100423459790000 | 26/08/2022 11:45 | TweetDeck       | <a href="https://twitter.com/RedactieD8/status/1563100423459786755/photo/1">https://twitter.com/RedactieD8/status/1563100423459786755/photo/1</a>                                                                                 | Loosduinseweg,Monstersestraat |
| 1562675961086490000 | 25/08/2022 07:39 | RTV Katwijk     | <a href="https://www.rtvkatwijk.nl/fietser-gewond-na-botsing-met-porsche/">https://www.rtvkatwijk.nl/fietser-gewond-na-botsing-met-porsche/</a>                                                                                   | leeg                          |
| 1562522900271600000 | 24/08/2022 21:31 | dlvr.it         | <a href="http://dlvr.it/SX9zCj">http://dlvr.it/SX9zCj</a>                                                                                                                                                                         | leeg                          |
| 1562520593068870000 | 24/08/2022 21:21 | RepobakMTV      | <a href="https://twitter.com/112HM/status/1562520593068867584/photo/1">https://twitter.com/112HM/status/1562520593068867584/photo/1</a>                                                                                           | leeg                          |
| 1562491273004860000 | 24/08/2022 19:25 | Twitter Web App | <a href="https://www.westlanders.nu/nieuws/fietser-gewond-na-botsing-met-scootmobiel-40507">https://www.westlanders.nu/nieuws/fietser-gewond-na-botsing-met-scootmobiel-40507</a>                                                 | leeg                          |

# Challenge: Convert webpages to a uniform, easy to scrape, format

Leidsch  Dagblad

Lees de digitale krant

HOME

REGIO'S

SPORT

EXTRA

PUZZELS

Leiden

Alphen aan den Rijn

Duin- en Bollenstreek

Uitgelicht

Home >

Regio >

Leiden

Twee gewonden bij botsing tussen scooter en fietser op kruispunt in Leiden



1/2

© Toon van der Poel

Internetredactie

Dinsdag 23 augustus 2022 om 15:40

LEIDEN

NET BINNEN

Het eind van de droogte komt in zicht:  
waterschappen draaien de noodkraan langzaam dicht

Blikjes, flesjes, zonnebrillen en zelfs een riool. Duikers halen honderden afvalresten van bodem Zegerplas naar boven: 'Helaas een recordopbrengst'

Premium De Nederlandse afgang in 1990; gaat de geschiedenis zich herhalen? Zou de oude Van der Poel het er met zijn zoon over gehad hebben? | column

Premium Bij brand neemt Peter Prinssen het trouwboekje van zijn overgrootouders meteen mee

Premium Britten staan eindeloos in de rij voor hun koningin

Premium Volgens een paranormaal medium kan je in mijn linkeroog 'heavy stuff' uit mijn verleden waarnemen. Wat moet ik daarmee? | column

Quick Boys herstelt zich na tik in derby:  
'Des te mooier is het dat we nu gewoon winnen en boven Katwijk staan'

Futuroloog Harari legt de mensheid uit

# Classical approach: find HTML-tags

Article title tag

```
395 value <article-detail__article-main>/a> >
<!--[START-block-widget] 276ca2bd-52ec-4860 25cf4304dcb1 --><!--serverside resolved fragment --><div <span class="attribute-
name">id=V7hhAMxt2UO" data-
value">articleDetail"><div class=hmc-
ld-1d70522a_marginbottom5 hmc-ld-a5a70c0_contentwrapper"><h1 <span class="attribute-
name">class=hmc-ld-1d70522a_marginbottom5 hmc-ld-26838ee5_paddingleft0 hmc-
ld-26838ee5_paddingright0" data-testid=article-header"><h1 class=hmc-ld-ebe0ecc6_root hmc-ld-
ebe0ecc6_heading3 hmc-ld-300db776_none hmc-ld-91174671_primary hmc-ld-66f0fec7_bold hmc-ld-1d70522a_marginbottom0 hmc-
ld-1d70522a_margintop0 hmc-ld-b48c4984_inherit" style=color:var(--color-text-
grey-base)" data-testid=article-title">Twee gewonden bij botsing tussen
scooter en fietser op kruispunt in Leidenh1><header><div class=hmc-ld-1d70522a_marginbottom4">&
lt;div class=hmc-ld-
d32fd143_swipercontainer"><div class=<a class="attribute-
value">hmc-ld-d32fd143_swiperwrapper" data-testid=carousel-wrapper"&
gt;<div class=hmc-ld-9bdd7365_stretch hmc-ld-
d32fd143_swiperslide" data-testid=article-carousel-item"><figure data-testid=article-image-wrapper" <span
```



# Solutions and techniques: Convert web-pages to PDF-format

- Challenges:
  - Avoid changing HTML-structures
  - Convert page to an universal format
- Solution: use a “HTML to PDF”- convertor and feed the collected web links
- Techniques:
  - Open source command line tool WK<html>TOpdf (<https://wkhtmltopdf.org/>)
  - Call this Tool from R or Python

## Techniques: Conversion of html to pdf

- Call `wkhtmltopdf` tool
- `for(t in 1:nr_of_links){`
- `cmd=paste0("wkhtmltopdf  
--disable-external-links  
--disable-javascript  
--encoding 'ascii' ",filename$url[t]," file_",t,".pdf")`
- `system(cmd)`
- `}`
- `"wkhtmltopdf --disable-external-links --disable-javascript --encoding 'ascii'  
https://www.westlanders.nu/nieuws/fietser-ernstig-gewond-bij-ongeluk-monsterseweg-40413 file_52.pdf"`

[Home >](#)[Regio >](#)[Duin- en Bollenstreek](#)

## Beschonken fietser met telefoon in hand botst al slingerend op scooterrijder in Voorhout



### ■ NET BINNEN

■ **Wetenschappers Leiden keuren uitspraken ON-presentatrice af**

■ **Uit de Speakers: Creatieve Franse rapper inspireert om op het voetbalveld ook creatief te zijn**

■ **Omvangrijke bomentransfer bij Schipholweg en Boyleweg op komst**

**Premium** ■ **Valkenburg krijgt toch geen tweede begraafplaats, al wordt het krap: 'Die mensen gaan ook een keer dood'**

**Premium** ■ **Wethouder en gedeputeerde geven Burgerinitiatief Vlietland weinig houvast: 'Kunnen geen beloftes doen'**

**Premium** ■ **'Hekken in Bollenstreek moeten worden vervangen door heggenlagen'**

■ **Hutspotmaaltijd voor 4.687 bezoekers**



## Beschonken fietser met telefoon in hand botst al slingerend op scooterrijder in Voorhout



### ■ NET BINNEN

Wetenschappers Leiden keuren uitspraken ON-presentatrice af

Uit de Speakers: Creatieve Franse rapper inspireert om op het voetbalveld ook creatief te zijn

Omvangrijke bomentransfer bij Schipholweg en Boyleweg op komst

**Premium** Valkenburg krijgt toch geen tweede begraafplaats, al wordt het krap: 'Die mensen gaan ook een keer dood'

**Premium** Wethouder en gedeputeerde geven Burgerinitiatief Vlietland weinig houvast: 'Kunnen geen belofes doen'

**Premium** 'Hekken in Bollenstreek moeten worden vervangen door hekwerk met lage hekjes'

Hutspotmaaltijd voor 4.687 bezoekers

## Beschonken fietser met telefoon in hand botst al slingerend op scooterrijder in Voorhout



1 / 5

Intermedactie  
Zaterdag 27 augustus 2022 om 08:41

### VOORHOUT

Een beschonken fietser is in de nacht van vrijdag op zaterdag frontaal in botsing gekomen met een scooter op het fietspad langs de Leidsevaart (N444) in Voorhout. De man verklaarde zelf te hebben gedronken en slingerend te hebben gefietst met een telefoon in zijn hand, hij wilde muziek uitzoeken.

Het ongeluk gebeurde even na 1.00 uur. Bij de botsing kwamen zowel de fietser als de scooterrijder en een bijrijder ten val. De politie en een ambulance werden vervolgens opgeroepen. Ambulancepersoneel heeft de twee opzittenden van de scooter nagekeken; ze raakten lichtgewond, maar hoefden niet mee naar het ziekenhuis. De fietser bleef ongedeerd. De schade aan de scooter is fors. Door de afhandeling van het ongeluk was het fietspad enige tijd niet toegankelijk. (Brom)fietsers konden via het gras wel passeren.



# Solutions and techniques:

## Convert PDF-document to plain text

- Challenges:
  - Remove pictures, banners, links
  - Convert page to readable format for text analyses
- Solution: multiple packages available for both R and Python, e.g. pdf tools
- Techniques:
  - Multiple tools available, both for R and Python usually based on libpopler
  - Call this Tool from R
  - Set text boundaries per source



"" , "x"

"1", "Home > Regio > Duin- en Bollenstreek

Beschonken fietser met telefoon in hand botst al slingerend op scooterrijder in Voorhout

© Foto's: Flashlight Fotografie

Internetredactie

Zaterdag 27 augustus 2022 om 08:41

VOORHOUT

Een beschonken fietser is in de nacht van vrijdag op zaterdag frontaal in botsing gekomen met een scooter op het fietspad

langs de Leidsevaart (N444) in Voorhout. De man verklaarde zelf te hebben gedronken en slingerend te hebben gefietst met

een telefoon in zijn hand, hij wilde muziek uitzoeken.

; <http://dlvr.it/SXK9JD>; leeg Het ongeluk gebeurde even na 1.00 uur.

Bij de botsing kwamen zowel de fietser als de scooterrijder en een bijrijder ten val. De politie en een ambulance werden vervolgens opgeroepen.

Ambulancepersoneel heeft de twee opzittenden van de scooter nagekeken; ze raakten lichtgewond, maar hoefden niet mee

naar het ziekenhuis. De fietser bleef ongedeerd. De schade aan de scooter is fors.

Door de afhandeling van het ongeluk was het fietspad enige tijd niet toegankelijk. (Brom)fietsters konden via het gras wel

passeren.

© Foto: Flashlight Fotografie

# Solutions and techniques: Analyse text, recognise crashes

- Challenges:
  - Remove remaining non-relevant text
  - Recognise vehicles, persons, actions, road characteristics
- Solutions:
  - Source dependent cleaning
  - Text analyses Techniques
  - Feedback from specialist in-depth team
- Techniques:
  - Python package spaCy
  - Develop criteria for relevance

# Time and Location

Town, village,  
hamlet

streets

| file_nr | date time        | source             | hashtags                      | Location1_town  | Location2_town  | Street1            | Street2         |
|---------|------------------|--------------------|-------------------------------|-----------------|-----------------|--------------------|-----------------|
| 1       | 27/08/2022 08:47 | dlvr.it            |                               | Voorhout        | Voorhout        | -                  | -               |
| 2       | 26/08/2022 13:29 | DBgallery          | Monstersestraat,DenHaag       | Den Haag        | Rotterdam       | Monstersestraat    | Loosduinsekade  |
| 3       | 26/08/2022 12:01 | dlvr.it            |                               | -               | -               | -                  | -               |
| 4       | 26/08/2022 12:45 | TweetDeck          | Loosduinseweg,Monstersestraat | -               | -               | Loosduinseweg      | Monstersestraat |
| 5       | 25/08/2022 19:09 | RTV Katwijk        |                               | Rijnsburg       | Katwijk         | Oranjelaan         | Floralaan       |
| 6       | 24/08/2022 21:00 | dlvr.it            |                               | Rijnsburg       | Rijnsburg       | Oranjelaan         | Floralaan       |
| 7       | 24/08/2022 21:00 | RepobakMTV         |                               | Gouderak        | -               | -                  | -               |
| 8       | 24/08/2022 19:23 | Web App            |                               | Honselersdijk   | -               | -                  | -               |
| 9       | 23/08/2022 15:47 |                    |                               | Leiden          | Leiden          | Driessenstraat     | -               |
| 10      | 23/08/2022 08:53 |                    |                               | Leiden          | Zoeterwoude     | -                  | -               |
| 11      | 20/08/2022 00:41 |                    |                               | Den Haag        | Den Haag        | Monsterseweg       | Lenstraweg      |
| 12      | 19/08/2022 22:02 | nieuwsopbeeld      |                               | -               | -               | -                  | -               |
| 13      | 18/08/2022 19:51 | RepobakMTV         |                               | Gouderak        | -               | -                  | -               |
| 14      | 16/08/2022 10:04 | dlvr.it            |                               | Noordwijkerhout | Noordwijkerhout | Leeweg             | -               |
| 15      | 15/08/2022 00:41 | FS-Poster-App      |                               | Den Haag        | Den Haag        | Kempstraat         | Wolmaransstraat |
| 16      | 14/08/2022 23:25 | Twitter for iPhone | Kempstraat,DenHaag            | -               | -               | Kempstraat         | -               |
| 17      | 14/08/2022 10:44 | TweetDeck          |                               | Rijswijk        | Rijswijk        | Beatrixlaan        | Brederodelaan   |
| 18      | 14/08/2022 10:08 | TweetDeck          | PrinsesBeatrixlaan,Rijswijk   | Rijswijk        | -               | PrinsesBeatrixlaan | -               |

Date and time

# Typical text analyses results

- Road users: cyclist, scooter, police and ambulance
- Persons:
  - drunk cyclist - the man
  - Scooter rider and co-rider
  - ambulance staff
- Location: "along the Leidsevaart ( N444 ) in Voorhout", the bicycle path
- Time: night from Friday to Saturday, after 1 a clock
- Actions:
  - Cycling, zigzagging with a phone in hand, In the collision
  - both the cyclist and the scooter rider and a co-rider fell

The logo for SWOV, consisting of the letters 'SWOV' in a bold, italicized, sans-serif font. The letters are white and positioned on a dark green background. A white curved line is visible behind the letters, extending from the left edge of the frame.

***SWOV***

# Discussion

A white wavy line that starts from the bottom right corner and extends towards the left, ending near the center of the bottom edge of the slide.

# Discussion

- Completeness: Missing Twitter-links for some messages
  - For some news sources 1 in 7 messages without a Tweet (feedback specialist)
- Some Twitter messages have no links
  - Use the original Twitter message and photo's
- Twitter pages can not be converted to pdf-format (unknown browser type)



## JavaScript is niet beschikbaar

We hebben gedetecteerd dat JavaScript in deze browser is uitgeschakeld. Schakel JavaScript in of stap over op een ondersteunde browser om twitter.com te blijven gebruiken. In ons Helpcentrum kan je een lijst bekijken van ondersteunde browsers.

[Helpcentrum](#)



# Discussion

- Use feedback from specialists to optimise results
- Recognize false positives, find characteristics
- Find missing news items
- Be aware of text quality issues for these news sites

# Additional features, further development

- Automatic relevance score (still under development)
- Use the location information to show crashes on a map
  - Requires linking to a street name database like OpenStreetmap/google maps
  - Link geolocation to Google street view or alternatives
  - Link to national road catalogue to derive maximum speed, road characteristics
  - Construct web interface to bundle information

## Further use of this production system

- Monitoring of specific types of crashes nationwide
- Find crashes missing in the police registration
- On demand delivery of news to specialists

**swov**

**Thank  
you!**

**swov.nl**