



Tekst-analyse met born-digital bronnen

DH Clinic Dag 4

Peter Boot

peter.boot@huygens.knaw.nl

7-11-2017



Inhoud

- Wat voor born-digital bronnen zijn er?
- Hoe verzamel je ze?
- Wat voor analyses kun je toepassen?



Bronnen

- Wat is born digital? Bijna alles.
- Hier beperking tot online materiaal
- Online communities, facebook, twitter, discussiesites, weblog en blogging sites, youtube, online tijdschriften en magazines, online winkels, ...
- Van belang voor sociologen, economen, psychologen, historici, letterkundigen ...



Voorbeelden online boekdiscussie

- <https://www.hebban.nl/>
- <https://www.facebook.com/groups/451488498379185/>
- <https://samenlezenisleuker.wordpress.com/>
- <https://twitter.com/SamenGelezen>
- <http://www.ezzulia.nl/>
- <http://bol.com>



Verzamelen van bronnen

- Waarom? Het staat toch online?

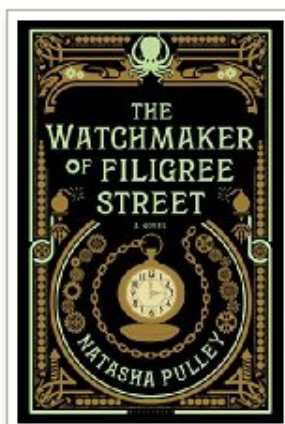


Verzamelen van bronnen

- Waarom? Het staat toch online?
- Maar kan veranderen
- Kan verdwijnen
- Online versie leent zich niet voor systematisch onderzoek
- Niemand anders doet het (itt bij historische teksten)

zondag 11 oktober 2015

Lotsbestemming of vrije wil



Natasha Pulley,
The Watchmaker of Filigree Street (GB 2015)

Roman, 336 pp.

Niet in het Nederlands vertaald.

28 september 2015

★★★★½

Op het eerste gezicht is dit historische fictie: een vol sfeer neergezet Londen van 1883, hoofdstukken die zich afspelen in negentiende-eeuws Japan, één van de eerste women's colleges in Oxford, bomaanslagen door Ierse nationalistes, koortsachtig onderzoek door Scotland Yard. Maar er blijkt na een tijdje iets vreemds aan de hand te zijn: een hooggeplaatste Japanner die zich als horlogemaker in Londen heeft gevestigd doet zijn intrede en maakt daar onwaarschijnlijk geavanceerde uurwerken en mechanische dieren en kan schijnbaar de toekomst zien - of 'zich herinneren' zoals hij zelf zegt. Was deze Keita Mori inderdaad in staat om het ontploffen van de Ierse bom te voorspellen vanwege helderziende vermogens, of is het gewoon omdat hij zelf achter de bom zit? Scotland Yard is uiteraard overtuigd van het laatste, maar de hoofdpersoon en de lezer zijn daar minder zeker van. *The Watchmaker of Filigree Street* is een origineel en intrigerend debuut, dat niet echt een historische roman is, ook niet helemaal fantasy of een thriller, maar iets geheel eigs.

De hoofdpersoon is Thaniel Steepleton, een telegrafist bij een ministerie, die zijn carrière als concertpianist heeft opgegeven om zijn zuster en haar twee zoontjes te onderhouden, nadat haar man is overleden. Dan is er nog Grace Carrow, dochter uit een aristocratische familie, die niets liever wil dan natuurkundige worden, maar haar kansen daarop ziet verdampen als ze er na haar studie in Oxford niet in slaagt een baan als onderzoeker te bemachtigen. Ze heeft weliswaar een flinke erfenis in het verschiep, maar voorwaarde is wel dat ze dan eerst moet trouwen - niet iets waar Grace veel zin in heeft. Thaniel en Grace zijn allebei gefrustreerd door onvervulde verlangens en een leven dat hen weinig keus laat. Dan komt de even bescheiden als raadselachtige Japanse Mr. Mori in hun leven en het lijkt erop dat hij gebeurtenissen en het lot wél naar zijn hand kan zetten, alleen: tot welk doel?

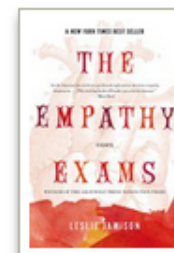
Anna leest...

... voor haar eigen plezier en in volstrekte onafhankelijkheid;
 ... accepteert geen gratis boeken in ruil voor een recensie en doet niet mee aan blogtournees;
 ... is ooit afgestudeerd in de Engelse letterkunde en werkzaam op het gebied van het Nederlands recht.

[\[Meer over mijn blog...\]](#)

[\[Meer over mijn achtergrond...\]](#)

Binnenkort te bespreken



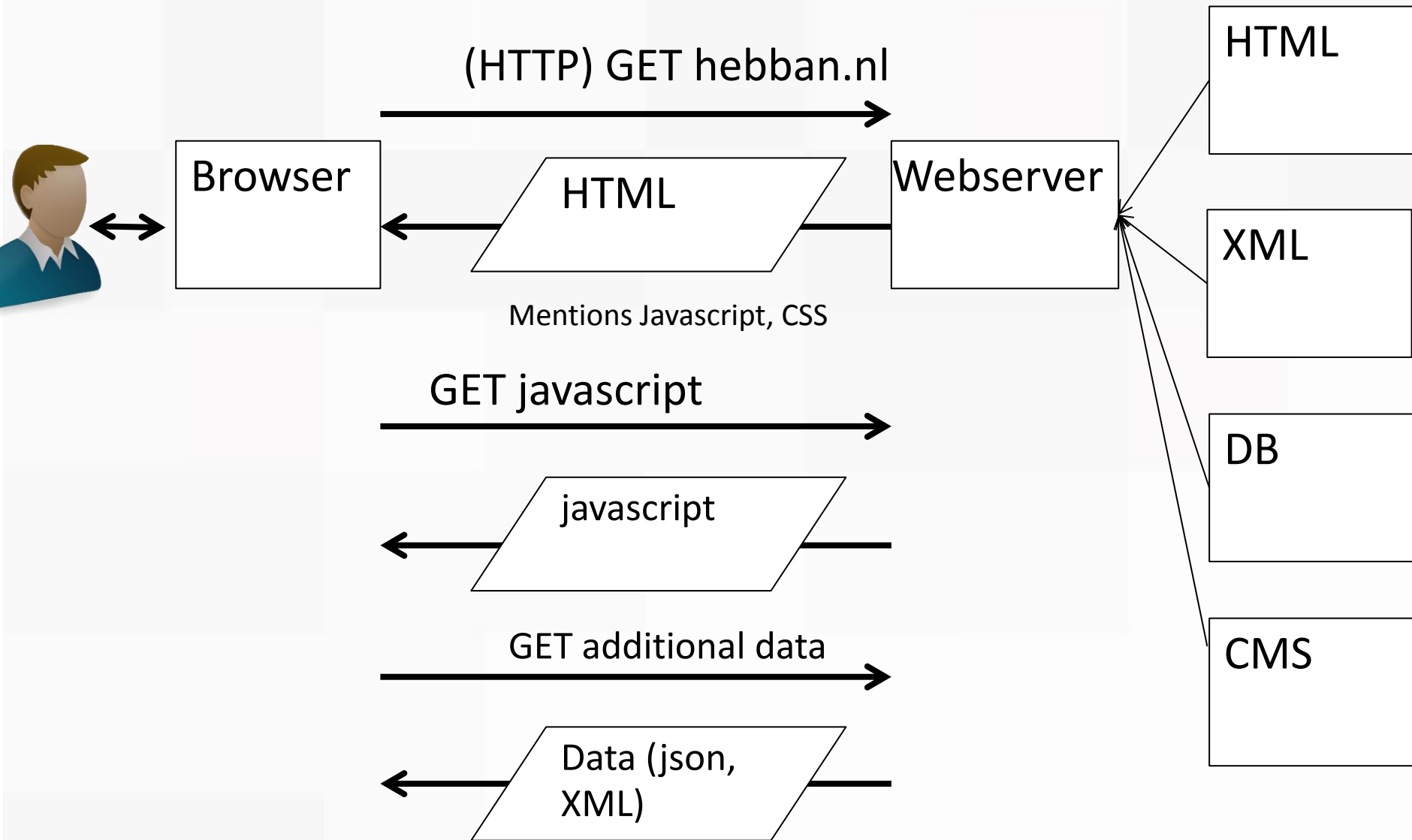
Leslie Jamison,
The Empathy Exams (VS 2014). Essays, 256 pp.



Elena Ferrante,
De geniale vriendin (Italië 2011). Roman, 334 pp.

```
828 <meta content='http://ecx.images-amazon.com/images/P/1620408333.01._SX142_SY224_SCLZZZZZZZ_.jpg' itemprop='image_url' />
829 <meta content='3926416932805183400' itemprop='blogId' />
830 <meta content='2335722628407698148' itemprop='postId' />
831 <a name='2335722628407698148'></a>
832 <h3 class='post-title entry-title' itemprop='name'>
833 Lotsbestemming of vrije wil
834 </h3>
835 <div class='post-header'>
836 <div class='post-header-line-1'></div>
837 </div>
838 <div class='post-body entry-content' id='post-body-2335722628407698148' itemprop='description articleBody'>
839 <div class="separator" style="clear: both; text-align: center;">
840 <a href="http://ecx.images-amazon.com/images/P/1620408333.01._SX142_SY224_SCLZZZZZZZ_.jpg" imageanchor="1" style="clear:
841 <b>Natasha Pulley, &nbsp;</b><br />
842 <b>The Watchmaker of Filigree Street (GB 2015)</b><br />
843 Roman, 336 pp.<br />
844 Niet in het Nederlands vertaald.<br />
845 28 september 2015<br />
846 
847 <br />
848 <br />
849 Op het eerste gezicht is dit historische fictie: een vol sfeer neergezet Londen van 1883, hoofdstukken die zich afspel
850 <a name='more'></a><br />
851 De hoofdpersoon is Thaniel Steepleton, een telegrafist bij een ministerie, die zijn carrière als concertpianist heeft
852 <br />
853 Het fantasy-element, dat wezenlijk is voor het verhaal, wordt acceptabel doordat Pulley het stevig verankerd heeft in
854 <br />
855 Thaniel, de telegrafist en ambtenaar, raakt op een bijzondere manier bevriend met Mr. Mori en maakt vervolgens geheel
856 <br />
857 Is de stelling van dit boek dat bovennatuurlijke verschijnselen echt bestaan en wetenschappelijk aangetoond kan worden
858 <br />
859 Er is wel wat aan te merken op &nbsp;<i>The Watchmaker of Filigree Street</i>. Het is duidelijk dat Pulley een grondige
860 <a href="http://www.mylivesignature.com/" target="_blank"><img src="http://signatures.mylivesignature.com/54487/186/A7
861 <br />
862 <i>PS Aarzel niet om je eigen commentaar toe te voegen. Ik stel het zeer op prijs als mensen de moeite nemen om reacti
863 <div style='clear: both;'></div>
864 </div>
865 <div class='post-footer'>
866 <div class='post-footer-line post-footer-line-1'><span class='post-comment-link'>
867 </span>
868 <span class='post-icons'>
```


What gebeurt er als we browsen?





Opslaan van webpagina's

- Bestaande webarchieven
- Tools voor opslaan 1 of beperkt aantal pagina's
- Tools voor downloaden hele sites / grote aantallen pagina's
- Tools voor downloaden en analyseren van gelijksoortige pagina's
- Tooldatabase Digital Methods Initiative:
<https://wiki.digitalmethods.net/Dmi/ToolDatabase>
 - Twitter Capture and Analysis Toolset (DMI-TCAT)
- Programmeren
 - Simuleer browser
 - Via Application Programming Interface (API)



Bestaande webarchieven

- Internet Archive: <http://archive.org>
- Koninklijke Bibliotheek



Tools voor beperkt aantal pagina's

- Enkele pagina: archive.is
 - <http://archive.is/tffes> archief van <http://govern.cat>
 - <http://archive.is/cyE3m> archief van <http://www.partitdemocrata.cat/>
- Een traject door één of meerdere sites: webrecorder.io
 - <https://webrecorder.io/pboot/catalonië/20171030093906/https://twitter.com/KRLS> archief van <https://twitter.com/KRLS>
- cURL

```

Selecteren Windows PowerShell
D:\Downloads\curl-7.56.1\AMD64>curl peterboot.nl/bibl.html>test.txt
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
           % Done    0         0         0     0      0:00:01 --:--:-- 0:00:01 450k

D:\Downloads\curl-7.56.1\AMD64>type test.txt
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE html
  PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html><head><meta http-equiv="Content-Type" content="text/html; charset=utf-8" /><link href="pb.css" rel="stylesheet" ty
pe="text/css" /><title>PB - Publications </title><script type="text/javascript">

  var _gaq = _gaq || [];
  _gaq.push(['_setAccount', 'UA-27806883-1']);
  _gaq.push(['_trackPageview']);

  (function() {
    var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async = true;
    ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www') + '.google-analytics.com/ga.js';
    var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga, s);
  })();

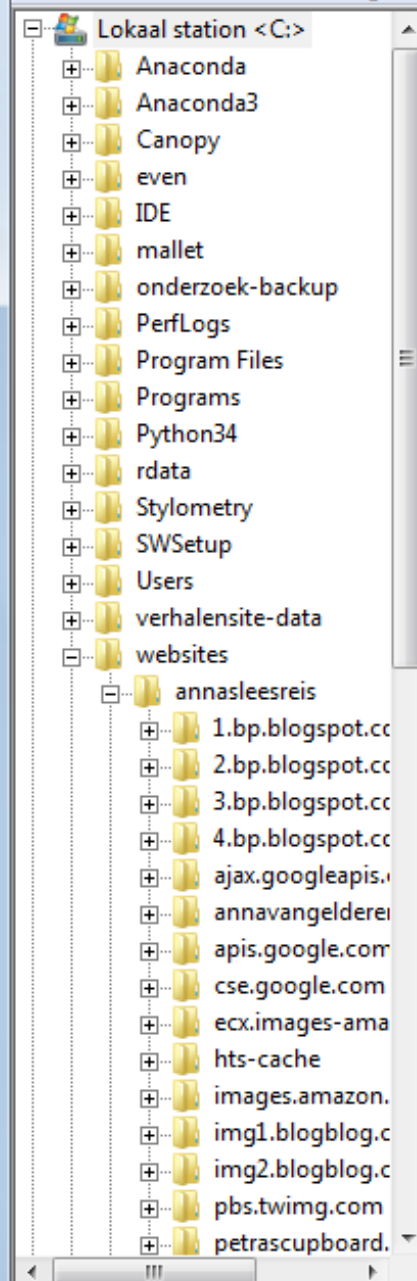
</script></head><body><table><tr><td style="width:150px;vertical-align:top" class="menuclass"><a href="index.html">Home<
/a><br /><a href="phd.html">Thesis</a><br /><a href="bibl.html">Publications </a><br /><a href="archive.html">Archive (D
utch)</a><br /></td><td style="vertical-align:top" class="bodyclass"><div class="bibliography"><div class="titlepage"><d
iv><div><h2 class="title"><a id="bibl"></a>Publicaties - Publications</h2></div></div></div>

    <p>If there is no hyperlink to an electronic version, <a href="mailto:pboot@xs4all.nl" class="ulink">mail me
</a> for a pdf.</p>
    <p>
      <a href="#y2017" class="link">2017</a>
      <a href="#y2016" class="link">2016</a>
      <a href="#y2015" class="link">2015</a>
      <a href="#y2014" class="link">2014</a>
      <a href="#y2013" class="link">2013</a>
      <a href="#y2012" class="link">2012</a>
      <a href="#y2011" class="link">2011</a>
      <a href="#y2010" class="link">2010</a>
      <a href="#y2009" class="link">2009</a>
      <a href="#y2008" class="link">2008</a>
      <a href="#y2007" class="link">2007</a>
      <a href="#y2006" class="link">2006</a>
      <a href="#y2005" class="link">2005</a>
      <a href="#y2004" class="link">2004</a>
      <a href="#y2003" class="link">2003</a>
      <a href="#y2002" class="link">2002</a>
    </p>
</div class="bibliodiv">
  <p><a id="y2017"></a>[2017]</p>
  <div class="bibliomixed"><a id="b2017a"></a><p class="bibliomixed">[b2017a]
```



Ophalen grote aantallen pagina's

- wget
- (win)httrack



- Mirroring Mode -

Enter address(es) in URL box

Action:

Download web site(s)

Web Addresses: (URL)

Add URL...

http://annavangelderen.blogspot.de/

URL list (.txt):

Preferences and mirror options:

Set options...

< Vorige

Volgende >

Cancel

Help

- Favoriete koppelingen
- Dropbox
 - Documenten
 - Computer
 - Meer >>

- Mappen
- websites
 - annasleesreis
 - 1.bp.blogspot.com
 - 2.bp.blogspot.com
 - 3.bp.blogspot.com
 - 4.bp.blogspot.com
 - ajax.googleapis.com
 - annavangelderens.blogspot.de
 - 2002
 - 2008
 - 2009
 - 2010
 - 2011
 - 2012
 - 2013
 - 2014
 - 2015
 - 01
 - 02
 - 04

Naam	Gewijzigd op	Type	Grootte	Labels
lotsbestemming-of-vri...	16-3-2016 16:07	Firefox HTML Doc...	134 kB	
oplichtertjes-tijdens-d...	16-3-2016 16:07	Firefox HTML Doc...	121 kB	
verdwenen-in-slade-h...	16-3-2016 16:07	Firefox HTML Doc...	153 kB	
wat-is-echt-wat-is-wa...	16-3-2016 16:07	Firefox HTML Doc...	140 kB	





Downloaden en analyseren van pagina's

- Goose: verwijderen van 'boilerplate'
- Octoparse: analyseren van de pagina in onderdelen
- Voorbeeld: <https://deleesclubvanalles.nl/book-review/recepten-los-angeles/>

```
2 from goose3 import Goose
3 url = 'https://deleesclubvanalles.nl/book-review/recepten-los-angeles/'
4 g = Goose()
5 article = g.extract(url=url)
6 print(article.title)
7 print(article.cleaned_text)
8 print(article.top_image.src)
```

D:/documents/data/Collegespraatjes/dhclinic/programs/goosetest.py/

UTF-8

Python

Ln: 5 Col: 29

Command Output Notifications Console Syntax Checking Status

Recepten uit Los Angeles

[Recensie] Wie ooit in Los Angeles is geweest zal in dit boek veel herkennen. Met name wat Victor Garnier Astorino in zijn voorwoord zegt over de stad. Los Angeles is een stad met onbegrensde mogelijkheden en veel culturen. Alles kan er en niets is er gek. De mensen zijn echt heel erg aardig en zullen zelden of nooit oordelen als iemand net even anders is dan anders. Wie zijn dromen waar wil maken moet in Los Angeles zijn. Volgens de schrijver van Recepten uit Los Angeles vind je in deze stad enkele van de creatiefste en gezondste gerechten ter wereld. Alhoewel je daar een paar voorbeelden van terug zult vinden in dit boek is 'gezonder' niet de insteek. Want uiteraard treffen we ook de hamburgers, de pancakes en de pizza's die op iedere straathoek verkrijgbaar zijn aan als recept.

Na een inleiding van twee pagina's gaan we direct door naar de recepten verdeeld over zeven hoofdstukken die zich in een bepaald gebied in of rond Los Angeles afspelen. Er is geen informatie over de desbetreffende gebieden. We krijgen uitsluitend recepten en foto's van de gerechten en van de bewoners van deze bijzondere stad.

En omdat er zo weinig informatieve tekst is, is er dus meer ruimte voor de recepten. Het zijn er dus veel. Overheerlijke gerechten die we als typisch Amerikaans kennen zoals hamburgers in verschillende soorten en maten, donuts en pizza's. Maar omdat in Los Angeles alles kan en niets gek is zien we ook een Pizza mac and cheese en een Sandwich met gebakken kip en kruidige aardbeienjam. Het overgrote deel van de recepten is helemaal niet zo traditioneel maar een afspiegeling van de erfenis van de vele culturen die de stad bevolken.

Heerlijke recepten van onder andere Taco's met garnalen, Tortilla met kip en kaas, Banana cream pie, Sandwich met asperges of een Focaccia met asperges en oesterzwammen. We zien een aantal Aziatische recepten en ook Mexicaanse invloeden zijn duidelijk aanwezig. Het boek sluit af met de hoofdstukken Basisrecepten en sauzen, Smaakmakers en als laatste de Bijgerechten waarin o.a. een recept om zelf cornedbeef te maken. Een handig register op ingrediënt maar ook één op recept sluiten het boek af. Zo kun je altijd alles snel terugvinden in dit fijne smulboek.

<https://deleesclubvanalles.nl/wp-content/uploads/2017/10/9200000079579026.jpg>

Define Fields: The data you choose to extract is shown in these fields.

Field names	Data Extracted	Extract Data Type	Delete
TT title	Recepten uit Los Angeles	Extract Text	
TT subtitle	Typisch Amerikaans	Extract Text	
TT reviewauthor	Culinette, Jeanette	Extract Text	
TT bookauthor	Garnier Astorino, Victor	Extract Text	
TT Field5	[Recensie] Wie ooit in Los Angeles is geweest zal in dit boek ...	Extract Text	
TT Field6	Typisch Amerikaans[Recensie] Wie ooit in Los Angeles is gew...	Extract Text	
TT Field7	Hamburgers	Extract Text	

Add Pre-defined Fields

Total number of fields: 12

Recepten uit Los Angeles | De Leesclub van Alles

https://deleesclubvanalles.nl/book-review/recepten-los-angeles/

☐ Block Pop-up



“ I have always imagined that Paradise will be a kind of library ” - Jorge Luis Borges



zoek artikelen, auteurs of recensenten

Zoeken

De Leesclub van Alles

[Home](#) [Over ons](#) [Auteurs](#) [Boeken](#) [DLVALive](#) [Nieuwsbrief](#) [Agenda](#) [Steun DLVA](#) [Adverteren](#) [Contact](#)

[Home](#) / [Artikelen](#) / [Recepten uit Los Angeles](#)

Like 0 [Share](#)

28-10-2017, Artikel door: [Culinette, Jeanette](#)

Auteur(s) boek: [Garnier Astorino, Victor](#)

Recepten uit Los Angeles

Typisch Amerikaans

[Recensie] Wie ooit in Los Angeles is geweest zal in dit boek veel herkennen. Met name wat Victor Garnier Astorino in zijn voorwoord zegt over de stad. Los Angeles is een stad met onbeperkte mogelijkheden en



Postbrieven

Bestand Start Invoegen Pagina-indeling Formules Gegevens Controleren Beeld Acrobat

Van Access Van web Van tekst Van andere bronnen Best. verbind. Alles vernw. Verbindingen Verbindingen

Verbinden Eigenschappen Kopp. bewerken

Sorteren Filter Wissen Opnieuw toep. Geavanceerd

Tekst naar Duplicaten Gegevensvalidatie Samenvoegen Wat-als-analyse Groeperen Degroep

Ext. geg. ophalen Sorteren en filteren Hulpmiddelen voor gegevens

A1	title				
A	B	C	D	E	
1	title	subtitle	reviewauthor	bookauthor	Field7
2	Recepten uit Los Angeles	Typisch Amerikaans	Culinette, Jeanette	Garnier Astorino, Victor	Hamburgers
3	Onvindbaar	Familiegeheimen	Liefaard, Joop	McHugh, Laura	Engelstalige Thriller (vertaald)
4	IJsvaarders van de Noordzeetoe	Mooi voorbeeld van Ziltpun	Klein Haneveld, Johan	Boekestein, Jaap	Nederlandstalige SF
5	Literalicious	Literaire meesterwerken op tafel geze	Stino, Marjo	Groote, Koen de	Kookboek
6	Is liefde lastig?		Verhasselt, Sarah	Pelseneer, Reine de	Nederlandstalige Roman
7	Ronny en het grote spinnenweb	Ik kan het ook	Stino, Marjo	Holslag, Anthonie	Kinderboek (1-4)
8	Thuis in de keuken	Thuis in de keuken gaat over gemak e	Louter, Marco	Withagen, Ans	Kookboek
9	Oude meesters	We mogen best ambitieus zijn en iets	Verplancke, Marnix	Vries, Joost de	Nederlandstalige Roman
10	Stille grond	Conflict in lawaaiig, druk en overvol R	Trouwborst, Jannie	Hassel, Sanneke van	Nederlandstalige Roman
11	Eva	Wanneer de wereld zich niet aan de regels houdt		Daniëls, Jessica	Autisme
12	Hoe alles moest beginnen	Associatieve roman over de kleine dra	Verplancke, Marnix	Verbogt, Thomas	Nederlandstalige Roman
13	De Jaren	"Ze was volmaakt gelukkig. De tijd sto	Lierop, Tea van	Woolf, Virginia	Engelstalige Roman (vertaald)
14	Moeras	Marathon in wervelend proza	Lierop, Tea van	Tolz, Steve	Engelstalige Roman (vertaald)
15	Het lichtatelier		Verhasselt, Sarah	Kessels, Marie	Nederlandstalige Roman
16	Lang, lang geleden	Bakstenen huis tussen de hutten	Trouwborst, Jannie	Jolley, Mike	Archeologie
17	De noordse mythen	De nazaten van Troje	Renaud, Sonja	Larrington, Carolyne	Mythen
18	Holodomor	Een toegedekte en ontkende genocid	Stino, Marjo	Gomme, Patti	Genocide
19	Eigenwijs vaderland	Populair nationalisme in negentiende	Oddens, Joris	Petterson, Anne	19e Eeuw
20	Romeinse legioenen Bert Overbe	Standaardwerk over 's werelds beste	Overbeek, Bert	Goldsworthy, Adrian	Militaire geschiedenis
21	De ondergang van het avondland	Iedere cultuur loopt ten einde	Heumakers, Arnold	Spengler, Oswald	Cultuurkritiek
22	Eeuwen van duisternis	IS avant la lettre	Verplancke, Marnix	Nixey, Catherine	Christendom
23	Pragmatisme	Hotel Diversiteit	Bossche, Marc Van den	James, William	Diversiteit
24	God bestaat niet en Jezus is zijn z	Christus voor atheïsten	Junge, Annabel	Hendrikse, Klaas	Atheïsme
25	De flexikok	Soms vegetarisch, soms veganistisch e	Esther	Brabanter, Veerle de	Kookboek
26	De lege schommel	Van Syrië naar Brussel	Trouwborst, Jannie	Elias, Bettie	Kinderboek (10-12 jaar)
27	De mooiste kindergedichten	'Dit is de heks van Sier-kon-fleks, ze w	Robben, Katelijne	Schmidt, Annie M. G.	Kinderboek (4-6 jaar)
28	De blauwe kamer		Verhasselt, Sarah	Simenon, Georges	Franstalige Thriller (vertaald)
29	Tosti & Toast	Brood roosteren	Esther	Green, Fenn	Kookboek
30	Dossier 64	Onopgeloste misdaden in Denemarke	Liefaard, Joop	Adler-Olsen, Jussi	Deense Thriller (vertaald)
31	Stille grond	De modderkluiten krijgen	Cuijpers, Anke	Hassel, Sanneke van	Nederlandstalige Roman
32	Broeder, schrijf toch eens!	Zeeuwse kronieken	Trouwborst, Jannie	Spruit, Rinus	Nederlangstalige Roman
33	De duistere winkel	De dromennotulist	Reinewald, Chris	Perec, Georges	Dromen
34	Dagboek van een boekverkoper	Zou ik boekverkoper de métier willen zijn?		Bythell, Shaun	Boekenverkoop



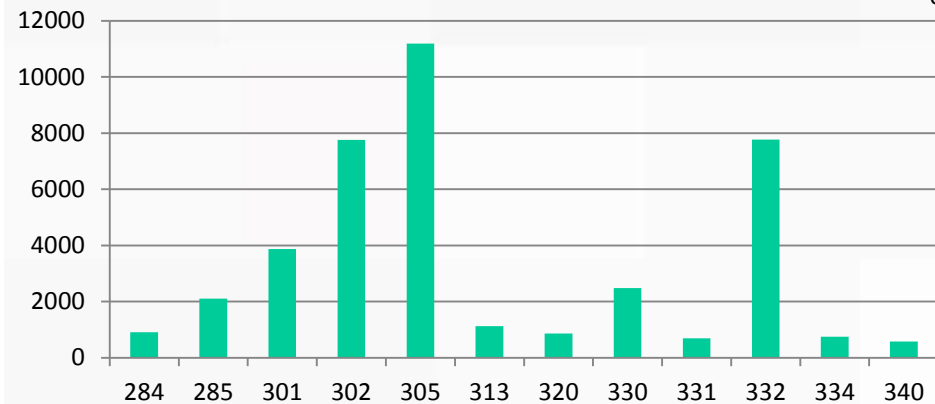
Analyses

- Tellen
- Concordantie (bijvoorbeeld AntConc)
- LIWC
- Stylometrische analyse
- Natural language processing
- Sentiment detectie
- Topic modelling

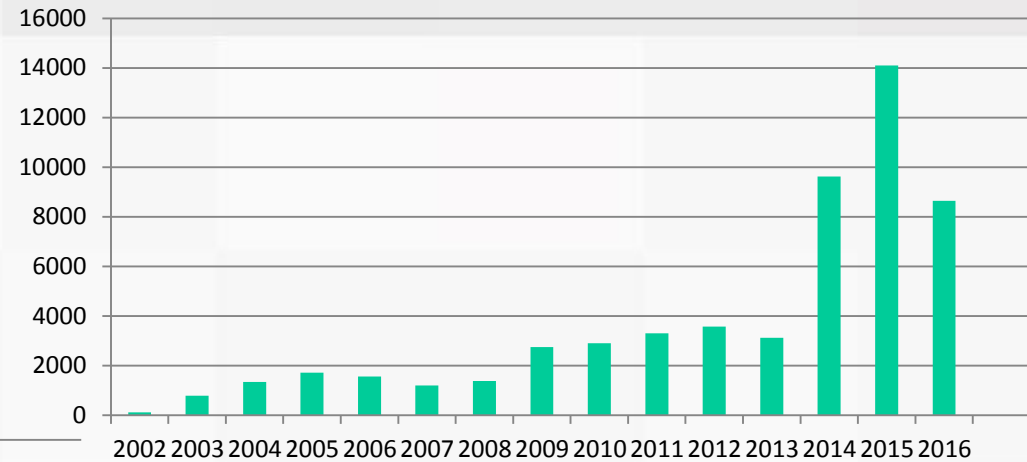


Tellen

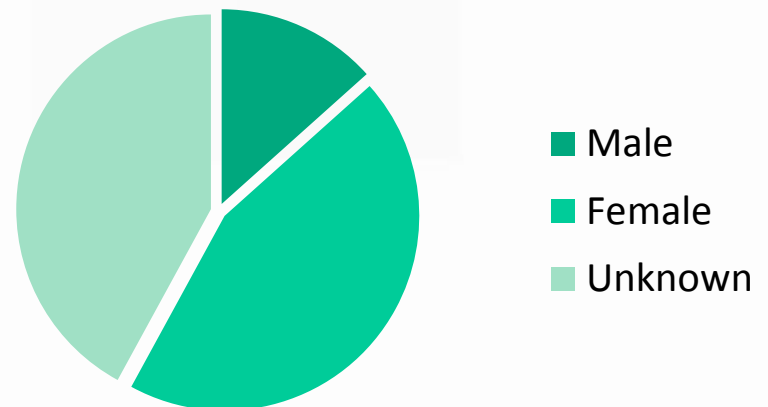
Aantal Hebban reviews naar genre (>500)



Aantal Hebban reviews per jaar



Aantal Hebban reviews naar geslacht



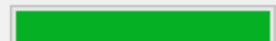
Corpus Files

321.txt
361.txt
503.txt
526.txt
653.txt
654.txt
676.txt
710.txt
719.txt
720.txt
725.txt
731.txt
736.txt
737.txt
815.txt
816.txt
817.txt
818.txt
819.txt
820.txt
821.txt
822.txt
823.txt
824.txt
835.txt
836.txt
837.txt
838.txt
839.txt
840.txt
841.txt

Total No.

2273

Files Processed



Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 91

Hit	KWIC	File
1	lerne psychologie van Sigmund Freud. Het hoeft dan ook niet te verbazen	2059.txt
2	eft dan ook niet te verbazen dat Freud zijn essay Das Unheimliche grotendeels	2059.txt
3	terpretatieve tour de force weet Freud dit verhaal in verband te brengen met	2059.txt
4	steeds uiterst relevant. Volgens Freud wordt iets unheimlich als de grens tussen	2059.txt
5	t Duitse idealisme en Peter Gay, Freud. Pionier van het moderne leven.	3174.txt
6	t Duitse idealisme en Peter Gay, Freud. Pionier van het moderne leven.	3175.txt
7	n. Minder geslaagd zijn de door Freud geïnspireerde speculaties over Da Vinci's	3345.txt
8	n. Ondertussen maakt Sigmund Freud zich zorgen over de dromen van allerlei	3369.txt
9	en op Schopenhauer, Nietzsche, Freud, Wittgenstein, Canetti en nog vele andere	3404.txt
10	t Duitse idealisme en Peter Gay, Freud. Pionier van het moderne leven.	3455.txt
11	t Duitse idealisme en Peter Gay, Freud. Pionier van het moderne leven.	3505.txt
12	n C.B. Vaandrager en Peter Gay, Freud. Pionier van het moderne leven.	3546.txt
13	n bij dit hybride genre. Dit keer: Freud. Pionier van het moderne leven van Peter	3589.txt
14	die zich serieus bezighoudt met Freud en zijn theoretische erfenis komt er een	3589.txt
15	dig beeld van de man Sigmund Freud en zijn werk. Freud. Pionier van het	3589.txt
16	an Sigmund Freud en zijn werk. Freud. Pionier van het moderne leven is simpel	3589.txt
17	oriteit als het gaat om Sigmund Freud, wat blijkt uit zijn vele studies over	3589.txt
18	ndlegger van de psychoanalyse: Freud: A Life for Our Time (1988), Freud for	3589.txt

Search Term ☒ Words ☐ Case ☐ Regex

Search Window Size

freud

Advanced

50

Start

Stop

Sort

Kwic Sort

☒ Level 1 1R ☒ Level 2 2R ☒ Level 3 3R

Clone Results



Concordantie met lijst woorden

niemand meer serieus, dus die kunnen we overslaan.
Maria's deugd: O Hemels Deur, o Morgenster!
s die hij veroordeelde tot de vergetelheid. [uit '
onder meer Doeschka Meijsing, Nicolaas Matsier en
nd, vol stille hartstocht. Een verhaal dat wijlen
e eeuw, sinds de hoogtijdagen van
s echt opengebroken door auteurs als Anna Blaman,
.â€™™ Van Mierlo is hiermee zeker niet de eerste.
van dit boek is overigens een uitspraak van
weet bijvoorbeeld dat ook de Nederlandse dichter
anrader. In dezelfde styl (hoewel minder goed dan
door de katholieken over Europa was uitgeroepen.
, vol als hij is van Ter Braak en
de associaties met Menno ter Braak en Eddy
nen) in vroeg stadium, in plaats van verwennerij.
de lijn van Arnon Grunberg en de vroege
ij daarin zeldzaam venijnig kan zijn. Niet alleen
faveure van De ontdekking van de hemel van

[Hella Haasse](#) op haar beurt propt zoveel nietszegge
[P.C. Boutens](#), uit: Beatrijs (1908)
[Nescio](#)'] Niet dat ik als middelbare scholier nog H
[Frans Kellendonk](#) behoorde Kooiman tot de
[Simon Carmiggelt](#) zonder meer graag geschreven zou
[W.F. Hermans](#) en (toen nog) Gerard van het
[Hugo Claus](#), Jan Wolkers, Harry Mulisch en W.F.
[Harry Mulisch](#), Jan Blokker en Henk Hofland (allen
[Multatuli](#) over een bepaald soort taalgebruik]. Voo
[Lucebert](#) meer exemplaren van dit boek in huis had
[Erasmus](#)) gaat Grunberg verder met het te kakken
[Erasmus](#) noemde Süleyman een barbaar van
[Du Perron](#) en Hermans en Mulisch. Wie in zo'
[Du Perron](#), hun Comité van Waakzaamheid tegen
[Erasmus](#) vergeleek een kind met een kneedbaar stuk
[Harry Mulisch](#). Het is een bewonderenswaardig strev
[A.F.Th. van der Heijden](#) kan in dit
[Harry Mulisch](#). Iris Pronk maakte zich er vorig jaa



Concordantie met reguliere expressies

\b(nieuwe? | nieuwste | recente? | beide? | overige | eerste? | eerdere? | andere? | latere? | laatste? | tweede | derde | minste | beste)\s+(boek(en)? | werk(en)? | romans? | (gedichten)?bundels? | titels?)

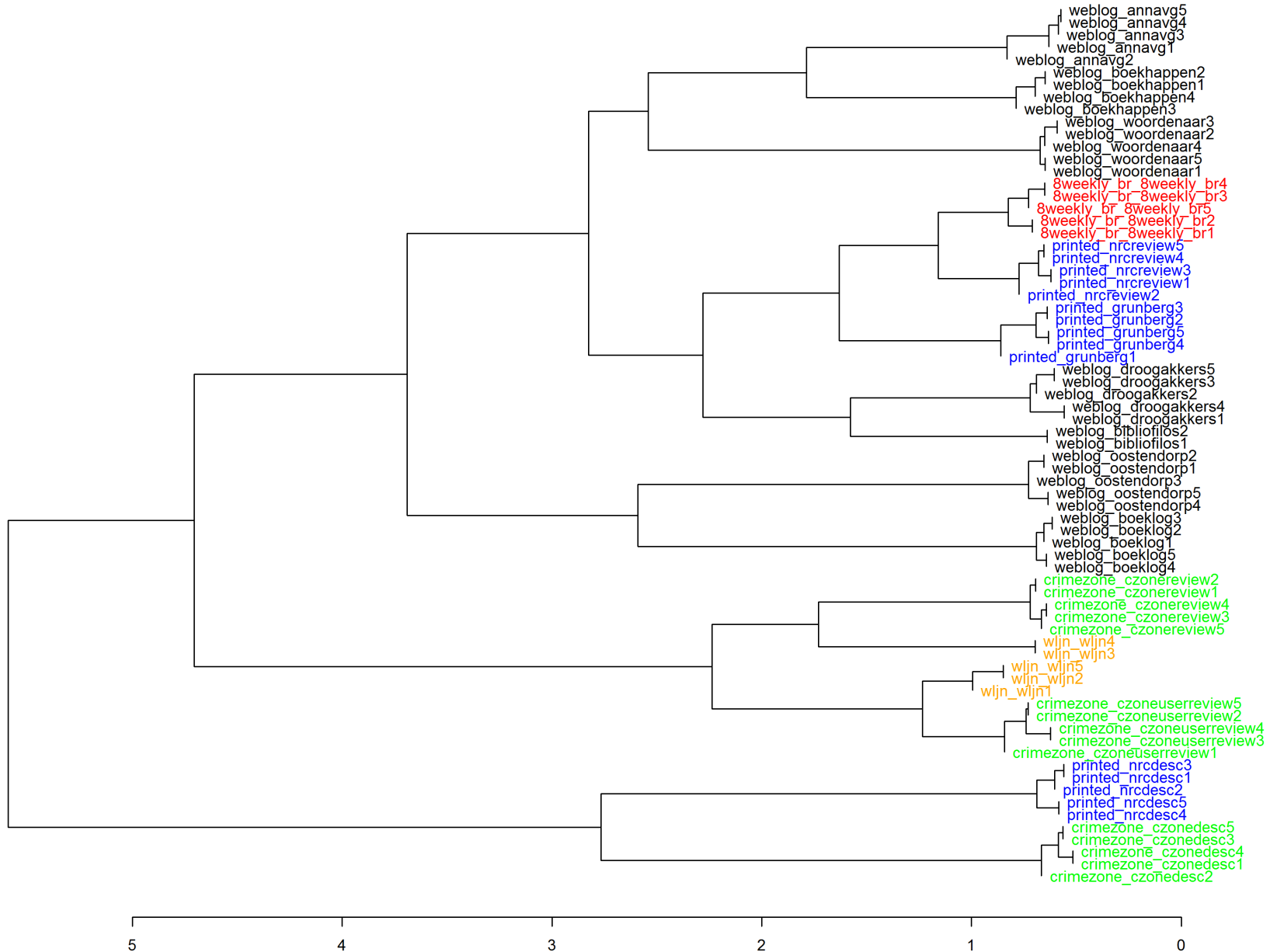
kenmerkt ook de rest van 1000 jaar Amsterdam, het [eerste boek](#) over de ruimtelijke ontwikkeling van Amsterdam, v
 ysteem konden verstoren. Het jaartal 1493 van het [laatste boek](#) van Charles Mann is dan ook wat willekeurig. Mann
 se indianen in de periode vā'ā'r Columbus. Manns [nieuwe boek](#) daarentegen is oeverloos. Het staat in het teken
 n 22-11-1963 slaagt veel beter dan veel van Kings [recente boeken](#) doordat hij de lezer een enorme gemeenschap aan
 jn bijna overbodig gewordenâ In Dave Eggersâ [nieuwe roman](#) heeft een man de taak een afwezige koning hologra
 oman van Arnon Grunberg. Je zou kunnen zeggen dat [beide boeken](#) voortkomen uit modern oriëntalisme. Nog steeds i
 Caribou Island) thematisch wat had uitgeput, deze [nieuwe roman](#) maakt aan die twijfel resoluut een einde. Geen A
 Maar er schuilt ook iets âdemonischâ in zijn [nieuwe bundel](#). Wat is een kerfstok ook alweer? Het is een stok
 ldregistratie. Ingmar Heytze gebruikt het in zijn [nieuwe bundel](#) Ademhalen onder de maan in een nieuwe betekenis.
 rd in 1990 stond. Vijftien jaar later kwam er een [nieuwe bundel](#), en nu is er dus weer een. Strange inleiding Als
 n, dommelenâ. Zoân joch levert nou eenmaal de [beste boeken](#) op. Ook nu weer. Jasper uit Al dat heerlijke verd
 n waarmee ook Anna Funder te maken kreeg in haar [eerste roman](#) All that I am, een dramatisch verslag van een gro
 t
 Alle duivels zijn hier, Ponsâ [tweede roman](#) na wat dichtbundels en journalistiek w
 aar bij herlezing blijkt wat Verhaar zelf in zijn [tweede bundel](#) Eeuwig tweede (1993) verwoordde: âEen metafoor
 â. Ontluistering is het trefwoord. Ook in zijn [latere bundels](#). Het meest persoonlijk in de cyclus âNawakkerâ
 Theo Verhaar zich tijdens het schrijven van zijn [laatste bundel](#), Valscherm voor Erasmus (2000), bewust was van z
 en fraaie afronding krijgen. Hoe zit het met zijn [nieuwe roman](#)? In Alles â de titel is pretentieuzer dan de in
 verhalen, muziek en beeldende kunst denken aan de [laatste romans](#) van K. Schippers (De hoedenwinkel, 2008) en D. H
 er.â Alles over de liefde heet Appignanesiâs [laatste boek](#) dan ook. Het is een ambitieuze, prettig leesbare
 esi over Anna Karenina en Madame Bovary, âalsof [beide romans](#) meteen al willen aangeven dat overspelige hartsto
 n verbazing in los te laten. Er zijn echoâs van [eerdere bundels](#), maar die echoâs passen wonderwel in de samenha
 vers die het hun lezers moeilijk maken. Voor haar [nieuwe roman](#) Als niet dan zou heeft Ali Smith bewust gekozen v
 beste, maar herhaaldelijk ook, en vooral in zijn [laatste bundels](#) steeds vaker, verkent hij expliciet zijn poëti
 . Opvallend dikwijls gebeurt dat â vooral in de [eerste bundels](#), maar later opnieuw â in drie kwatrijnen, al da
 weest. Zes verzen kregen in de verzamelbundel een [andere titel](#) dan in hun oorspronkelijke uitgave. In een daarva
 lk worden afgesloten met een âEpiloogâ, en de [laatste bundel](#) met een âEnvoiâ. In de twee tussenbundels ont
 r Men verwachtte Rushdie-achtige taferelen bij de [nieuwe roman](#) van Mohammed Hanif. Dat viel mee, al klaagt hij z
 geschuwd. âAlice in Sharia-land,â zo werd de [nieuwe roman](#), Onze lieve vrouw van Alice Bhatti, van Mohammed



Stylometrische analyse

- Detectie van auteur (of geslacht, genre, ...)
- Gebaseerd op vergelijking frequenties van hoogfrequente elementen in taal
 - Functiewoorden
 - Letter n-grams
- Bereken afstand tussen teksten gebaseerd op verschillen tussen frequenties
- Cluster teksten gebaseerd op die afstand
- Teksten van dezelfde auteur clusteren vaak
- Populair R package van Eder en Rybicki

Book Response stijlen Cluster Analysis





Natural Language Processing

- Taken bijvoorbeeld:
 - Part of speech tagging
 - Ontleden
 - Word sense disambiguation
 - Co-reference resolution
- Online tools: <https://webservices-lst.science.ru.nl/>



These webservices for Language & Speech Technology are offered by the research group **Language Machines** and the **Centre for Language and Speech Technology**, both part of the Centre of Language Studies of **Radboud University Nijmegen**. We work in close collaboration with the **Tilburg Centre for Creative Computing (TiCC)**, **Tilburg University**, and will include web-interfaces to various relevant tools developed there.

Here we make available various open-source computational linguistic tools for the web, through the software [CLAM](#).

Many of these webservices support input and/or output in the [FoLiA](#) XML format, some of them will be explicitly designed for Dutch only.

You will need an account to access the tools. An account can be obtained free of charge for small-scale non-commercial use, so feel free to register. The webservices can be accessed by both human end-users, offering an in-browser interface, as well as by client software. [Apply for an account](#) here, staff members may need to review your application and activate your account. If you already have an account but lost your username or password, you can [reset your password](#) here.

Service Name

Description

Ucto Tokeniser

Ucto is a unicode-compliant tokeniser. It takes input in the form of one or more untokenised texts, and subsequently tokenises them. Several languages are supported, but the software is extensible to other languages.

WEBSERVICE: <https://webservices-lst.science.ru.nl/ucto/>

REST API SPECIFICATION: <https://webservices-lst.science.ru.nl/ucto/info/>

HOMEPAGE: <https://languagemachines.github.io/ucto/>

Frog

Frog is a suite containing a tokeniser, PoS-tagger, lemmatiser, morphological analyser, and dependency parser for Dutch, developed at Tilburg University. It is the successor of Tadpole.

WEBSERVICE: <https://webservices-lst.science.ru.nl/frog/>

REST API SPECIFICATION: <https://webservices-lst.science.ru.nl/frog/info/>

HOMEPAGE: <https://languagemachines.github.io/frog/>

Valkuil

Valkuil is a Dutch spelling correction system.



Voorbeeld: Frog

Voorbeeldtekst uit 8Weekly:

De afgelopen paar weken zag je ze overal, maar nu je ze nodig hebt, blijven ze thuis. De politici. Zwevende kiezers die nog steeds niet weten wat ze willen, moeten zelf op zoek naar antwoorden. Gelukkig is in de krant, in de Soundmixshow en het muziektijdschrift Oor nog van alles te lezen waardoor je je keuze kunt bepalen. De politiek is overal, dus kon 8WEEKLY niet achter blijven.

8 vragen, variërend van 'hoe vaak komt u nog in de kerk?' tot 'Verwacht u een linkse of een rechtse meerderheid in de Tweede Kamer na de verkiezingen?' gingen naar de lijsttrekkers van de tien belangrijkste partijen van dit moment. SP, CDA en Lijst Pim Fortuyn gaven helaas geen enkele reactie op onze vragen, terwijl GroenLinks, Leefbaar Nederland en de secretaresse van Ad Melkert vriendelijk bedankten voor de eer.

<div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> <div></div> </div> <div> <div>Bestand</div> <div>Start</div> <div>Invoegen</div> <div>Pagina-indeling</div> <div>Formules</div> <div>Gegevens</div> <div>Controleren</div> <div>Beeld</div> <div>Acrobat</div> </div>										
<div> <div>A1</div> <div></div> <div>f_x</div> <div>1</div> </div>										
	A	B	C	D	E	F	G	H	I	J
1	1	De	de	[de]	LID(bep,stan,rest)	0.877577	O	B-NP	3	det
2	2	afgelopen	aflopen	[af][ge][loop][en]	WW(vd,prenom,zonder)	0.952381	O	I-NP	3	mod
3	3	paar	paar	[paar]	N(soort,ev,basis,onz,stan)	0.958015	O	I-NP	5	su
4	4	weken	week	[week][en]	N(soort,mv,basis)	0.992995	O	I-NP	3	mod
5	5	zag	zien	[zie]	WW(pv,verl,ev)	0.998897	O	B-VP	0	ROOT
6	6	je	je	[je]	VNW(pers,pron,nomin,red,	0.993967	O	B-NP	5	su
7	7	ze	ze	[ze]	VNW(pers,pron,stan,red,3,r	0.820513	O	B-NP	5	obj1
8	8	overal	overal	[overal]	VNW(onbep,adv-pron,obl,v	0.907407	O	B-ADVP	5	mod
9	9	,	,	[,]	LET()	1.000.000	O	O	8	punct
10	10	maar	maar	[maar]	VG(neven)	0.903981	O	B-CONJP	0	ROOT
11	11	nu	nu	[nu]	BW()	0.875000	O	B-SBAR	15	mod
12	12	je	je	[je]	VNW(pers,pron,nomin,red,	0.993967	O	B-NP	15	su
13	13	ze	ze	[ze]	VNW(pers,pron,stan,red,3,r	0.848485	O	B-NP	15	su
14	14	nodig	nodig	[nodig]	ADJ(vrij,basis,zonder)	0.993805	O	B-ADJP	15	predc
15	15	hebt	hebben	[heb][t]	WW(pv,tgw,met-t)	0.998615	O	B-VP	10	cnj
16	16	,	,	[,]	LET()	1.000.000	O	O	15	punct
17	17	blijven	blijven	[blijf][en]	WW(pv,tgw,mv)	1.000.000	O	B-VP	0	ROOT
18	18	ze	ze	[ze]	VNW(pers,pron,stan,red,3,r	0.697368	O	B-NP	17	su
19	19	thuis	thuis	[thuis]	BW()	0.992108	O	B-ADVP	17	predm
20	20	.	.	[.]	LET()	1.000.000	O	O	19	punct
21	21	De	de	[de]	LID(bep,stan,rest)	0.975904	O	B-NP	22	det
22	22	politici	politicus	[politicus]	N(soort,mv,basis)	0.999112	O	I-NP	17	su
23	23	.	.	[.]	LET()	1.000.000	O	O	22	punct
24	24	Zwevende	zweven	[zweef][end][e]	WW(od,prenom,met-e)	0.995349	B-LOC	B-NP	25	mod
25	25	kiezers	kiezer	[kies][er][s]	N(soort,mv,basis)	0.999597	O	I-NP	17	su
26	26	die	die	[die]	VNW(betr,pron,stan,vol,pe	0.574101	O	B-SBAR	25	mod



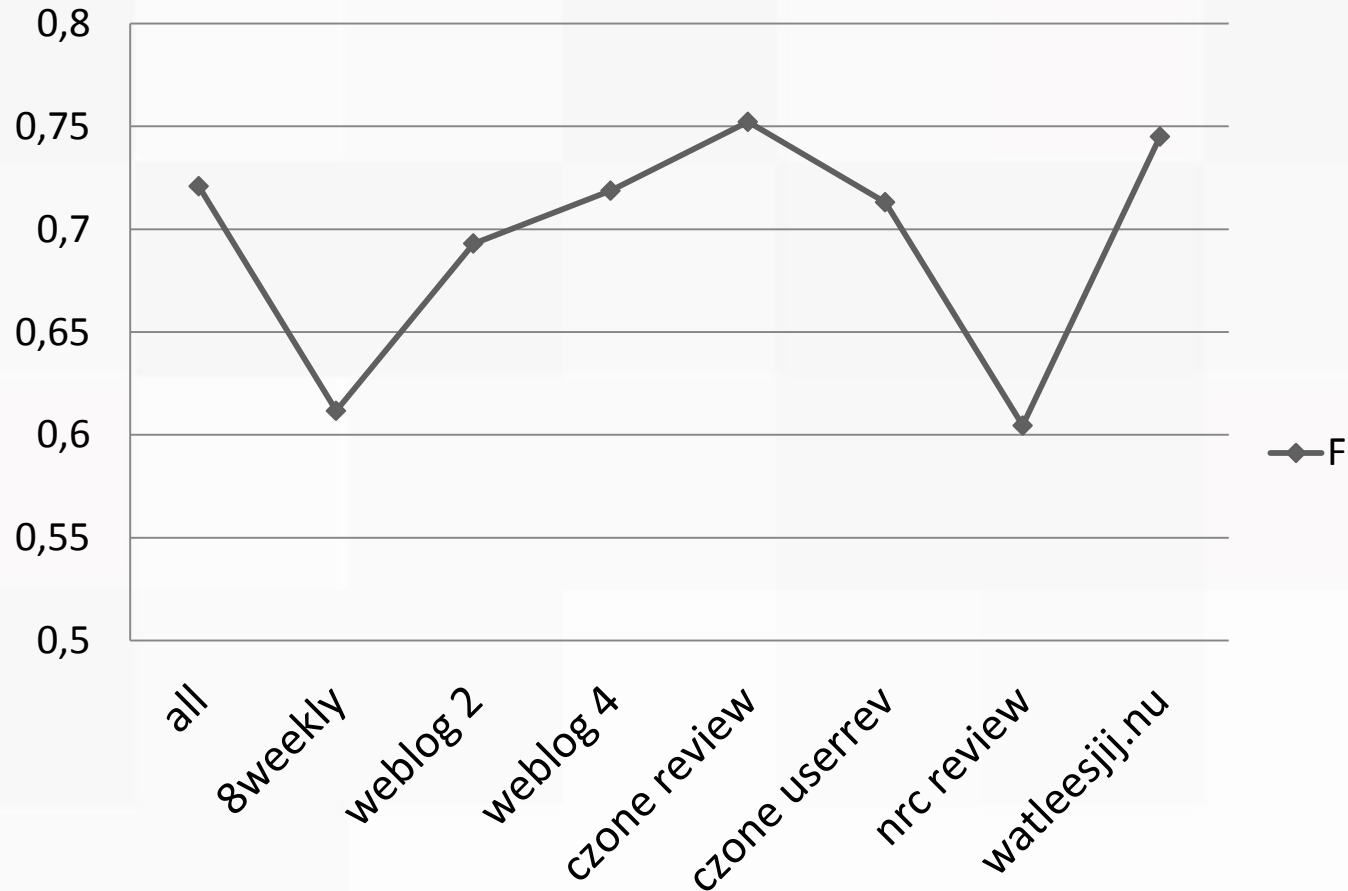
Sentiment detectie

= bepalen polariteit (positieve of negatieve houding)

Één methode hiervoor: gebruik woordenboek van bijvoeglijk naamwoorden met polariteit en intensiteit (Desmedt and Daelemans)

amateuristisch	polarity="-0.4"	subjectivity="0.8"	intensity="1.0"
vreselijk	polarity="-0.4"	subjectivity="0.8"	intensity="2.0"
mooi	polarity="0.7"	subjectivity="1.0"	intensity="1.0"
...			

Bereken polariteit van tekst op basis van bijvoeglijk naamwoorden en bijwoorden



Accuratesse van voorspelde houding
gebaseerd op sentiment analyse van reviews
per collectie



Topic modelling

- Welke thema's komen er in een collectie teksten aan de orde?
- Topic modelling
 - Maakt groepen woorden (=onderwerpen)
 - En kent die toe aan teksten
- Populaire techniek: Latent Dirichlet Allocation
- Hulpmiddelen: Mallet, Gensim



Topic modelling New York Times

music
band
songs
rock
album
jazz
pop
song
singer
night

book
life
novel
story
books
man
stories
love
children
family

art
museum
show
exhibition
artist
artists
paintings
painting
century
works

game
knicks
nets
points
team
season
play
games
night
coach

show
film
television
movie
series
says
life
man
character
know

theater
play
production
show
stage
street
broadway
director
musical
directed

clinton
bush
campaign
gore
political
republican
dole
presidential
senator
house

stock
market
percent
fund
investors
funds
companies
stocks
investment
trading

restaurant
sauce
menu
food
dishes
street
dining
dinner
chicken
served

budget
tax
governor
county
mayor
billion
taxes
plan
legislature
fiscal



Bedankt voor de aandacht!