WETENSCHAP

LITERATUURONDERZOEK
DISTANT READING WINT HET VAN CLOSE READING

de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de
de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de de van
van van van van van van van van van van van van van van van van van van van van van van van van van van van van van van van van van van
van van van van van van van van van van van van van van van van van van van van van van van van van van van van van van van van van van
van van van van van van van het het het het het het het het het het het het het het het het het het het het het het het het het het het het het het
het het het het het het het het het het het het het het het het het het het het het het het het het het het het het het het in in in in in in in in in in in
in in in in in in in in in in in in in in in in in in in in in in in in in in in in in in een een een een een een een een een een een een een een een een een
een een een een een een een een een een een een een een een een een een een een een een een een en en en en en en en en en en en en en en en
en en en en en en en en en en en en en en en en en en en en en dat dat dat dat dat dat dat dat dat dat dat dat dat dat dat dat dat dat dat dat dat
dat dat dat dat dat dat dat dat dat dat dat dat dat dat dat met met met met met met met met met met met met met met met met met met met
met met met met met met met met met zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn zijn te te
te te te te te te te te te te te te te te te te te te te te voor voor voor voor voor voor voor voor voor voor voor voor voor voor voor voor voor voor
voor voor is is is is is is is is is is is is is is is is is is is is niet niet niet niet niet niet niet niet niet niet niet niet niet niet niet niet niet die die die
die die die die die die die die die die die die die die die boeken boeken boeken boeken boeken boeken boeken boeken boeken boeken boeken
boeken boeken boeken boeken boeken boeken boeken op op op op op op op op op op op op op op op op op als als als als als als als als als als
als als als als als als als jockers jockers jockers jockers jockers jockers jockers jockers jockers jockers jockers jockers jockers jockers jockers joc-
kers maar maar maar maar maar maar maar maar maar maar maar maar maar maar hij hij hij hij hij hij hij hij hij hij hij hij hij hij onderzoek onder-
zoek onderzoek onderzoek onderzoek onderzoek onderzoek onderzoek onderzoek onderzoek onderzoek ze ze ze ze ze ze ze ze ze
ze ze woorden woorden woorden woorden woorden woorden woorden woorden woorden woorden dalen dalen dalen dalen dalen dalen dalen
dalen dalen dalen naar naar naar naar naar naar naar naar naar kan kan kan kan kan kan kan kan kan over over over over over over over over nog
nog nog nog nog nog nog je je je je je je je jaar jaar jaar jaar jaar jaar jaar hun hun hun hun hun hun hun er er er er er er er zo zo zo zo zo zo wel
wel wel wel wel wel teksten teksten teksten teksten teksten teksten ook ook ook ook ook ook om om om om om om of of of of of of moretti mo-
retti moretti moretti moretti moretti kunnen kunnen kunnen kunnen kunnen kunnen kestemont kestemont kestemont kestemont keste-
mont heeft heeft heeft heeft heeft heeft hebben hebben hebben hebben hebben hebben doen doen doen doen doen doen computers computers
computers computers computers computers computer computer computer computer computer computer auteurs auteurs auteurs auteurs auteurs
auteurs zou zou zou zou zou zoals zoals zoals zoals zoals zelf zelf zelf zelf zelf wetenschappers wetenschappers wetenschappers wetenschappers
wetenschappers we we we we we was was was was was the the the the the literatuur literatuur literatuur literatuur literatuur dan dan dan dan dan
aantal aantal aantal aantal aantal aan aan aan aan aan werken werken werken werken wat wat wat wat waarin waarin waarin waarin veel veel veel
veel uit uit uit uit toen toen toen toen tien tien tien tien tellen tellen tellen tellen staat staat staat staat software software software software reading
reading reading reading omdat omdat omdat omdat meer meer meer meer maken maken maken maken kenmerken kenmerken kenmerken kenmer-
ken grote grote grote grote groot groot groot groot functiewoorden functiewoorden functiewoorden functiewoorden dus dus dus dus door door
door door dit dit dit dit busa busa busa busa boek boek boek boek auteur auteur auteur auteur analyseren analyseren analyseren analyseren al al al
al zinnen zinnen zinnen zich zich zich worden worden worden werd werd werd waren waren waren vragen vragen vragen vakgebied vakgebied vak-
gebied vaak vaak vaak universiteit universiteit universiteit tot tot tot stellen stellen stellen soort soort soort schrijvers schrijvers schrijvers rowling
rowling rowling project project project positieve positieve positieve ontdekte ontdekte ontdekte nu nu nu nodig nodig nodig negatieve negatieve
negatieve miljoen miljoen miljoen meter meter meter literatuurwetenschappen literatuurwetenschappen literatuurwetenschappen lijst lijst lijst lezen
lezen lezen komen komen komen haar haar haar goed goed goed der der der close close close belangrijk belangrijk belangrijk begint begint begint
andere andere andere alles alles alles zowel zowel zodat zodat zocht zocht zin zin zij zij zegt zegt zeer zeer weten weten werkt werkt werk werk
waarom waarom waar waar voeren voeren vindt vindt vele vele valt valt uitspraken uitspraken twee twee tussen tussen tijd tijd tekens tekens tegelij-
kertijd tegelijkertijd stond stond stanford stanford slechts slechts schrijft schrijft samen samen rol rol robert robert reve reve resultaten resultaten
patronen patronen ontwikkelingen ontwikkelingen onderzoeksresultaten onderzoeksresultaten noveltm noveltm not not net net na na moment mo-
ment moeilijk moeilijk meest meest maakte maakte maakt maakt literatuurwetenschappers literatuurwetenschappers literatuurwetenschapper litera-
tuurwetenschapper literatuurwetenschap literatuurwetenschap literatuuronderzoek literatuuronderzoek literary literary literaire literaire lezing lezing
lengte lengte later later lang lang lading lading kwamen kwamen kleine kleine klein klein karina karina juola juola juist juist jane jane invloedrijkst in-
vloedrijkst ik ik hoopt hoopt honder honderd hoeveelheden hoeveelheden hoeveel hoeveel hoe hoe handvol handvol had had grams grams gevon-
den gevonden geven geven geval geval gespecialiseerd gespecialiseerd geschreven geschreven geschiedenis geschiedenis geen geen gedurende
gedurende gebruiken gebruiken gebruik gebruik galbraith galbraith emotionele emotionele emoties emoties elk elk duidelijk duidelijk decennia de-
cennia daarvoor daarvoor daarom daarom daarmee daarmee cuckoo cuckoo computationele compatutionele combinatie combinatie collega collega
co co calling calling brown brown blijkt blijkt bij bij betrouwbare betrouwbare bestuderen bestuderen beperkt beperkt beautiful beautiful auteursat-
tributie auteursattributie austen austen antwoord antwoord anderen anderen and and analyseerde analyseerde amerikaanse amerikaanse altijd altijd
alle alle zwarte zwaar zullen zover zolang zogenoemde zoekt zoek zijzelf ziet zeventig zette zes zeker zeggen zal woordlengte woordenboekenmaker
woman wist wilden wilde wil wie wetten wetenschapper wetenschappelijk werkwijze werkloosheid wereldoorlog werden wens welke weleens weinig
wegen weg weer waverley walter wakker waarvoor vrouw vroeg vriend vraag vorig vorderingen voorkomt voorkomende voorgedragen voordeel vol-
gen vol voedsel vlak vingerafdruk vinden vind vijftig vier viel verzoek verwijzingen verwerkt verwerken vervolgens vervelend vertraging vertoonde
vertonen vertelt vertellen verraden vermoeden verkoopt verklaringen verifi vergelijkbaar vergeleken vergeleek vereist verband veranderen vaststellen
vastgeplakt vast valide vacancy usb uitte uitleggen uitgevonden typische type tweede transparanter traditionele traditie ton toevalsfactor toe toch to
titel tipje tinten times tijdperk tijdens thomisticus thomas theorie thema testen terug term ter tekst tegemoet technologisch technologie tapes talen
sweet succesboeken subsidie stuk studenten stoppen stijl stick sterk stem steeds statistisch statistici stapel staan spot spoor speelt spaans snel sluier
sinds she sensibility sence seks scott schrijven schrijfstijl schreef routineus rood roms romans roman roberto riddle ren rekende regelmatig referenties
redelijke recept raadselen raadsel quality publiek pseudoniem provocatie promovenda prominent programmeur programmeren programma proef-
schrift problematisch probleem probeert prijs pride prejudice potter positiever ponskaart plot pipet pieken periode patroon patrick past pageturners
pact paar overgeleverde overeenkomsten oud oskam origineelst oprichtte opgevoed opbloeien oog oawetenschappelijke ontwikkelden onszelf on-
misbaar onleesbaarheid ongeveer onderzoekt onderzoeksmethode onderzoeksidee onderzoekers onderzochten onderzocht onderlegde onder om-
slaan ogen nvloed nut nteresseerd normale noemt noch nieuwste nieuws nieuwe nieuw neem nederlandse nederland nauwkeurig nauwelijks natuurlijk
namen naast mogelijk moet model minstens minst minder mike mij middeleeuwse methodologie mensen mens melo meeste meerdere mee matthew
markt marketing marek man makkelijker magneetband magie m lo literair lijken liefst liefde lichten leven leukste leeslijst langste lange landgenoot
lachje lab laatste kwam kunst kritiek kregen kreeg krant korte kort koppen kon komende kom koffie klopt kinderschoenen kilometer kijken kestemonts
keerden karel jullie juiste jonge jodies jodie jagt ja ivanhoe italiaan invloed introduceerde intrede interessante instituut ingezet informatietechnologie
informatielaag inflatie indirect index inderdaad incluis iets identificeren idee huygens humanities hulpjes hulp huizinga houden hoop hoogte hoogle-
raar hoog honger hobby herhaalbaar hem helpen helft hele heldere heilige heette heet heen harry hand hadden grunberg grijs grammaticale gram
grafiek graal gooien goedkoop goede gigabyte geweest getiteld getipt genoemd generaties gemiddelde gemakkelijk gemaakt gelijk geleden gekopi-
eerd geesteswetenschapper geesteswetenschappen gedwongen gedigitaliseerde gedeelde gedaan gebruikt gebied geanalyseerd ge functies franco
fiets fictie extreem euro ervaringen ervan ermee eos enkele enige enig engels ellende elkaar eigenlijk eigen eeuw eerste dvd duizenden duizend duivel
duits drie doet distant discussies dingen digital diep dichter deze dertig deed data dankzij daaruit daarin daarachter cover controleerbaar controle
concludeerde combineren cijfers cd casual canon buitenwereld brittanni britse brengen breken branden boven boektitels boekdrukkunst blote blijven
bleken bleek binnen bijvoorbeeld big bibliotheek bibliografie bevinding betekent betekeniswoorden betekenissen betekenisloze betekenis bestsellers
bestond bestand bestaat bestaande besloot beschouwd beschikking beschikken beschikbaar beruchte bently beloftevolle belangrijkste belangrijker
bekend behulp begrijpen begonnen beeld bedacht be automatiseringstechnologie auteursattributies arnon archief archer aquino antwerpen anoniem
anomalie andersom ander amsterdam amerikaan alleen algemene algemeen alexander alberto afstellen afnam afhankelijk affectie af achterhoofd ach-
tergrond achter acerbi aannamen aankondiging aangeven

Alle woorden van het verhaal hiernaast, van de meest naar de minst gebruikte.

Met computers kun je duizenden boeken tegelijk bestuderen: big data hebben hun intrede gedaan in literatuuronderzoek. ‘Seks verkoopt, te veel seks niet.’

Door **Robert-Jan Friele**

De leesmachine

Een leeslijst van dertig boeken? Zeventig? Een paar honderd? Matthew Jockers vindt het weinig. 1.200 boeken schreef de Amerikaanse literatuurwetenschapper zijn studenten voor op de universiteit van Stanford. 3.500 boeken onderzocht hij om te kijken welke auteurs in de 19de eeuw het invloedrijkst waren. En voor zijn nieuwste onderzoek zal Jockers een miljoen boeken bestuderen. Samen met enkele collega’s, dat wel.

Voor het project NovelTM hebben Jockers en co geen grote bibliotheek nodig, noch vele koppen koffie om wakker te blijven gedurende het lezen. Niet zij, maar computers werken zich de komende zes jaar door de stapel boeken van 1800 tot nu heen, op zoek naar, ja, naar wat eigenlijk? De wetten van de literatuur?

Statistiek heeft zijn intrede in de geesteswetenschappen gedaan. *Digital humanities* heet het nieuwe vakgebied. In de literatuurwetenschappen betekent dat het met software analyseren van grote hoeveelheden gedigitaliseerde teksten en bibliografieën. Het aantal woorden, de lengte van woorden, hoeveel verwijzingen een roman heeft naar de bestaande buitenwereld, of er veel namen in staan: de computer kan alles tellen.

‘En dan vinden we patronen die je met het blote oog niet ziet’, zegt Karina van Dalen-Oskam van het Huygens Instituut voor Nederlandse Geschiedenis en hoogleraar computationele literatuurwetenschap aan de Universiteit van Amsterdam. Net als Jockers doet ze onderzoek naar wat ze zelf noemt ‘de raadselen der literatuur’. Van Dalen hoopt met het project The Riddle of Literary Quality, waarvoor 400 romans worden geanalyseerd, antwoord te geven op de vraag of boeken die als literatuur worden beschouwd, gedeelde kenmerken hebben.

Jockers is geïnteresseerd in de plot van boeken en de rol van emoties. Daarvoor gebruikt hij een lijst van woorden met een negatieve en positieve lading. Met die lijst in het achterhoofd kan de computer analyseren waar in het boek positieve emoties omslaan in negatieve, of andersom. Dat vereist wel enig afstellen van de software. ‘Neem de zin: *She was not a beautiful woman*. Het negatieve *not* en het positieve *beautiful* in één zin: dat is voor de computer moeilijk te begrijpen.’ Met een lachje: ‘Daarom vind ik Jane Austen zo vervelend; die lange zinnen van haar zijn moeilijk te analyseren.’

De computer vindt patronen die je met het blote oog niet ziet

Jockers ontdekte dat de pageturners van Dan Brown meer emotionele pieken en dalen vertonen dan literaire werken. Toen hij een grafiek maakte van de gemiddelde emotionele lading, bleken de boeken van Brown wel een stuk positiever.

Van Dalen en Jockers breken met de traditie waarin hele generaties literatuurwetenschappers – zijzelf incluis – zijn opgevoed: het zeer nauwkeurig lezen en analyseren van een zeer beperkt aantal boeken, ook wel *close reading* genoemd. Niet voor niets introduceerde de Italiaan Franco Moretti een kleine tien jaar geleden, nog voor hij met Jockers het Stanford Literary Lab oprichtte, de term *distant reading*. Het was zowel een provocatie als de aankondiging van een nieuw tijdperk.

‘Het probleem van close reading’, schrijft Moretti, ‘is dat het afhankelijk is van een extreem kleine canon.’ Wetenschappers weten alles van een handvol boeken. Daarmee zeggen ze indirect dat al die andere duizenden werken er niet toe doen. Dat valt niet vol te houden, vindt Moretti. Maar de mens kan in zijn leven slechts een beperkt aantal boeken lezen, dus is ‘een klein pact met de duivel nodig’: computers.

Die maken het mogelijk teksten snel een informatielaag te geven, met daarin het aantal ▸

► woorden, woordlengte, grammaticale functies, betekenis. De wens om dat te doen, bestaat al decennia.

Moretti's landgenoot Roberto Busa wilde vlak na de Tweede Wereldoorlog de werken van Thomas van Aquino bestuderen. Hij besloot voor elk van de 13 miljoen woorden een ponskaart te maken – de enige automatiseringstechnologie die hem op dat moment ter beschikking stond. Busa rekende later zelf uit hoe groot zijn archief zou zijn geweest: 90 meter lang, 1.20 meter hoog, 1 meter diep. En 500 ton zwaar.

Zover kwam het niet: in 1955 werd de magneetband uitgevonden. Toen had Busa slechts 1.800 tapes nodig, van samen 1.500 kilometer lang. Dat was tot 1992, het jaar waarin Busa zijn *Index Thomisticus* op een handvol cdroms kon branden. En in 2003 zette hij alles op één dvd, een bestand van 1,92 gigabyte. Weer tien jaar later past dat op een goedkope usb-stick.

'Veel onderzoeksideeën zijn decennia oud, maar nu beschikken we over de technologie om ze uit te voeren', zegt Mike Kestemont van de Universiteit Antwerpen. Hij is gespecialiseerd in auteursattributie, het identificeren van de schrijvers van anoniem overgeleverde teksten.

Juist de 'betekenisloze' woorden verraden de auteur

Al in 1964 ontwikkelden twee Amerikaanse statistici daarvoor de theorie dat het tellen van 'functiewoorden' als *de*, *het*, *boven*, *onder* en *naast* de auteur kunnen verraden. Dus niet 'betekeniswoorden' als *man*, *vrouw*, *fiets*, *groot*, *klein* of *rood* zijn belangrijk, maar juist de 'betekenisloze' woorden. Het idee daarachter is dat auteurs functiewoorden zo routineus gebruiken dat ze er nauwelijks controle over hebben.

Omdat dit soort onderzoek begint met het statistisch in beeld brengen van meerdere teksten, zijn computers onmisbaar om binnen redelijke tijd tot resultaten te komen. De vorderingen van dit type onderzoek kwamen vorig jaar prominent in het nieuws toen de Amerikaan Patrick Juola op verzoek van *The Times* het boek *The Cuckoo's Calling* analyseerde. Op de cover stond dat de auteur Robert Galbraith heette, maar de Britse krant was getipt dat J.K. Rowling, bekend van Harry Potter, het had geschreven.

Juola vergeleek *The Cuckoo's Calling* met Rowlings eerste boek voor volwassenen, *The Casual Vacancy*, en met drie boeken van andere auteurs. Hij zocht niet naar 'typische J.K. Rowling-woorden', maar juist naar algemene kenmerken, zoals de honderd meest voorkomende woorden. Daaruit concludeerde hij dat de schrijfstijl van Robert Galbraith inderdaad de meeste overeenkomsten vertoonde met die van J.K. Rowling. In Nederland werd vergelijkbaar onderzoek ingezet toen in 2002 het vermoeden bestond dat Marek van der Jagt een pseudoniem was van Arnon Grunberg.

Kestemont is zelf gespecialiseerd in middeleeuwse teksten. Auteursattributie is belangrijk voor dat vakgebied, omdat boeken vaak met de hand werden gekopieerd door hulpjes van de auteurs. Zij wilden nog weleens zinnen veranderen, zodat het belangrijk is te kunnen vaststellen wie wat heeft geschreven.

Sinds kort maakt Kestemont ook gebruik van *n-grams*. De *n* staat voor een aantal tekens dat achter elkaar staat. De computer zoekt bijvoorbeeld hoe vaak de combinatie van de vier tekens '. *To*' voorkomt in een tekst, omdat deze 4-gram zou kunnen aangeven dat een auteur regelmatig zinnen begint met 'Toch'. Dat blijkt een minstens zo betrouwbare onderzoeksmethode als het tellen van functiewoorden en heeft als voordeel dat het ook werkt in talen waarin de computer de functiewoorden niet goed kan tellen, omdat ze aan andere woorden zijn vastgeplakt – zoals het Spaans, waarin *lo* in *dámelo* – 'geef me het', staat voor 'het'.

Kestemont kan alleen nog niet goed uitleggen waarom onderzoek met n-grams werkt. 'We kunnen er betrouwbare uitspraken mee doen over auteurs, maar we weten niet waarom. Het is een soort zwarte magie.'

Veel van het computationele onderzoek naar literatuur begint nog wel met close reading, vertelt Van Dalen. 'Dan valt je iets op, dat je vervolgens met de computer kunt testen. Vaak klopt maar de helft van je aannamen en zo kom je de leukste dingen op het spoor.'

De kunst is de juiste vragen te stellen over de gevonden cijfers en patronen. Zo ontdekte Moretti dat de lengte van boektitels in Groot-Brittannië sterk afnam tussen 1740 en 1850. Op zichzelf geen interessante bevinding, maar, zoals Moretti schrijft: 'Elk patroon is een vingerafdruk van de geschiedenis.' In dit geval: dankzij de ontwikkelingen in de boekdrukkunst in die periode kwamen er meer boeken

WENSDROOM EN NACHTMERRIE

Voor literatuuronderzoekers mag een droom uitkomen nu duizenden boeken digitaal beschikbaar zijn, voor schrijvers is het een nachtmerrie: hun werk is ineens gratis te raadplegen.
Google begon tien jaar geleden miljoenen boeken in te scannen en online beschikbaar te maken (in het geheel als de auteursrechten geëindigd zijn, in fragmenten als dat niet het geval was). De Hathi Trust is een samenwerking van Amerikaanse universiteiten die hun digitale collecties delen. De database van Google bevat inmiddels bijna 30 miljoen boeken, die van de Hathi Trust 13 miljoen; 4,9 miljoen daarvan zijn publiek beschikbaar, de andere alleen voor onderzoek.
Volgens het schrijversgilde is het digitaliseren van al die werken 'zonder toestemming van de rechthebbenden' gebeurd. Het spande rechtszaken aan tegen beide organisaties, verloor, maar ging in hoger beroep. De uitspraak in de Google-zaak wordt begin 2015 verwacht.
In Nederland is in 1999 de Digitale Bibliotheek voor de Nederlandse Letteren (DBNL) opgericht. Die bevat inmiddels elfduizend teksten. In de gebruikersvoorwaarden staat dat die alleen beschikbaar zijn voor onderzoek en niet vrij gedeeld mogen worden. De boeken die Karina van Dalen wilde onderzoeken voor The Riddle of Literary Quality waren niet opgenomen in de DBNL. Ze benaderde daarom de uitgevers en vroeg om digitale kopieën. Ze kreeg ze vrijwel allemaal, maar mag de werken niet delen met andere onderzoekers. 'Ik kan straks alleen mijn onderzoeksresultaten delen, niet mijn onderzoeksmateriaal.' Dat maakt verificatie van haar onderzoek lastig, en is dus wetenschappelijk gezien problematisch. 'Maar als ik mijn tijd aan auteursrechten moet besteden, houd ik geen tijd over voor onderzoek.'

Franz Eybl (1806-1880): Lezend meisje (1850).

op de markt, werd marketing belangrijker en daarmee een korte, heldere titel.

Alexander Bently en Alberto Acerbi onderzochten het verband tussen inflatie en werkloosheid en de 'literaire ellende' in boeken. Zowel in het Engels als in het Duits bleek dat er duidelijk te zijn, maar met een vertraging van ongeveer tien jaar. Zolang doen schrijvers over het (literair) verwerken van hun ervaringen.

Jockers deed onderzoek naar de invloed van schrijvers op collega's. Hij maakte een lijst van 650 kenmerken – zoals affectie, referenties aan voedsel en honger – die hij voor 3.500 boeken analyseerde. Hij ontdekte dat Jane Austen (bekend van *Pride and Prejudice*, *Sence and Sensibility*) en Walter Scott (*Ivanhoe*, *Waverley*) de auteurs waren die het minst beïnvloed zijn door anderen – en dus het origineelst – en tegelijkertijd zelf het invloedrijkst waren: hun stijl en thema's keerden gedurende de langste tijd terug in andere boeken.

Jockers en co kregen voor het project NovelTM 2,3 miljoen euro subsidie. Kestemont is met zijn proefschrift *De stem van de auteur* als eerste geesteswetenschapper voorgedragen voor de Eos Pipet 2014, de prijs voor de meest belovende jonge wetenschapper van het moment.

Tegelijkertijd staat het vakgebied nog in zijn kinderschoenen. De verklaringen voor valide onderzoeksresultaten zijn niet altijd duidelijk, zoals in het geval van Kestemonts auteursattributies. Ook zijn veel onderzoeksresultaten niet herhaalbaar en dus controleerbaar door anderen. Dat heeft ermee te maken dat onderzoekers als Jockers hun eigen software schrijven. Na hun onderzoek gooien ze die weg.

Karina van Dalen: 'We worden gedwongen om onszelf vragen te stellen

Weet de literatuurwetenschap nu wat een goed boek goed maakt?

over de methodologie. Als je onderzoek niet kunt verifiëren, is dat wetenschappelijk problematisch.'

Van Dalen hoopt daarom dat steeds meer wetenschappers hun programma's beschikbaar zullen stellen voor algemeen gebruik. Dat zou het onderzoek transparanter maken en ook gemakkelijker uit te voeren. 'Ik hoop dat over tien jaar veel meer mensen dit soort werk doen, maar wetenschappers kunnen niet alle discussies in de literatuurwetenschappen volgen én op de hoogte zijn van de laatste ontwikkelingen op het gebied van informatietechnologie. Gemakkelijk te gebruiken software zou kunnen helpen.'

Nu speelt de toevalsfactor nog een grote rol bij het opbloeien van de liefde voor literatuuronderzoek met computers. Jockers en Kestemont hebben een achtergrond als traditioneel literatuurwetenschapper, maar hadden als hobby programmeren en begonnen die twee te combineren. Van Dalen zocht als woordenboekenmaker met computers naar woorden in grote hoeveelheden teksten, zodat ze zeker wist alle betekenissen te hebben verwerkt. Ze bedacht dat die werkwijze ook in literatuuronderzoek nuttig was en kreeg hulp van haar vriend, een programmeur.

Met hun combinatie van literatuur en computers lijken wetenschappers als Van Dalen en Jockers tegemoet te komen aan de kritiek die Karel van het Reve in 1978 uitte tijdens zijn roemruchte Huizinga-lezing. In die lezing, getiteld *Het raadsel der onleesbaarheid*, viel hij de literatuurwetenschappers aan op hun in zijn ogen onwetenschappelijke werk. Waarom vertellen jullie mij niet wat goede boeken goed maakt, vroeg hij zich af.

Van Dalen: 'Van het Reve had natuurlijk gelijk, dat is een van de belangrijkste vragen van de literatuurwetenschap.' Met haar onderzoek probeert ze, net als Jockers, dichter bij het antwoord te komen. Het is de heilige graal van de literatuurwetenschappen.

Of is die al gevonden? Een promovenda van Jockers, Jodie Archer, heeft met behulp van 685 kenmerken drieduizend bestsellers vergeleken met tienduizend 'normale' boeken. Zo wil ze uitspraken doen over het recept voor succesboeken. De resultaten zijn nog niet publiek gemaakt, maar een tipje van de sluier kan Jockers wel vast oplichten: 'Seks in fictie verkoopt, maar niet altijd. Er is een *sweet spot* van hoeveel ervan je in een boek moet stoppen.' En *Vijftig tinten grijs* dan? 'Dat is een anomalie, blijkt uit Jodies onderzoek.' ●