



Universiteit  
Leiden

The Netherlands

## Scholarship in interaction: case studies at the intersection of codework and textual scholarship

Zundert, J.J. van

### Citation

Zundert, J. J. van. (2022, September 27). *Scholarship in interaction: case studies at the intersection of codework and textual scholarship*. Retrieved from <https://hdl.handle.net/1887/3464403>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3464403>

**Note:** To cite this publication please use the final published version (if applicable).

## Samenvatting

De hermeneutiek heeft een lange geschiedenis die in Europa onlosmakelijk is verbonden met de tekst- en editiewetenschap. Beiden vinden hun wortels uiteindelijk in de klassieke filosofie en de latere Bijbel-exegese. Zowel het postmodernisme met haar relativistische karakter als de sterke opkomst van digitale technologie – evenzo veranderlijk en onbestendig als invloedrijk – zijn twee ‘disruptieve’ krachten die inwerken op een editiewetenschap die zich bezighoudt met het culturele ‘geheugen’ van de mens en zijn cultuur en die daardoor gepreoccupeerd is met feiten en stabiliteit. Maar ook feiten en stabiliteit zelf zijn op hun best romantische idealen. Geen enkele vorm van editiewetenschap kan ontsnappen aan haar principiële hermeneutische karakter – er is niets dan interpretatie.

Nu een post-digitale werkelijkheid zich aandient, is een grondig begrip van de relatie tussen hermeneutiek, editiewetenschap en digitale technologie een belangrijke en urgente vereiste. Een diep begrip van de werking van softwarecode en hoe deze de editiewetenschap beïnvloedt is nodig, maar er is geen structurele wetenschappelijke agenda om dit probleem te bestuderen. Ten dele is dit het gevolg van de animositeit die in de laatste twee decennia van de twintigste eeuw tussen de ‘kwantitatieven’ en de kwalitatieve benadering van literatuur- en editiewetenschap ontstond en die de ontwikkeling van een serieus discourse omtrent digitale dan wel computationele hermeneutiek binnen de (digitale) geesteswetenschappen in de weg stond. Recentelijk zijn er wel aanwijzingen dat een serieuzer discourse zich kan ontwikkelen (Rockwell, Ramsay, Capurro), maar over het algemeen is er nog steeds een tendens om het potentieel van digitale en computationele technieken te reduceren tot een methodology van determinisme, reductivisme en kwantitatieve empiriek (Johanna Drucker is bij uitstek een voorbeeld van iemand die zulke perspectieven schetst). Het verwijt van reductivisme is snel gemaakt, maar

het feitelijke karakter van digitale geesteswetenschappen is hybride en interdisciplinair, en er is niet a priori sprake van een breuk met de traditionele hermeneutiek. Zelfs al is het zo dat veel computationele methoden in de digitale geesteswetenschappen momenteel neigen naar reductieve en kwantificerende technieken, dan nog zijn hun formalisering en analyses allesbehalve vrij van hermeneutiek. Precies zoals de filologische praktijk fundamenteel niet kan ontsnappen aan interpretatie en daarmee aan hermeneutiek, kunnen ook digitale modelering en kwantificering niet ontsnappen aan de hermeneutiek die gemoeid is met de keuzes en aannames die ten grondslag liggen aan gekozen formalisering, en al helemaal niet aan de hermeneutiek die betrekking heeft op de interpretatie van analyseresultaten. We moeten – refererend aan de woorden van Katherine Hayles – de spanning tussen analyse met behulp van algoritmen en de hermeneutiek van “close reading” dan ook niet overdrijven. Maar het probleem is – zoals onder andere David Berry en Federica Frabetti – hebben aangegeven dat we de precieze innerlijke en impliciete hermeneutische werking van code en computationele technieken eigenlijk niet goed begrijpen. Dat doen we op dit moment met betrekking tot de tekst- en editiewetenschap zeker niet. Dat begrip kunnen we alleen kweken door een serieuze en langdurige omgang met en diepgravend onderzoek naar digitaliteit en software source-code in de context van tekst- en editiewetenschap.

Sommige effecten van de interactie tussen digitale technologie en de tekst- en editiewetenschap bevorderen het ontwikkelen van dat begrip niet. Die effecten kunnen we beter in beeld krijgen en begrijpen door middel van methoden uit de “Science and Technology Studies” (STS). Deze bieden onder meer systematische technieken voor het observeren van en reflecteren op wetenschappelijke processen. Wanneer we dat doen, blijkt bijvoorbeeld dat grafische interfaces met name een fictieve transparantie van modellen en paradigma’s bieden. Een grafische schil verhult een analytisch model even zozeer als dat hij dat model bruikbaar maakt. Een grafische schil is niet het model zelf, maar een visualisatie ervan met zijn eigen beperkingen, interpretaties en gebruiksmogelijkheden. Als grafische interfaces vooral worden vormgegeven als expressies van al bestaande en bekende paradigma’s, modellen en metaforen – denk aan het “bureaublad” op een PC – kunnen ze een aanmerkelijke hindernis vormen voor het leren kennen en leren

omgaan met nieuwe modellen, perspectieven en werkwijzen. Hiervan is bijvoorbeeld sprake in de editiewetenschap waar verreweg de meeste digitale edities in essentie digitale metaforen van het boek zijn. Deze volgen een logica die volledig boek-mimetische is en waarin de digitale editie niets meer of minder is dan een zo nauwkeurig mogelijke imitatie van het fysieke document. Deze benadering negeert vrijwel alle nieuwe intrinsieke omgangsvormen met het tekstmateriaal die andere digitale modellen mogelijk bieden. Dit effect heb ik “paradigmatische regressie” genoemd, en het is één van de redenen waarom we – in de context van tekst- en editiewetenschap – weinig experimenten met alternatieve modellen voor digitale tekst aantreffen. Zulke experimenten zouden zelfs al gebaseerd kunnen zijn op de toepassing van een zeer elementaire en wijdverspreide Internet-technologie zoals de http-link. Bijvoorbeeld voor het beschrijven en modelleren van intertekstualiteit. Deze mogelijkheid mag ook binnen de gemeenschap van tekst- en editiewetenschappen – bij monde van bijvoorbeeld Theodor Nelson en George Landow – genoegzaam bekend verondersteld worden, maar tot geïnspireerde experimentele digitale edities heeft dit niet geleid. Natuurlijk is het regressie-effect niet alleen een gevolg van technologie. Het volgt ook uit bewuste theoretische en praktijkmatige keuzes door individuele editoren en van keuzes binnen hun academische gemeenschap. De sterke neiging van TEI-XML tot het imiteren van documenten en de levenloos geboren “social edition” suggereren dat deze keuzes in zekere zin conservatief zijn en in lijn met de idee dat het mandaat van editoren en editie nog altijd volledig geclaimd wordt door de wetenschappelijk editoren en een tekstmodel nog steeds geheel is onderworpen aan de vorm en beperkingen van het fysieke boek.

Er is weinig bewijs te vinden dat de interactie tussen editiewetenschappers en digitale technologie resulteert in belangwekkende technisch-methodologische innovatie. Een verwacht effect zou de ontwikkeling van een methodisch-technische mengtaal zijn, een methodologisch pidgin-dialect dat duidt op de uitwisseling van terminologie en werkwijzen tussen de domeinen. De case studies in deze dissertatie, die met name dergelijk ontwikkelingen beschouwen binnen de context van het Huygens Instituut, getuigen niet van dergelijke uitwisseling behoudens enige terminologie uit het IT-projectmanagement die niet strikt betrekking heeft op tekst- en

editiewetenschappelijke methodologie of op computationele methoden. De methodologische veranderingen waarnaar gewezen kan worden lijken eerder nogal onzichtbaar te worden doorgevoerd op het niveau van programmatuur, softwarematige modellen en statistische methoden, waarvan de technische details opnieuw worden toegedekt door grafische interfaces en visualiseringen.

Het paradigmatisch regressie-effect is een kracht die de scholen van “close” en “distant reading” eerder uit elkaar dan naar elkaar toe drijft. Terwijl in theorie XML-gebaseerde edities ook in dienst staan van computationeel literair onderzoek, zijn het in de praktijk onontwarbare kluwen van XML-tags. Van de andere kant hebben computationeel literair analisten een voorkeur voor het ‘on the fly’ annoteren en analyseren van niet-onderhoudbare digitale wegwerpteksten. Een gedeeld tekstmodel c.q. formaat is noodzakelijk als de expertise en kunde van beide domeinen elkaar blijvend willen versterken. Kennisgrafien (“knowledge graphs”) lijken hier uitkomst te bieden.

Het lijkt vanzelfsprekend dat de editiewetenschappen vanuit principe geïnteresseerd zouden moeten zijn in betere modellen voor de beschrijving en analyse van digitale objecten (zoals digitale teksten en edities). Ten eerste omdat er een algemene trend van digitalisering van cultuur en maatschappij gaande is waardoor culturele objecten en processen steeds vaker digitale objecten en processen zijn. Ten tweede omdat deze ‘softwarisatie’ ook zijn weerslag heeft op de methodes van de editietechniek, zij het meer onzichtbaar dan zichtbaar. Wanneer geesteswetenschappelijke activiteit en tekstkritische beslissingen worden gedelegeerd aan softwareontwikkelaars en deze door softwareontwikkelaars opnieuw gedelegeerd worden naar de code die zij schrijven, dan is het belangrijk dat de editiewetenschap beschikt over methoden en technieken om zulke gedelegeerde processen en taken te toetsen op wetenschappelijke correctheid. Dit is van des te groter belang naarmate ontwikkelde software specifiek geschreven is voor eenmalige analytische toepassing (zogenoemde “bespoke” of “tailor-made” code). In tegenstelling tot wat gebruikelijke maar misplaatste metaforen willen doen geloven, is softwarecode geen neutrale amorele techniek, maar een product van menselijke creativiteit met ingebedde aannames en perspectieven op de realiteit. Dus om de wetenschappelijkheid en kwaliteit van het editieproces te waarborgen dient er een voldoende systematiek

te zijn voor softwarekritiek en voor de “peer review” van softwarecode. Op deze wijze kan de wetenschappelijke verantwoordelijkheid en ethiek van softwareontwikkelaars die dergelijk code produceren bevorderd en gewaarborgd worden. Het is momenteel niet strikt eenduidig vast te stellen welke categorieën van code onder een dergelijke regime zouden moeten vallen en welke niet. Het is evident dat een computertaal zoals R niet nog eens door tekstwetenschappers gecontroleerd hoeft te worden – zo zij dit al zouden kunnen. Zo’n techniek is ruim voldoende geëvalueerd door computerwetenschappers en statistici. Maar software die specifiek geschreven is – bijvoorbeeld in R – om een specifieke editietaak uit te voeren in de context van een specifiek en éénmalig editieproject, zou wel degelijk aan code peer review onderhevig moeten zijn als onderdeel van de wetenschappelijke evaluatie. Niettegenstaande het fascinerende werk dat al gebeurt in softwarestudies, ccs (“critical code studies”) en mediastudies met betrekking tot de kritische bestudering van de rol van software in cultuur en maatschappij, is er geen in het wetenschappelijk proces ingebedde systematiek voor code peer review in de tekst- en editiewetenschappen, noch in de geesteswetenschappen in brede zin. Voor code peer review is het kunnen lezen en begrijpen van code echter noodzakelijk en deze vaardigheden zijn zeldzaam in de geesteswetenschappen in het algemeen en in de editiewetenschap in het bijzonder.

Wanneer bepaalde editiewetenschappelijke taken gedelegeerd worden aan softwarecode is het opportuun de rol van diegenen die die softwarecode schrijven nader te beschouwen. In een zekere zin worden programmeurs ook wetenschappelijk auteurs en editeurs, wat een beschouwing omtrent auteurschap en editeurschap relevant maakt. Extreem verkort kan de intellectuele en filosofische geschiedenis van het auteurschap gezien worden als een discourse omtrent wie het mandaat heeft op het definiëren van de betekenis van tekst. Afhankelijk van tijd en context heeft het antwoord op die vraag gevarieerd van “goden” en “God” in vroeger tijden tot “de auteur” en uiteindelijk “de lezer” in meer recente tijden. Het claimen van een dergelijk mandaat is gerelateerd aan het claimen van zekere (soms absolute) waarheden. Postmodern relativisme en de doodverklaring van de auteur door onder andere Derrida hebben duidelijk gemaakt dat dit soort claims altijd (inter)subjectief zijn.

De editiewetenschap heeft grosso modo op twee verschillende manieren gereageerd op het postmoderne relativisme en de daaruit volgende ‘crisis’ met betrekking tot feit en waarheid. Eén manier is om de editie voor te stellen als een vorm van archief waarbij alle relevante bronnen en interpretaties betrokken kunnen en idealiter moeten worden. Deze richting accepteert volledig de intersubjectiviteit van interpretatie en het procesmatige karakter van schrijven, lezen en editeren, alsmede het ‘fluïde’ karakter van tekst. Deze benadering is inclusief met betrekking tot oorspronkelijk materiaal maar ook met betrekking tot de lezer c.q. gebruiker, die uitgenodigd wordt om interpretaties en annotaties toe te voegen, waardoor een ‘open science’ en ‘open einde’ editie ontstaat. Deze richting in het editeren wijst dan ook graag naar digitale (Web)technologie om deze open vormen van interactiviteit te realiseren. De andere reactie is meer reactionair en streeft er naar het model van het boek opnieuw als de dominante vorm te bevestigen. Deze stroming streeft ernaar de “filologische feiten” voorrang te verlenen boven een hegemonie van interpretatie en fluïditeit. Wat betreft hun eigen voorkeur is het niet ongebruikelijk voor editeurs om zowel in theorie als in de praktijk van het editeurschap heen en weer geslingerd te worden tussen deze extreme posities, zoals bijvoorbeeld in het geval van Jerome McGann.

Uit nadere beschouwing van de rollen van auteur, editeur en softwareontwikkelaar blijkt dat het schrijven van code ook een vorm van auteurschap is. In de context van tekst- en editiewetenschappen zijn zowel editeur als programmeur in wezen revisionaire auteurs van oorspronkelijke werken. Maar in verreweg de meeste gevallen blijkt het schrijven van code in het geval van digitale edities niet geclaimd te worden als een vorm van auteurschap – vaak wordt alleen de claim van de editeur op het werk erkend. In vergelijking met traditionele vormen van editiewetenschap voegt het schrijven van code tenminste één laag van interpretatie toe aan het proces van editeren. Dit introduceert in ieder geval twee nieuwe problemen. Ten eerste worden tekst en interpretatie nog meer fluïde en daarmee meer instabiel, wat ingaat tegen wat nog vaak als doel van filologie en tekstwetenschap wordt gezien: het op een zekere manier stabiliseren van de tekst. Natuurlijk is het idee van de stabiele tekst an sich al erg problematisch, maar de toevoeging van het schrijven en toepassen van softwarecode creëert hoogstens meer mogelijkheden

om de tekst en zijn interpretatie op verschillende manieren te veranderen en te vermeerderen en daarmee te destabiliseren. Temeer omdat code door ingebouwde heuristiek in zekere zin gedelegeerde (wetenschappelijke) activiteit kan verrichten, en in bepaalde gevallen autonome beslissingen neemt. Daarbij heeft code een retoriek en performativiteit die tot nu toe weinig systematisch onderzocht is en dus weinig begrepen. Daarbij is het Belangrijk om te constateren dat geen van deze effecten effectief gecontroleerd wordt als onderdeel van het (editie)wetenschappelijk proces omdat programmeren door noch ontwikkelaar noch geesteswetenschapper erkend wordt als schrijven of als editeren. Hierdoor blijft de invloed van programmeren op de editiewetenschappelijke methodologie grotendeels onbekend en ongeëvalueerd.

De introductie van de programmeur – of computer science expert, ontwikkelaar, IT-specialist, of hoe degene ook wordt genoemd die de code schrijft waarvan gezegd kan worden dat deze een geesteswetenschappelijke activiteit uitvoert – in het geesteswetenschappelijk proces leidt tot een tot nog toe onopgelost probleem van wetenschappelijke verantwoordelijkheid. Wanneer digitale objecten en softwarecode integraal onderdeel worden van een geesteswetenschappelijke argumentatie, dan dienen de producenten van die originele bijdrage een claim te leggen op de creatie ervan. Niet alleen om gerechtvaardigd claim te leggen op de academische waarde en waardering voor dat werk, maar vooral om de wetenschappelijke verantwoordelijkheid te nemen voor de wetenschappelijke argumentatie die zij zo co-creëren. De intentie en betekenis van het revisionaire auteurschap dat code produceert spreekt niet voor zich. Dit houdt voor elke vorm van auteurschap. Al in de jaren 1980 beargumenteerde Donal Knuth dat code retorische functies heeft en dat programmeurs een, zij het vaak impliciete, poëtica hanteren. Hij produceerde zelfs een model en een implementatie van een computertaal om deze aspecten expliciet te kunnen maken. Helaas is zijn betoog en motivering later meesttijds verkeerd begrepen, weinig tot niet onderwezen en al helemaal zelden toegepast. Maar deze meer fundamentele theoretisch retorische beginselen van code verdienen hernieuwde aandacht als beginselen van het revisionaire auteurschap van code dat programmeurs vaker en actiever zouden moeten claimen.

De case studies in deze dissertatie laten zien dat een interdisciplinaire gemeenschap van ontwikkelaars en tekstwetenschappers actief op zoek



is naar de mogelijk interacties tussen geesteswetenschappelijke analyse en digitale modellen voor tekst. Verschillende modellen benadrukken verschillende aspecten van datgene wat ze modelleren. TEI-XML lijkt in de meeste gevallen het model van de codex (boek) en de tekststructuur van fysiek gedrukte werken op de voorgrond te plaatsen. De statistische modellen uit de “distant reading” methodologie benaderen een tekst als een reeks van tekens en tokens en benadrukken de meetbaarheid daarvan. Beide modellen stellen onderzoekers in staat tot hoogwaardige en verfijnde analyse en interpretatie. Maar beide gaan ook uit van een erg nauwe en armoedige opvatting van wat tekst is: structuren of strings. Ze verliezen beide aantoonbaar veel informatie ten aanzien van elke fysiek object dat ze beschrijven, wat eenvoudig te demonstreren valt door te wijzen op het feit dat ze bijvoorbeeld veelal alle materiële eigenschappen van een fysieke bron negeren. Dat wil niet zeggen dat die modellen die aspecten fundamenteel niet kunnen representeren, maar dat hun specifieke compositie ze bijzonder geschikt maak om bepaalde aspecten te beschrijven waardoor zij die aspecten uitlichten en benadrukken. Wanneer je een ‘lens’ toepast die bijzonder geschikt is om tekststructuur bloot te leggen, dan zal tekststructuur analytisch gezien op de voorgrond geplaatst worden. Als je een ‘lens’ gebruikt die een tekst als een losse verzameling woorden beschrijft, dan zullen alle teksten er als losse verzamelingen van woorden uit gaan zien. Als je alleen een hamer hebt, zien verdacht veel dingen eruit als spijkers.

De conventionele analytische gereedschappen van de tekst- en editiewetenschap, waartoe bijvoorbeeld het “apparatus criticus” behoort, zijn ook modellen. Deze modellen geven hun eigen specifieke focus aan interpretatie en analyse. Nieuwe en andere modellen introduceren nieuwe mogelijkheden, perspectieven en uiteindelijk kennis. Zoals uitgelegd vertoont de editiewetenschap een zekere mate van paradigmatische regressie die het haar lastig maakt veel met nieuwe modellen te experimenteren. Een deel van de verklaring voor deze relatieve neiging tot regressie moet gezocht worden in de overweldigende vrijheid van modelleren die digitale technologie biedt. Die bijna absolute vrijheid staat in groot contrast tot de specifieke en bewust opgelegde en gehanteerde beperkingen van het “apparatus criticus” dat over verscheidene eeuwen heen ontwikkeld is, en mede is ingegeven en vormgeven door de grenzen en beperkingen van het fysieke boek. Maar

met het oog op het onderzoeken van een nieuwe digitale of computationele epistemologie zou het nuttig zijn om tekst ook soms te bevrijden van deze welbekende modellen en de epistemologische mogelijkheden van meer en andere modellen voor tekst te onderzoeken.

Een aantal case studies wijst in de richting van grafen als een krachtige expressie van de multidimensionale aspecten van tekst. Zulke grafen zouden dus een goede kandidaat kunnen zijn voor een nieuw tekstmodel in de editiewetenschap en tekstkritiek. De voorbeelden in de case studies tonen aan dat zulke graafgebaseerde modellen specifieke voordelen bieden boven conventionele modellen – bijvoorbeeld een praktische oplossing voor het probleem van overlap in TEI-XML, de representatie van multidimensionaliteit, en het inzichtelijk in kaart brengen van tekstvariatie.

Een belangrijke wetenschappelijke verantwoordelijkheid voor elk wetenschappelijk veld is de routinematige reflectie op de eigen methodologie. De sociale wetenschappen, en meer speciaal de Science and Technology Studies als onderdeel daarvan, beschikken over uitstekende methoden en technieken voor dergelijke reflectie. Bijvoorbeeld in de vorm van etnografie en auto-etnografie. Voordat er generaliserende conclusies getrokken kunnen worden omtrent waardevolle wijzen van digitaal-technologische innovatie in het hermeneutisch domein zullen er nog veel case study gebaseerde onderzoeken nodig zijn naar de interactie tussen computerwetenschap, softwareontwikkeling en de geesteswetenschappen. Etnografisch en auto-etnografisch onderzoek naar pogingen tot zulke innovatie in de editiewetenschap in de context van het Huygens Instituut toont aan dat het erg moeilijk is om methodologie in dit domein te innoveren op basis van digitale methoden met zorgvuldig behoud van de hermeneutische benadering. Het blijkt dat zowel geesteswetenschappers als digitaal specialisten en ontwikkelaars ernaar moeten streven zich meer bewust te zijn van de eigen methodologische beperkingen en meer zouden moeten proberen om elkaars wetenschappelijke sterke punten te waarderen als aanvulling op de eigen methoden en technieken. Een auto-etnografie van vijftien jaar ervaring op het kruispunt van deze wetenschappelijke domeinen laat zien dat zowel geesteswetenschappers als digitaal-technologische specialisten zich meer respectvol bewust en geïnteresseerd kunnen tonen voor elkaars wetenschappelijke kwaliteiten, en dat beide zich bewuster kunnen zijn van hun

## *Samenvatting*

individuele en gemeenschappelijke wetenschappelijke verantwoordelijkheid ten aanzien van methodologische reflectie en innovatie.

De huidige computationele benaderingen die worden toegepast in het hermeneutische domein zijn behept met de retoriek van snelheid en schaalvergroting uit de industriële automatisering. Deze retoriek en de bijbehorende benaderingen creëren een blinde vlek voor een belangrijke en echte computationele uitdaging: dat van het probleem van de computationele hermeneutiek. Computationeel onderzoekers zouden aangemoedigd moeten worden deze uitdaging aan te gaan. Van de andere kant doen geesteswetenschappers er goed aan om te pogen minder geïntimideerd te raken door hun computationele collega's en deze computationele ontdekkingstocht aan te gaan vanuit het vertrouwen en het volle besef dat hun werk en methoden wortelt in de onmetelijk waardevolle wetenschappelijk kracht van meer dan twintig eeuwen hermeneutiek.