

'n Inligtingstegnologie-gesentreerde gebruikerskoppelvlak vir navorsingsdoeleindes binne die geesteswetenskappe met spesifieke verwysing na die Afrikaanse letterkunde

Burgert Senekal

Burgert Senekal: Departement Afrikaanse en Nederlands, Duits en Frans, Universiteit van die Vrystaat

Opsomming

Die inligtingsontploffing is 'n bekende probleem wat alle instansies konfronteer wat betrokke is by inligting- en kennisbestuur, insluitend akademiese navorsing en militêre intelligensie. In 'n groot mate te danke aan die astronomiese Amerikaanse verdedigingsbegroting (in 2009 was die Amerikaanse intelligensie-begroting alleen \$48 miljard (Aldrich 2009:889)), is militêre intelligensie aan die voorpunt van tegnologiese ontwikkeling wat betref inligtingsbestuur, en aangesien die akademiese navorser ook gekonfronteer is met die "data-stortvloed", is die lesse wat geleer word in militêre intelligensie ook nuttig vir die akademiese gemeenskap. Hierdie artikel integreer militêre-intelligensie-tegnologie en konsepte binne die raamwerk van die militêre-intelligensie-siklus met die navorsingsproses ten einde die kwantitatiewe en kwalitatiewe voordele te illustreer wat die uitgebreide inlywing van inligtingstegnologie vir navorsingsdoeleindes in die geesteswetenskappe inhou, met spesifieke verwysing na die Afrikaanse letterkunde. Dit word aangevoer dat konsepte van militêre intelligensie van toepassing is op akademiese navorsing, spesifiek die gebruik van data-ontginning, visualisering en sosiale-netwerk-analise (SNA). Die skema wat deur Pirolli en Card (2005:3) ontwikkel is, word gebruik in samewerking met NAVO se intelligensie-siklus om 'n proses uiteen te sit waarin inligtingstegnologie opgeneem is in elke fase van die navorsingsproses: beplanning, versameling, ontleding, en verspreiding. Sommige militêre-intelligensie-ontledingsprogrammatuur word getoets en bespreek, terwyl goedkoper (en gratis) weergawes ook gebruik word om die voordele van die integrasie van hierdie programmatuur vir navorsingsdoeleindes te beskryf, en neem hierdeur Senekal (2011) se voorstelle heelwat verder. Uiteindelik argumenteer die artikel vir 'n wyespektrum-integrasie van inligtingstegnologie in akademiese navorsingsaktiwiteite, en illustreer die potensiaal van inligtingstegnologie met 'n aantal voorbeelde waardeur tendense binne die Afrikaanse literêre sisteem voorlopig geïdentifiseer en visueel verteenwoordig word. Die artikel maak egter geen literêre bewerings op sigself nie: dit maak bloot voorstelle oor hoe inligtingstegnologie gebruik kan word om navorsing te doen in die Afrikaanse letterkunde in die 21ste eeu.

Trefwoorde: militêre intelligensie; sosiale-netwerk-analise; visualisasie; data-ontginning; geesteswetenskaplike navorsing

Abstract

Academic intelligence: An information technology-centred user interface with cyberspace for research purposes within the humanities

The information deluge is a well-known problem that confronts all institutions involved in information and knowledge management, including academic research and military intelligence. The terrorist attacks on the World Trade Center in New York on 11 September, 2001 highlighted the problem of integrating vast amounts of data in a coherent manner, and in response the US intelligence community in particular has sought to develop software applications to process the data deluge more efficiently. While computer platforms are continually being developed to store and manage larger amounts of information, human development lags behind because of limited “scalability”. Roberts (2011:9) states, “Although it is a matter of time [...] before the fundamental limitations of physics are encountered in computer and graphics technology, we are asymptotically approaching the limits of human capability to process data that are collected.” As early as 1996 Geyer (1996:xxiii) identified this problem as the postmodern manifestation of meaninglessness within Seaman’s (1959) model of alienation and argued that we need to develop “adequate new scanning mechanisms to gather the goal-relevant information one needs, as well as more efficient selection procedures to prevent being overburdened by the information one does not need, but is bombarded with on a regular basis”.

Each document considered during any research project has the potential to support or disprove the relationships between entities within a system, and therefore every document, every text, has value for the end result (Graham and Carley 2008:9), which means that the more documents that are taken into consideration, the greater the chances that ultimate conclusions will be accurate. However, a human researcher, or even a team of researchers, cannot integrate knowledge dispersed over such vast databases as exist now, since information stored digitally (notwithstanding analogue collections) now comprise zettabytes’ worth of data globally. The technology responsible for the information explosion is therefore the only viable way to deal with such volumes of data, but also requires adjustments of the researcher – a new information technology-centred user interface with cyberspace.

Due to a large extent to the astronomical USA defence budget (in 2009 the USA intelligence budget alone was \$48 billion (Aldrich 2009:889)), military intelligence is at the forefront of technological development in terms of information management, and since the academic researcher is also confronted with the data deluge, the lessons learned in military intelligence are also useful for the academic community. Thomas and Cook (2006:10) argue precisely that the core technologies developed within military intelligence will also have much wider applicability and the potential to add value in almost any situation in which information overload and complexity hampers effective, fast analysis and communication of results.

This article therefore integrates military intelligence technology and its derivatives and concepts within the framework of the military intelligence cycle to illustrate the quantitative and qualitative benefits that streamlining research through the extensive incorporation of information technology holds for research within the humanities, with specific reference to Afrikaans literature. It is argued that concepts from military intelligence are transferrable to academic research, specifically the use of data mining, visualisation and Social Network Analysis (SNA), especially when attempting to analyse the Afrikaans literary system.

The schema developed by Pirolli and Card (2005:3) is used in conjunction with NATO’s intelligence cycle to construct a process in which information technology is incorporated in every phase of research: direction, collection, analysis and distribution. Some military

intelligence analysis programs are tested and discussed, while cheaper (and free) civilian versions are also used to describe the benefits of integrating these for research purposes. Treverton et al. (2006:25) write, "If it is to stay relevant, intelligence needs to be like human vision – binocular, providing depth to what we can be seen in front; and peripheral, which provides warning." The authors thus argue that information should be analysed deeper and wider within a military intelligence context, but the same applies to academic research and the study of literature.

Especially within systems theory, attempts have been made to study literature in a broader capacity, which necessarily means that larger data sets have to be considered. Within this paradigm, information technology is particularly suitable, especially SNA and other forms of visualisation. Visualisation creates the opportunity to analyse large amounts of data at a single glance, after which conclusions can be drawn and trends can be investigated in greater depth. It can expose features and patterns that can stimulate further research when questions are identified that were not even asked previously. Although visualisation is therefore based on quantitative methods, it also speeds up and facilitates qualitative research, as less time spent on identifying trends translates into more time becoming available to analyse and explain trends.

This view is akin to Pirolli and Card's (1999) concept of the researcher/analyst as an "information predator": within the current academic setting, where budgets and time become more and more limited, greater output can be generated in a more cost-effective manner, provided that information technology is used across the entire research platform, over and above the ability of information technology to make a qualitative contribution to research.

The article takes Senekal's (2011) suggestions to fruition by using digital publications for new research. Ultimately the article argues for a full-spectrum integration of information technology in academic research activities, illustrating its potential with a few examples whereby trends within the Afrikaans literary system are provisionally identified and represented visually. However, the article makes no literary claims in itself: it merely suggests avenues in which information technology could be employed to conduct research in Afrikaans literature in the 21st century.

Keywords: military intelligence; social network analysis; SNA; visualisation; data mining; Afrikaans literary system; humanities research

1. Inleiding

[T]here are known knowns; there are things we know we know.

We also know there are known unknowns; that is to say we know there are some things we do not know.

But there are also unknown unknowns – the ones we don't know we don't know.

Hierdie stelling is in 2002 deur Donald Rumsfeld, destyds die Amerikaanse minister van verdediging, gemaak met verwysing na sogenaamde massavernietigingswapens in die aanloop tot die inval van Irak. Alhoewel heelwat spot gedryf is met hom hieroor, het hy 'n bekende probleem rondom inligting geformuleer, naamlik dat daar heelwat inligting is

waarvan mense nie eens bewus is dat hulle dit nie het nie, bo en behalwe dit wat ons wéét ons nie weet nie.

In reaksie op Rumsfeld se uitlating onderskei Slavoj Žižek 'n vierde kategorie van inligting: “unknown knowns” – dinge wat ons nie besef ons wel weet nie (Žižek 2004), wat onder andere die gevolg is van vooroordele en teoretiese of selfs morele en etiese invalshoeke. Johnsson (1981:238) het byvoorbeeld drie dekades gelede opgemerk: “[T]he theoretical frame of reference that governs recognition is a constitutive element in the blindness of any interpretative insight.” Rumsfeld en Žižek se kategorieë van inligting kan dus wees:

1. “known knowns” – dinge wat bekend is, byvoorbeeld dat die Grensoorlog plaasgevind het
2. “known unknowns” – dinge wat ons weet onbekend is, byvoorbeeld of die aarde die enigste planeet is waarop lewe voorkom.
3. “unknown unknowns” – dinge wat met verdere navorsing blootgelê word sonder dat daar juis na hierdie inligting gesoek word
4. “unknown knowns” – dinge wat as gevolg van huidige invalshoeke misgekyk word.

Terwyl kategorie vier buite die bestek van die huidige artikel val en eerder oorgelaat behoort te word aan wetenskapsfilosowe en -teoretici, word kategorieë twee en drie in hierdie artikel betrek ten einde aan te toon hoe belangrik die gebruik van inligtingstegnologie binne geesteswetenskaplike navorsing is, alhoewel dit soms misgekyk word. Die Amerikaanse National Visualization and Analytics Centre (NVAC) poog juis “to detect the expected and discover the unexpected from massive and dynamic information streams and databases consisting of data of multiple types and from multiple sources, even though the data are often conflicting and incomplete” (Thomas en Cook 2006:10).

Inligting is 'n wesenlike probleem in die 21ste eeu – meer so as in vorige tydperke waar relevante, akkurate inligting uit kleiner datastelle ontgin moes word. Bo en behalwe die problematiek rondom die verifiëring van inligting en die impak wat die leser se vooroordele het, word die hedendaagse mens met 'n data-tsoenami gekonfronteer. Reeds in 2005 het die NVAC (2005:2) gewaarsku: “our ability to collect data is increasing at a faster rate than our ability to analyse it”. In 2002 alleenlik is 5 exagrepe (5×10^{18} bytes) wêreldwyd gestoor in die vorm van film, papier en elektroniese media, met 'n verdere 18 exagrepe se stromende data wat gegenereer is (NVAC 2005:24). Die hoeveelheid data wat digitaal beskikbaar is, niteenstaande dit wat nog digitaal beskikbaar gestel moet word, beslaan reeds verbysterende volumes: daar word geskat dat daar teen 2007 195 eksagrepe (10^{18}) se data wêreldwyd digitaal gestoor is, en daar is verwag dat hierdie volume teen 2011 tot 1,8 zettagrepe (10^{21}) sou styg (Darvill 2011:5). Roberts (2011:9) skryf:

Although it is a matter of time (some experts say 10 years or more) before the fundamental limitations of physics are encountered in computer and graphics technology, we are asymptotically approaching the limits of human capability to process data that are collected.

Neri en Pettoni (2009:35) verwys na die moderne paradoks rakende inligting: die beskikbaarheid van 'n groot hoeveelheid inligting lei daartoe dat mense deur inligting oorlaai word wat meestal geen bruikbare kennis beteken nie. Meer inligting beteken dus nie noodwendig meer kennis nie; soms juis die teendeel.

Een van die probleme met die inligtingsontploffing is “human scalability”: die mens kan nie vinnig genoeg aanpas by die inligtingsvereistes wat nou aan hom gestel word nie. Die brein se kapasiteit het volgens Darvill (2011:5) in 2,5 miljoen jaar verdubbel, maar dit is hopeloos te stadig om met die data-tsoenami tred te hou, aangesien Moore se Wet bepaal dat verwerkerspoed en geheuedigtheid elke 18 maande verdubbel (NVAC 2005:25). Een manier om hierdie probleem aan te spreek, is deur spanwerk aan te moedig, maar selfs groot spanne navorsers kan nie die hoeveelheid inligting hanteer wat beskikbaar is nie. Met vrye en onmiddellike toegang tot internasionale publikasies word die navorser oorval deur data: word 'n item soos J.M. Coetzee byvoorbeeld op Google Scholar gesoek, lewer dit 1,850,000 resultate in 0,15 sekondes op – heelwat meer artikels as waardeur selfs 'n span navorsers kan lees; tegnologie is onontbeerlik. Graham en Carley (2008:9) skryf:

The history of the progress of man is about how he has leveraged technology to expand his limited cognitive capacities using tools – the microscope, the telescope and now in the modern age, tools that help him gain access to physical/knowledge ghosts whose presence cannot be doubted but can only be seen through the application of quantitative methods.

Die tegnologie wat vir die inligtingsontploffing verantwoordelik is, is dus ook die enigste werkbare manier om met sulke volumes inligting om te gaan (Senekal 2011), maar die omgang verg aanpassings van die navorser – 'n nuwe inligtingstegnologie-gesentreerde gebruikerskoppelvlak¹ met die kuberruimte. Pirolli en Card (1999:3) beklemtoon:

Our adaptive success depends to a large extent on a vast and complex tributary of cultural tasks that engage our physical and social environments. These tasks require increasingly sophisticated information-gathering, sense-making, decision-making, and problem-solving strategies.

Schreibman, Siemens en Unsworth (2004:xxvi) voer aan dat dieselfde vir die geesteswetenskappe geld, waar inligtingstegnologie 'n navorser in staat stel om verbande tussen tekste en terme, patrone, saamvoegings en afwesighede te identifiseer wat die navorser nie daarsonder sou kon herken nie. Nie alleen verskaf inligtingstegnologie dus 'n noodsaaklike manier om groot volumes inligting te hanteer nie, maar dit skep ook die geleentheid om vanuit 'n ander invalshoek na bronne te kyk. Besser (2004:558) skryf: “Though the promise of digital technology in almost any field has been to let one do the same things one did before but better and faster, the more fundamental result has often been the capability of doing entirely new things.” Die huidige artikel bou veral voort op navorsing wat verrat is in Schreibman e.a. (2004).

Militêre intelligensie is een gebied waar die belangrikheid van tegnologiese ontwikkeling in antwoord op die data-tsoenami oor die afgelope dekade deurlopend beklemtoon is. Ter

illustrasie: in 2005 was die teikens van dataverwerking in die VSA se intelligensiedienste die tempo van een miljard nuwe gestruktureerde boodskappe of transaksies per uur, en een miljoen nuwe ongestruktureerde boodskappe of dokumente per uur (NVAC 2005:24). Die bestuur van 'n oormaat inligting is 'n sentrale probleem in hierdie milieu, waar juis 'n ontoereikende bestuur van intelligensie te blameer is vir die terroriste-aanvalle van 11 September 2001. Na 11 September is sagteware doelgerig ontwikkel om groot hoeveelhede data oor wydverspreide databasisse en in verskillende formate te kan integreer (Splivalo 2009 en Darvill 2011:5), en die Amerikaanse departement van binnelandse veiligheid het in 2004 die National Visualization and Analytics Center gestig: “with the mission of stimulating next-generation technologies and talents to reduce the risk of terrorism” (Thomas en Cook 2006:10).

In 'n groot mate te danke aan die VSA se astronomiese verdedigingsbegroting (volgens Aldrich 2009:889 was die VSA se intelligensiebegroting alleen in 2009 \$48 miljard) staan militêre intelligensie aan die voorpunt van tegnologiese ontwikkeling wat betref inligtingsbestuur, en aangesien die akademiese navorser ook gekonfronteer word met die data-wolkbreek, is die lesse wat in militêre intelligensie geleer word, bruikbaar vir die akademie. Thomas en Cook (2006:10) voer juis aan dat die kerntegnologieë wat binne militêre intelligensie ontwikkel word, ook veel wyer toepaslikheid sal hê, met die potensiaal om waarde toe te voeg in bykans enige situasie waarin inligtingsoorlading -en kompleksiteit effektiewe, vinnige ontleding en kommunikasie van resultate belemmer. Hierdie artikel stel dus pertinent ondersoek in na die potensiaal en bruikbaarheid van sulke programmatuur vir navorsingsdoeleindes, met spesifieke verwysing na die Afrikaanse letterkunde, maar verskaf ook 'n oorsig oor ander verwante rekenaarprogramme wat heelwat goedkoper (en selfs gratis) is. Wanneer programmatuur in hierdie artikel gebruik word om tendense in die Afrikaanse letterkunde te ontleed, of om die belangrikheid van spesifieke skrywers of uitgewers aan te toon, is die oogmerk egter nie om iets voor te stel oor die belangrikheid van spesifieke skrywers of tendense in die Afrikaanse literêre sisteem² oor die algemeen nie, maar eerder om die potensiaal van inligtingstegnologie te illustreer. Hierdie artikel gaan dus slegs verkennend om met data ten einde die potensiaal van inligtingstegnologie te illustreer, en maak geen afleidings na aanleiding van die ontleding van data nie.

2. Die intelligensieproses

Die VSA se departement van verdediging definieer *intelligence* as: “information and knowledge obtained through observation, investigation, analysis, or understanding” (aangehaal in Roberts (2011:5)). Intelligensie in hierdie sin van die woord is dus *inligting* wat *verwerk* is, soos Bose (2008:510) beklemtoon: “Intelligence differs from data and information because it requires some form of analysis, whose purpose is to derive some meaning from the piles of data and information that bury every organization.” Brei (1996:4) beskryf die verskil tussen inligting en intelligensie meer omvattend:

[I]ntelligence is more than information. It is knowledge that has been specially prepared for a customer's unique circumstances. The word knowledge highlights the need for human involvement. Intelligence collection systems produce [...] data, not intelligence; only the human mind can provide that special touch that makes sense of data for different customers' requirements. The special processing that partially defines intelligence is the continual collection, verification, and analysis of information that allows us to understand the problem or situation in actionable terms and then tailor a product in the context of the customer's circumstances. If any of these essential attributes is missing, then the product remains information rather than intelligence.³

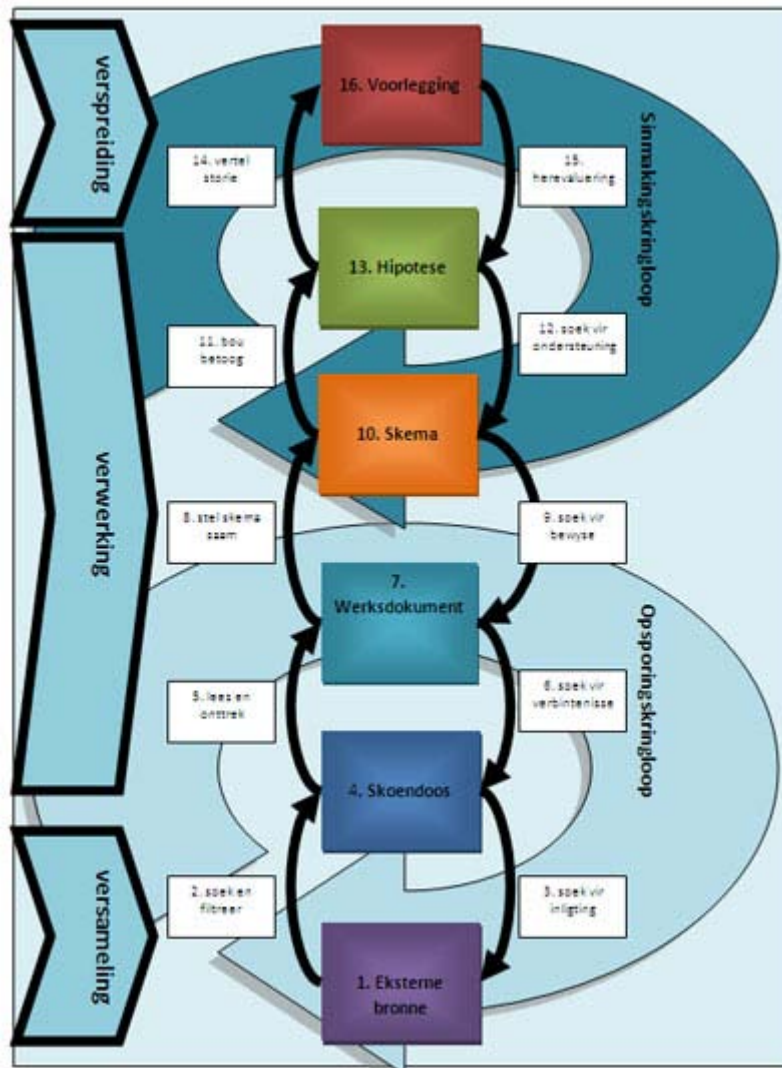
Krizan (1999:7) noem dat die sakektor gewoonlik die term *inligting* bo *intelligensie* verkies, maar sy verkies self die term *intelligensie*, juis omdat dit die toevoeging van waarde deur ontleding beklemtoon. As sodanig kan intelligensie vergelyk word met navorsing, waar nuwe betekenis geskep word in die verwerking van onder andere bestaande teorieë, historiese gegewens, en feite, en die representasie daarvan in 'n afgehandelde formaat ('n publikasie).⁴

Die sogenaamde intelligensiesiklus verskaf vier stappe in die versameling en verwerking van intelligensie, wat ook van toepassing op navorsing binne die akademiese konteks is: beplanning, versameling, verwerking en verspreiding (vgl. byvoorbeeld Mouton en Marais 1990:25 se vyf tipiese stappe in die navorsingsproses, of Vermeulen, Lategan en Litheko 2011:15 se soortgelyke navorsingsiklus). Biermann, Korsnes, Rohmer en Ündeger (2004:3) gee 'n opsomming van hierdie siklus vanuit die Noord-Atlantiese Verdragsorganisasie (NAVO) se glossarium van terme en definisies (AAP-6) [6], wat hier aangepas word vir akademiese navorsing (sien ook Treverton en Gabbard 2008:3):

- *Beplanning* – Die bepaling van watter inligting nodig word, beplanning van navorsing, en delegering aan relevante persone (byvoorbeeld assistente).
- *Versameling* – Die gebruik van alle bronne tot die navorser se beskikking om die nodige data te versamel.
- *Verwerking* – Die evaluering, ontleding, integrasie en interpretasie van data, met ander woorde die omskakeling van data na intelligensie.
- *Verspreiding* – Die doelgerigte verspreiding van bevindinge in die gepaste formaat, byvoorbeeld die publikasie van artikels, monografieë en boeke.

Terugvoering is 'n vyfde stap wat Bose (2008:512–3) byvoeg wanneer hy na besigheidsintelligensie verwys. Vir die doeleindes van hierdie artikel word egter volstaan by Bierman e.a. (2004) se vier stappe, aangesien terugvoering reeds geïntegreer is in die model soos hier onder uiteengesit.

'n Meer gedetailleerde manier om die proses visueel voor te stel is deur die gebruik van Pirolli en Card (2005:3) se skema. In die hieropvolgende diagram word hulle skema met NAVO se intelligensiesiklus geïntegreer, maar let daarop dat die beplanningsfase nie ingesluit word nie, omdat dit die proses voorafgaan, en die aanduiding van deurlopende terugvoering vervang in hierdie geval die sikliese uitbeelding van die proses.



Figuur 1. Die intelligensieproses

Volgens Pirolli en Card (2005:2–3) bestaan die proses uit die versameling van inligting, die verwerking daarvan in 'n vorm wat ontleiding vergemaklik, die ontwikkeling van insig, en die skep van 'n kennisproduk. Die datavloei verteenwoordig die omskakeling van inligting uit 'n ruwe toestand na 'n vorm waar kundigheid toegepas kan word en dan na 'n verwerkte vorm wat geskik is vir kommunikasie aan ander wat nie kenners van die betrokke veld is nie.

Binne 'n akademiese opset kan *eksterne bronne* alles behels wat tot die navorser se beskikking is: artikels en resensies wat aanlyn beskikbaar is, koerantuitknipsels, boeke, en selfs video's van onderhoude op YouTube (sien byvoorbeeld Mouton en Marais 1990:79). Die *schoendoos* is dan 'n versameling van daardie bronmateriaal, dit waarop die navorser sy studie gaan baseer – waarskynlik 'n lêer (op die rekenaar). Hieruit kan 'n *werksdokument* saamgestel word met opsommende notas vanuit die *schoendoos*. Wanneer dieselfde werksdokument vorm aanneem soos gedagtes georden word en 'n betoog opgebou word, kan hierna verwys word as 'n *skema*. Met verdere skawing ontwikkel die *skema* in 'n *hipotese*, en

uiteindelik in 'n finale produk: 'n artikel of boek wat gereed is vir voorlegging vir publikasie.⁵

Twee kringlope kan hiervolgens onderskei word: 'n opsporingskringloop wat hoofsaaklik gemoeid is met die opspoor en filtrering van bronne, en 'n sinmakingskringloop wat hoofsaaklik die vormgewing van die data behels. Patterson, Roth en Woods (2001) het waargeneem dat ontleders geneig is om te begin met 'n breë stel dokumente, byvoorbeeld een wat deur 'n hoë-herroeping- / lae-presisie-navraag opgespoor is (byvoorbeeld deur meer algemene soekterme), en dan die bronne verminder het tot kleiner, hoër-presisie-datastelle, voordat die dokumente gelees en ontleed is. Terugvoering is hier belangrik, aangesien die ontleder/navorsers in 'n voortdurende “dialoog” met sy inligting verkeer (NVAC 2005:73): soos meer inligting oor 'n onderwerp bekend word, kan beter soekterme gebruik word om beter resultate op te lewer (lae herroeping / hoë presisie).

Pirolli en Card (2005:5) verwys na koste – “the absolute and relative time costs of information operations” – en in 'n vroeëre publikasie (1999:7) verwys die skrywers na twee vorme van koste wat hier ter sake is: benewens hulpbronskoste (die tydsinset wat 'n inligtingsoektog verg) is daar 'n verdere “geleentheidskoste”: dit wat prysgegee word (ander geleentheid) deur die huidige aktiwiteit te voltooi. In eenvoudige terme: hoe langer dit neem om navorsing te doen, hoe minder tyd is beskikbaar vir klasgee en studentenavrae, gesinne, ensovoorts. Die skrywers (2005:5) voer aan dat tegnologie deur die verloop van die hele intelligensieproses aangewend kan word om die “koste” (tydsinsette) te beïnvloed. Hoe meer inligting betrek word, hoe kleiner is die kans dat belangrike inligting misgekyk word,⁶ maar dit gaan natuurlik gepaard met die prys van hoër tydsinsette in beide die opsporings- en sinmakingskringloop.

Die ideaal is dus om groter datastelle te kan ondersoek, maar dit kan slegs met inligtingstegnologie vermag word. Hossain, Andrews, Ramakrishnan en North (2011) skryf dat waar sommige rekenaarprogramme, soos IN-SPIRE, Jigsaw, ThemeRiver, en NetLens fokus op die versamelingsfase (opsporingskringloop), programme soos i2 Analyst's Notebook, Sentinel Visualizer, Entity Workspace en Palantir op die verwerkingsfase (sinmakingskringloop) fokus. Future Point Systems se Starlight Visual Information System (VIS) skakel egter beide kringlope, aangesien die Data Engineer-komponent meer gerig is op die opsporingskringloop, terwyl die sinmakingskringloop deur die Starlight-program gedek word (dit kan as pakket aangeskaf word). Let egter daarop dat Pirolli en Card (2005:6) beklemtoon dat tegnologie 'n ondersteunende rol speel: dit is die kennis en ondervinding van die ontleder (navorsers) wat deurslaggewend en onontbeerlik is (sien ook NVAC 2005:39; Hossain e.a. 2011:22): programmatuur is bloot 'n stuk gereedskap wat ontleding vergemaklik; dit kan dit nie vervang nie.

Die vier fases van die proses word nou in verskillende onderafdelings toegelig.

3. Beplanning

Die beplanningsfase van die intelligensieproses kan vergelyk word met die navorsingsontwerp (sien byvoorbeeld Mouton en Marais 1990:59–73). Aangesien tegnologie nie ’n deurslaggewende rol in hierdie fase speel nie, word hier volstaan daarmee om te noem dat dit natuurlik ’n noodsaaklike fase in sowel die intelligensie- as die navorsingsproses is, want “once one knows what he or she is looking for, data sourcing, collection methods, and the entire research process becomes more efficient and focused” (Bose 2008:515). Wikipedia is hier bruikbaar vir navorsingsdoeleindes, en een student het opgemerk: “Wikipedia is my presearch tool” (Head en Eisenberg 2009:12) – die “eerste kontak” met ’n onderwerp. Wikipedia is nuttig wanneer ’n nuwe onderwerp ontdek word, en skakels kan gevolg word wat natuurlik tot meer gesaghebbende bronne lei.

4. Versameling

Die samestelling van bruikbare datastelle is ’n sleutelkomponent in militêre intelligensie (Graham en Carley 2008:1), en so ook in enige vorm van navorsing. In bostaande diagram word voorgestel dat brondokumente in ’n “skoendoos” versamel word waaruit verdere verwerking plaasvind. Hierdie versameling is dus ’n voorlopige samevoeging van brondokumente vanuit die globale digitale korpus in ’n meer bruikbare hoeveelheid wat volgens ’n tema versamel word. Bose (2008:516) onderskei tussen passiewe versameling, “supporting ongoing informational needs”, en aktiewe versameling, “surveying knowledge domains and targeting specific questions”.

4.1 *Passiewe versameling*

Bose (2008:518) skryf dat daar soms na die passiewe versameling van inligting verwys word as “inligting-stoot” (“information push”), vergelykbaar met wat McKee, Koltutsky en Vaska (2009:3) “current awareness alerting” noem, terwyl aktiewe versameling van inligting die intrek van inligting in die organisasie se sisteem behels. Passiewe versameling behels die opstel van die nodige infrastruktuur sodat relevante inligting die organisasie binnestroom, byvoorbeeld die intekening op relevante nuusbriewe, of die opstel van rekenaarprogrammatuur om die internet outomaties te monitor vir nuwe inligting en dit dan in die navorser se skoendoos of databasis in te trek. Dit het die primêre voordeel dat ’n wye verskeidenheid onderwerpe gedek kan word in plaas van die aktiewe versameling wat gewoonlik geskied in reaksie op ’n spesifieke aanvraag en doelgerigte inligting oplewer. In die geval van die Nasionale Afrikaanse Letterkundige Museum en Navorsingsentrum (NALN) was die opbou en instandhouding van die knipselversameling juis so ’n “passiewe-versameling”-aktiwiteit (slegs in Bose se sin van die term!): koerante, tydskrifte en joernale is afgelewer, deurgegaan vir relevante artikels, en die artikels dan uitgeknipt, geïndekseer en op die databasis ontsluit. In die 21ste eeu bestaan daar ’n wye verskeidenheid elektroniese hulpmiddele om passiewe versameling van digitale inligting te bespoedig, byvoorbeeld Starlight Data Engineer, wat deel

uitmaak van die VIS-pakket. Data Engineer kan onder andere op die volgende maniere opgestel word om inligting binne die stelsel in te trek en te verwerk:

- Die “Filesystem – Directory Monitor” ondersoek voortdurend een of meer geselekteerde lêers op die rekenaar en kopieer enige nuwe dokumente wat bygevoeg word na spesifieke lêers, byvoorbeeld rakende ’n spesifieke onderwerp.
- Die “Internet – RSS Feed Harvester”-module ondersoek voortdurend gespesifiseerde plekke op die internet vir nuwe stories en skakel dan outomaties die nuwe stories in individuele XML-rekords om vir verdere ontleding. Hierdie funksie kan gebruik word om die navorser op datum te hou met nuus en tendense met behulp van internet-inligtingsbronne, en kan onbepaald in die agtergrond loop.
- Die “Internet – Web Crawler”-module deursoek die internet vanaf ’n gegewe beginpunt, en laai sowel die huidige webblad as al die skakels op die webblad af. Dan gaan dit deur elke skakel wat versamel is en laai ook daardie bladsye en skakels af. Dit gaan voort met hierdie proses totdat die gebruiker-gespesifiseerde diepte bereik is. ’n Voorbeeld: as ’n mens begin by www.futurepointsystems.com en gaan tot ’n diepte van 3 (3 grade van die oorspronklike bladsy), sou jy nou honderde dokumente besit wat uit ’n verskeidenheid HTML-bladsye, beelde, teksdokumente, en PDF-dokumente bestaan.
- Die “Internet – YouTube Search”-module stel ’n mens in staat om inligting oor video’s in te samel, soos die titel, beskrywing, publikasiedatum, laaste bywerking, gradering, die aantal kommentar, sleutelwoorde, kategorieë, lengte van die video, inligting oor die video-outeur, ouderdom, geslag, ligging, verwante skakels na die video-outeur, die aantal kere wat die video gekyk is, kommentaar oor die video, en inligting oor die mense wat kommentaar op die video gelewer het.

Heelwat ander programme stel ook die gebruiker in staat om onder andere webblaai outomaties te monitor vir bygewerkte inligting, asook om ’n hele internet-soektog outomaties by te werk. Dit beteken dat die navorser outomaties in kennis gestel word wanneer nuwe inligting oor ’n onderwerp beskikbaar is, byvoorbeeld oor ’n spesifieke skrywer. Sommige van sulke programme is gratis, terwyl ander ’n paar honderd rand kos:

- Copernic Agent (<http://www.copernic.com/en/products/agent>)
- ChangeDetect (<http://www.changedetect.com>)
- MetaProducts Offline Explorer (http://www.metaproducts.com/mp/offline_explorer.htm)
- Check&Get (<http://activeurls.com>)
- HTTrack (<http://www.httrack.com>).

Vir hierdie artikel is Offline Explorer Pro gebruik om Versindaba (<http://versindaba.co.za>) se tuisblad en eerste twee stelle skakels (die skakels vanaf die tuisblad en die skakels vanaf die skakels wat vanaf die tuisblad bereik is) af te laai en te indekseer. Webblaai kan dan op die rekenaar gestoor word vir argiveringsdoeleindes en verdere navorsing, met die voordeel dat daar sonder ’n internetverbinding deur datastelle gesoek kan word.

Ander programme wat spesifiek vir die akademiese milieu ontwerp is, is RefAware en IngentaConnect (McKee, Koltutsky en Vaska 2009), wat deurlopend die internet monitor vir nuwe inligting oor 'n gegewe onderwerp, en soos Copernic Agent dan die navorser in kennis stel van nuwe inligting wat gevind word. Sulke programme maak dit maklik om op hoogte te bly van die jongste ontwikkelinge op 'n navorser se gebied. Ander gratis opsies is ticTocs en Google Alerts.

4.2 Aktiewe versameling

Die versameling van akkurate, betroubare inligting binne 'n koste-effektiewe tydsraamwerk is een van die sleutels tot suksesvolle navorsing en word soms na verwys as inligtingsherwinning (Information Retrieval of IR): “the interdisciplinary science of searching for documents, for information within documents and for documents’ metadata in databases and on the World Wide Web” (Moisil 2009:25). Pirolli en Card (1999:11) beklemtoon dat die aktiewe versameling van inligting egter gedryf word deur die noodsaak om die soektog so koste-effektief as moontlik af te handel:

[O]ne may think of an information forager as an information predator whose aim it is to select information prey so as to maximize the rate of gain of information relevant to their task. These (sic) information prey might be relevant documents or document collections.

Pirolli en Card (2005:3) onderskei verder tussen versameling van inligting wat van onder na bo gedryf word, met ander woorde van data tot gevolgtrekkings (wat Mouton en Marais 1990:105 die induktiewe strategie noem), en versameling wat van bo na onder gedryf word, met ander woorde hipotese-gedrewe versameling (wat Mouton en Marais 1990:105 die deduktiewe strategie noem). In eersgenoemde doen die navorser 'n soektog oor relevante databasisse, doen navraag, ensovoorts, en stel dan 'n versameling dokumente saam in die skoendoos vir verdere verwerking. Wanneer dan deur hierdie dokumente gelees word, word daar uiteraard telkens nog inligting bekom wat die navorsers noodsaak om terug te keer na die versamelingsfase wanneer nuwe bronne ontdek word of teoretiese benaderings teëgekomp word. Dan word 'n aktiewe soektog deur die internet of op databasisse soos EBSCO Host gedoen deur sleutelwoorde, skrywer(s), publikasies of artikels se titels te gebruik.

'n Soektog wat van bo na onder plaasvind, kan geskied wanneer bestaande opvattinge bevraagteken of bevestig moet word (Pirolli en Card 2005:4). Hernude (aktiewe) soektogte lewer dan ondersteunende of ondermynende inligting op, verhoudinge word bevraagteken, ensovoorts. So kan 'n versamelingsoektog van bo na onder byvoorbeeld geskied wanneer literêre opvattinge verander, byvoorbeeld die herskrywing van die Afrikaanse literêre kanon⁷ om voorheen gemarginaliseerde stemme in te sluit.

Alhoewel daar vele soekenjins beskikbaar is waarmee oor die internet gesoek kan word, is Google een van die markleiers, “the world’s largest and most powerful search engine” (Noruzi 2005:171; sien ook Ripple 2006:98), wat teen 2003 reeds 'n werkwoord geword het. Vir akademiese doeleindes word <http://scholar.google.co.za> aanbeveel (maar nie daartoe

beperk nie), aangesien dit reeds 'n spesialisering van die algemene Google is wat op akademiese publikasies fokus, hetsy akademiese joernale, webblaaie of boeke. Hierdie wetenskaplike been van Google is ontwikkel deur die Indiese rekenaarwetenskaplike Anurag Acharya, en ontleed outomaties watter artikels die meeste na verwys word in bibliografieë, wat dit maklik maak vir die navorser om te bepaal wie die leiers op 'n bepaalde terrein is (Noruzi 2005:171).

4.3 Akkuraatheid en versameling

In sowel die intelligensie- as die navorsingskonteks is akkuraatheid van bronmateriaal natuurlik 'n sentrale gegewe en Krizan (1999:26–8) stel voor dat versameling 'n aantal faktore in ag moet neem. Eerstens moet die betroubaarheid van bronne bepaal en as sodanig aangedui word. Binne 'n akademiese navorsingskonteks sal byvoorbeeld voorkeur gegee word aan publikasies in geakkrediteerde akademiese joernale bo webblaaie, maar soos met militêre intelligensie is die skrywer van die inligting ook van belang: 'n aanlyn-resensie, geskryf deur 'n welbekende en gerekende letterkundige, kan byvoorbeeld as betroubaar gesien word, byvoorbeeld een wat op <http://versindaba.co.za> voorkom. Alhoewel die akademiese gemeenskap baie skepties is oor die internet, en veral Wikipedia, is daar heelwat navorsers wat let op die bruikbaarheid van Wikipedia binne 'n akademiese omgewing (Black 2008; Head en Eisenberg 2009; Tardy 2010), selfs al word dit slegs as wegspringplek gebruik. Die internet is nie onbetroubaar op sigself nie.

Tweedens stel Krizan (1999:26) voor dat voorkeur gegee word aan bronne wat nader aan die oorspronklike gebeurtenis is; in die geval van letterkundige studie is die lees van oorspronklike teoretici se werk soms verkieslik bo die lees van interpretasies daarvan. Hier is veral Google Books nuttig, aangesien vele welbekende boeke hierdeur beskikbaar is, en in 'n mate gratis aanlyn gelees kan word.

Toepaslikheid is 'n ander faktor wat volgens Krizan (1999:27) by die versameling van inligting die versamelingsproses behoort te beïnvloed. Alhoewel 'n individu 'n kenner op een gebied is, is hy nie noodwendig 'n gesaghebbende op 'n ander gebied nie. Soos met die gedrukte media verg elektroniese navorsing dus ook dat die skrywer en uitgewer in ag geneem word.

Dit is belangrik om te onthou dat die gebruik van die internet geensins die gesaghebbendheid van bronmateriaal ondermyn nie: dieselfde materiaal wat in druk verskyn, is meesal aanlyn beskikbaar, en dieselfde reëls geld as in gedrukte media: die skrywer en publikasie self verhoog die gesaghebbendheid van bronne, nie die medium nie.

5. Verwerking

Elke dokument het die potensiaal om die verhouding tussen entiteite binne 'n sisteem te ondersteun of te weerlê, en gevolglik het elke dokument, elke teks, waarde vir die

eindresultaat (Graham en Carley 2008:9). Hier word die navorser gekonfronteer met data-integrasie, hetsy as individu of as span, wanneer 'n wye verskeidenheid inligting as 'n sinvolle geheel oorsigtelik beskou moet word sodat patrone geïdentifiseer kan word. In hierdie fase word inligting omskep in intelligensie:

Analysis encompasses a systematic examination of relevant data, information, and knowledge collected, for applicability or significance, and the transformation of the results into actionable intelligence that will improve planning and decision making or will enable the development of strategies that offer a sustainable competitive advantage. (Bose 2008:513)

Militêre intelligensie staan in hierdie opsig weer eens aan die voorpunt van tegnologiese ontwikkeling. Science Applications International Corporation se Pathfinder-tekstontledingsagteware (<http://www.saic.com/contractcenter/gsa-it/pathfinder.html>), soos gebruik deur die VSA se Army National Ground Intelligence Center, kan byvoorbeeld glo binne slegs 'n paar minute deur meer as 500 000 dokumente sorteer en patrone identifiseer (Roberts 2011:8). Programmatuur soos Starlight VIS het soortgelyke vermoëns, en alhoewel daar geen vervanging is vir die lees van bronmateriaal deur 'n navorser nie, bestaan daar heelwat ander tegnologiese hulpmiddele om groter stelle dokumente onder die loep te neem, onder andere deur middel van data-ontginning en visualisasie, soos vervolgens in meer besonderhede verduidelik word. Selfs programmatuur wat die bestuur van bronmateriaal bespoedig, soos Zotero (<http://www.zotero.org/>) en Qiqqa (<http://www.qiqqa.com/>), kan 'n beduidende impak op die navorsingsproses hê deur die versameling en ontleding van brondokumente rekenaarmatig te bespoedig.

5.1 Data-ontginning en visualisasie

Volgens die VK se National Centre for Text Mining (aangehaal in Borgman 2010:8) is data-ontginning “the process of identifying patterns in large sets of data [...] to uncover previously unknown, useful knowledge” (sien ook Ceruti en McCarthy 2000:174). Bose (2008:523) skryf:

Text mining looks for things like word proximity and sentence structure, sifting through billions of text (sic), placing it into some common format (referred to as normalization), and filtering or ranking the text using statistical methods. Text mining is not the same thing as a search engine on the web. In a search, one tries to find what others have prepared. With text mining, one wants to discover new patterns, nuggets of knowledge that may not be obvious or known.

Visuele ontleding word gedefinieer as “the science of analytical reasoning facilitated by interactive visual interfaces” (NVAC 2005:4). Wanneer ons iets verstaan, sê ons in Afrikaans of Engels: “Ek sien wat jy bedoel” of “Do you see what I mean?” Dit is verwoording van die ingebore verband wat ons lê tussen visie, visualisasie en ons redenasie-prosesse (NVAC 2005:33). Programmatuur wat in visualisasie spesialiseer, poog dus om bestaande denkprosesse te ondersteun:

Visual representations and interaction technologies give users a gateway into their data, letting them see and understand large volumes of information at once. To facilitate analytical reasoning, visual analytics builds on the human mind's ability to understand complex information visually. (Thomas en Cook 2006:11; sien ook Roberts 2011:17.)

Daar bestaan talle produkte op die mark en Kirschenbaum (2007:4) noem hoe Spotfire (<http://spotfire.tibco.com>), TextArc (<http://www.textarc.org>) en ander al vrugbaar aangewend is binne die letterkunde, terwyl Athenikos (2009) beskryf hoe visualisasie deur middel van Prefuse (<http://prefuse.org>) kan help met die studie van filosofie. 'n Ander program wat spesifiek vir kwalitatiewe navorsingsdoeleindes ontwikkel is, is QSR se Nvivo (<http://www.qsrinternational.com/>), wat ook die ontleding van data deur middel van visualisering bemagtig, en veral ontwerp is om met multimedia te werk (video, oudio en teks). Kirschenbaum (2007:4) voer aan:

The importance of visualization as a means of accessing and studying the results of the text analysis cannot be over-emphasized. When we look at a painting or picture, we grasp the entirety of it within our optical field. The eye can easily move from one region of the image to the next, looking for patterns and correspondences which aid in interpretation. In the case of a novel (or even a short story or a long poem), however, we cannot hold the entirety of the text within our visual field. Indeed, the physical form of the codex book itself mitigates against this, as the text is arbitrarily broken up into discrete units divided by pages. Visualization, which essentially makes the text a picture, is capable of bringing a novel into focus as a unified visual event.

Voyeur (<http://voyeurtools.org>) is 'n gratis aanlyn-teksontledingsprogram wat deur Stefan Sinclair en Geoffrey Rockwell ontwikkel is as deel van die Hermeneuti.ca-projek. Ten einde die potensiaal van Voyeur te illustreer, is 'n korpus Afrikaanse akademiese artikels saamgestel wat alles insluit van *Stilet* wat elektronies beskikbaar is (2002–2010, 231 artikels) en alles van *LitNet Akademies* wat in PDF beskikbaar is (2008–2010, 92 artikels). Uit die ontleding van hierdie 326 artikels deur Voyeur, wat 2 625 634 woorde bevat (en 6 061 bladsye beslaan), sien 'n mens wat die Afrikaanse woorde met die grootste gebruiksfrekwensie is:

Tabel 1. Woordfrekwensies

Nommer	Woord	Aantal kere gebruik
1	die	164 379
2	van	85 417
3	en	64 556
4	in	58 334
5	'n	48 988

Hierdie inligting dra egter nie veel by tot 'n beter begrip van die Afrikaanse literêre sisteem nie, en gevolglik is hierna eerder spesifieke woorde nageslaan. 'n Groter korpus⁸ bestaande uit wat digitaal beskikbaar is⁹ van *Literator* (507 artikels), *Tydskrif vir Geesteswetenskappe*

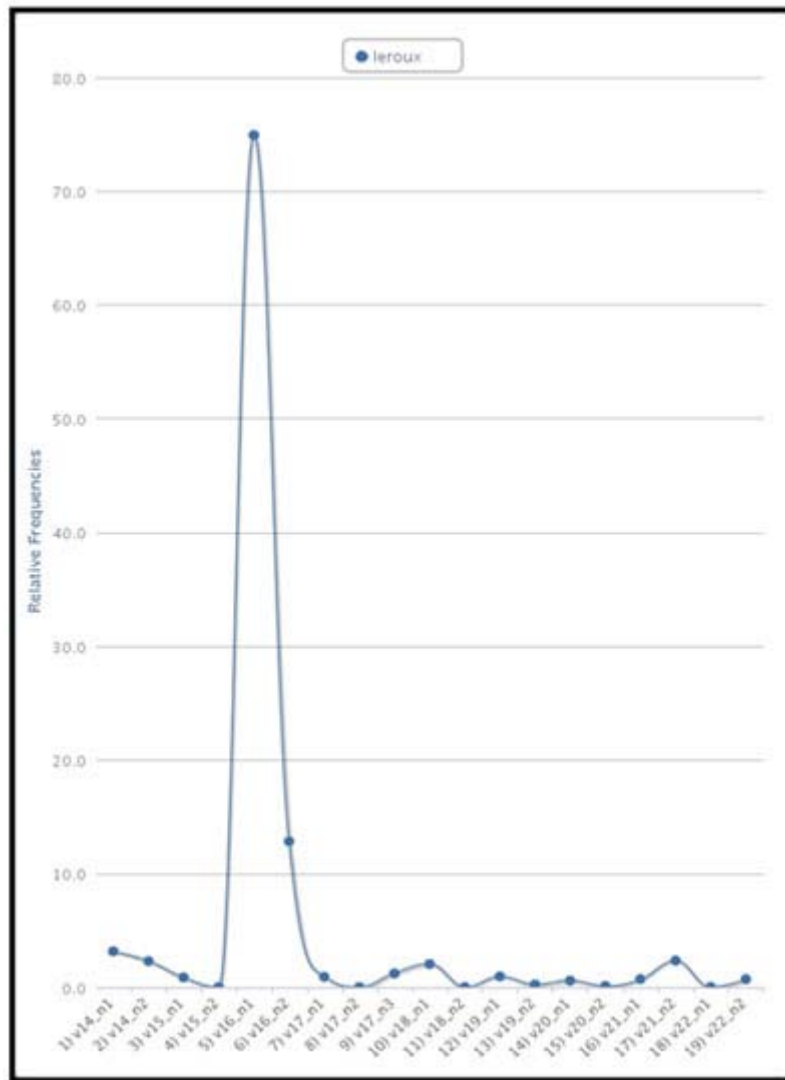
(426), *Tydskrif vir Letterkunde* (409) en *Tydskrif vir Literatuurwetenskap* (274) is saamgestel, met 1 942 artikels en 11 012 962 woorde. In hierdie korpus kom J.C. Kannemeyer se naam 953 keer voor en H.P. van Coller s'n 1 172 keer.

Teksontleding is reeds aangewend in die studie van literêre prestige (Taboada, Gillies, McFetridge en Outtrim 2008), en in navolging hiervan en van Van Coller (2004) is 'n aantal Hertzogprysweners nageslaan:¹⁰

Tabel 2. Aantal kere wat skrywers genoem word

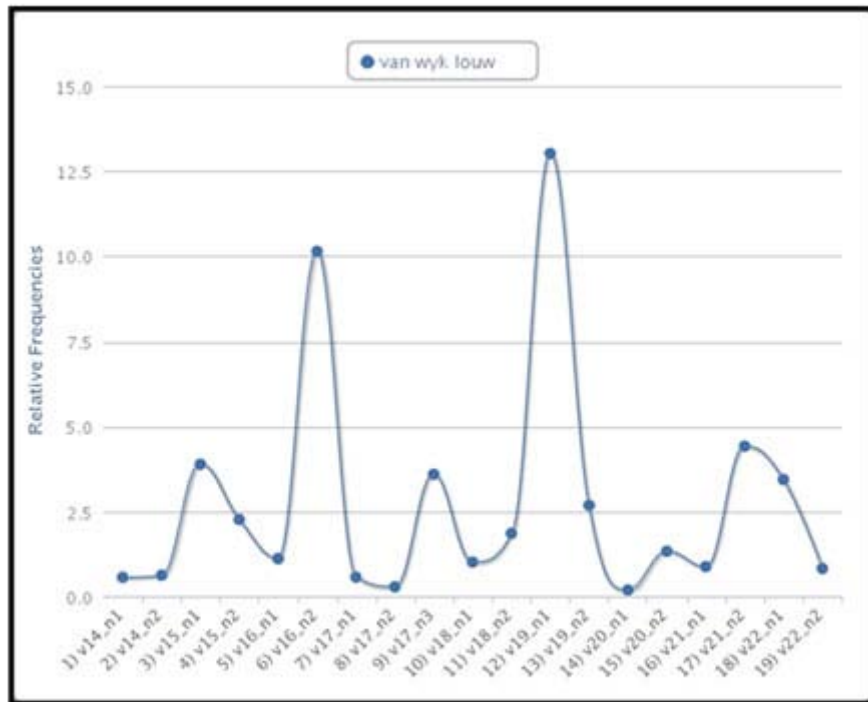
Skrywer	Aantal kere genoem
André P. Brink	2 729
Antjie Krog	2 212
Breyten Breytenbach	2 198
N.P. Van Wyk Louw	1 587
Ingrid Winterbach	1 054
C.L. Leipoldt	637
Hennie Aucamp	566
Elisabeth Eybers	424
M.E.R.	314
C.J. Langenhoven	303
D.F. Malherbe	296
Wilma Stockenström	283
Uys Krige	256
Totius	221
W.E.G. Louw	149
Sheila Cussons	133
Boerneef	109
Jochem van Bruggen	80
J.F.W. Grosskopf	53
H.A. Fagan	35
Sangiro	15

Uit bostaande is dit duidelik watter skrywers die sigbaarste is in akademiese tydskrifte, hetsy in hul rol as skrywers of as kritici. Voyeur skep ook outomaties grafieke waardeur skrywers se teenwoordigheid in publikasies visueel voorgestel word, byvoorbeeld Etienne Leroux in die bogenoemde korpus van *Stilet*-artikels:



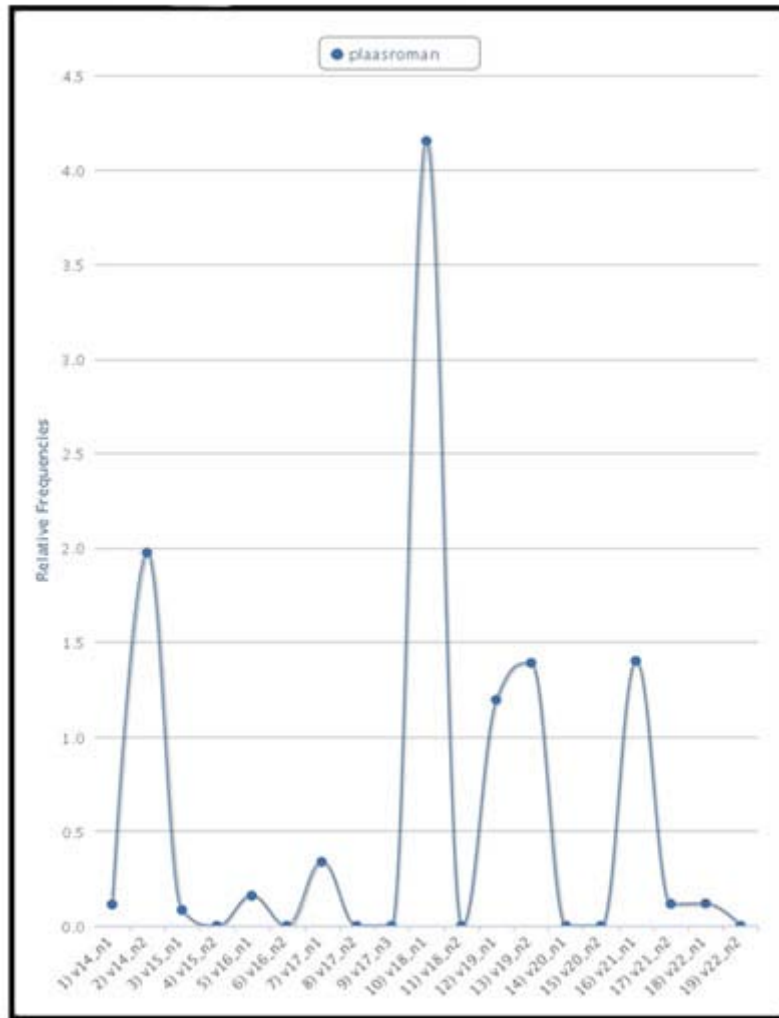
Figuur 2. Frekwensie waarteen Leroux in *Stilet* genoem word

In hierdie grafiek is dit duidelik dat *Stilet* 16(1) spesiale aandag aan Leroux gewy het (74,91 per 10 000 woorde), en dit is juis in hierdie uitgawe dat Van Coller (2004) vra of Leroux 'n vergete skrywer is. Was Leroux daarna vergete? *Stilet* 21(2) (2009) noem Leroux met 'n frekwensie van 2,32 per 10 000 woorde; sien daarteenoor byvoorbeeld N.P. Van Wyk Louw in dieselfde korpus:



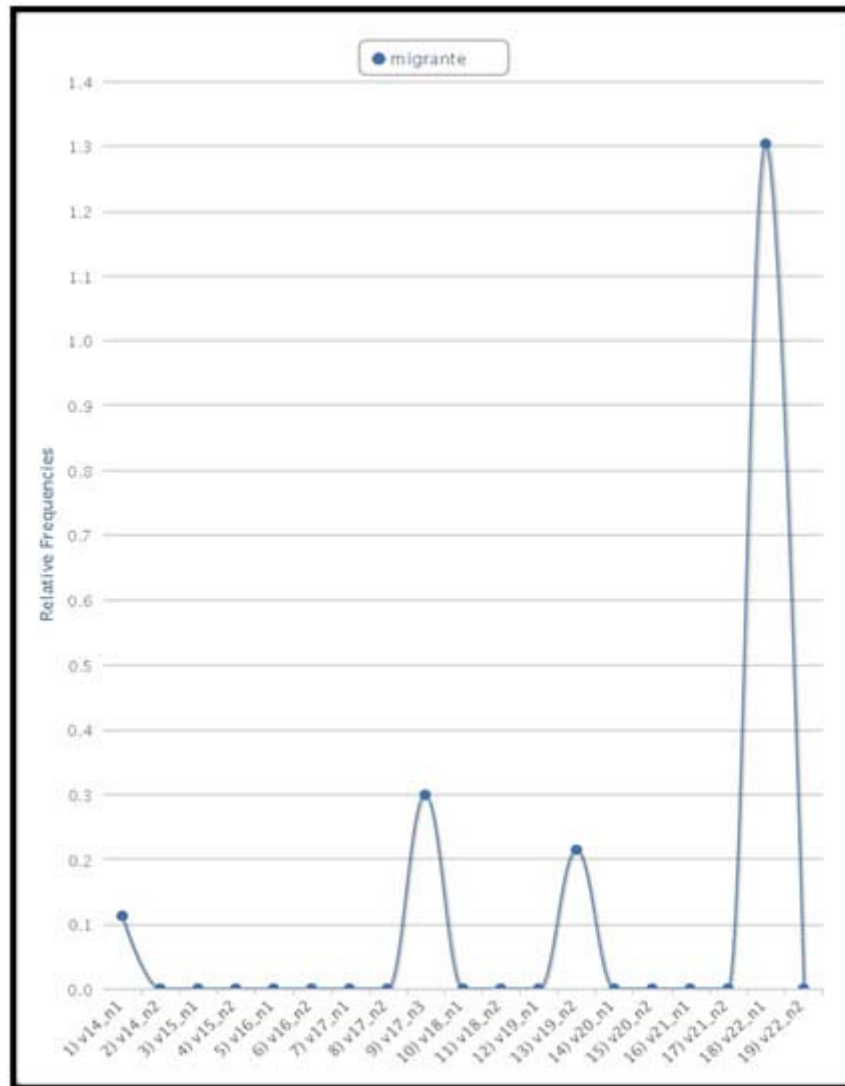
Figuur 3. Frekwensie waarteen Van Wyk Louw in *Stilet* genoem word

Tendense in die letterkunde kan ook deur middel van sulke grafieke ondersoek word. Kyk byvoorbeeld na die aantal verwysings na die plaasroman in dieselfde korpus:



Figuur 4. Frekwensie waarteen “plaasroman” in *Stilet* genoem word

Vergelyk dit met migranteliteratuur:



Figuur 5. Frekwensie waarteen “migrante” in *Stilet* genoem word

Daar bestaan natuurlik heelwat probleme met so ’n kwantitatiewe werkswyse. Hier is konteks weggelaat; die aantal kere wat ’n skrywer in een artikel genoem is, is nie verreken nie, en dergelike. My oogmerk is nie om iets voor te stel oor die belangrikheid van spesifieke skrywers of tendense in die Afrikaanse literêre sisteem nie, maar eerder om die potensiaal van inligtingstegnologie te illustreer, veral wanneer ’n meer oorsigtelike blik oor die letterkunde verkry wil word. Toekomstige navorsing kan hierdie kwessies in meer diepte ondersoek. Soos wat Van Coller (2004) die posisie van Etienne Leroux binne die Afrikaanse literêre sisteem navors, kan inligtingstegnologie byvoorbeeld gebruik word om dieselfde studie vinniger te doen, en daarom kan ander skrywers se loopbane vergelykend gebruik word, soos trouens hier gedoen word met Van Wyk Louw. Ook kan inligtingstegnologie die navorser in staat stel om tendense in die Afrikaanse letterkunde te identifiseer, wat tot verdere ontleding kan lei, soos die geval hier met *migrante* – dit help die navorser om te sien waaraan hy aandag kan gee, waar gapings is, waar die meeste inligting oor ’n onderwerp is, ensovoorts. Sou ’n nagraadse student byvoorbeeld rigting soek oor waar om te begin

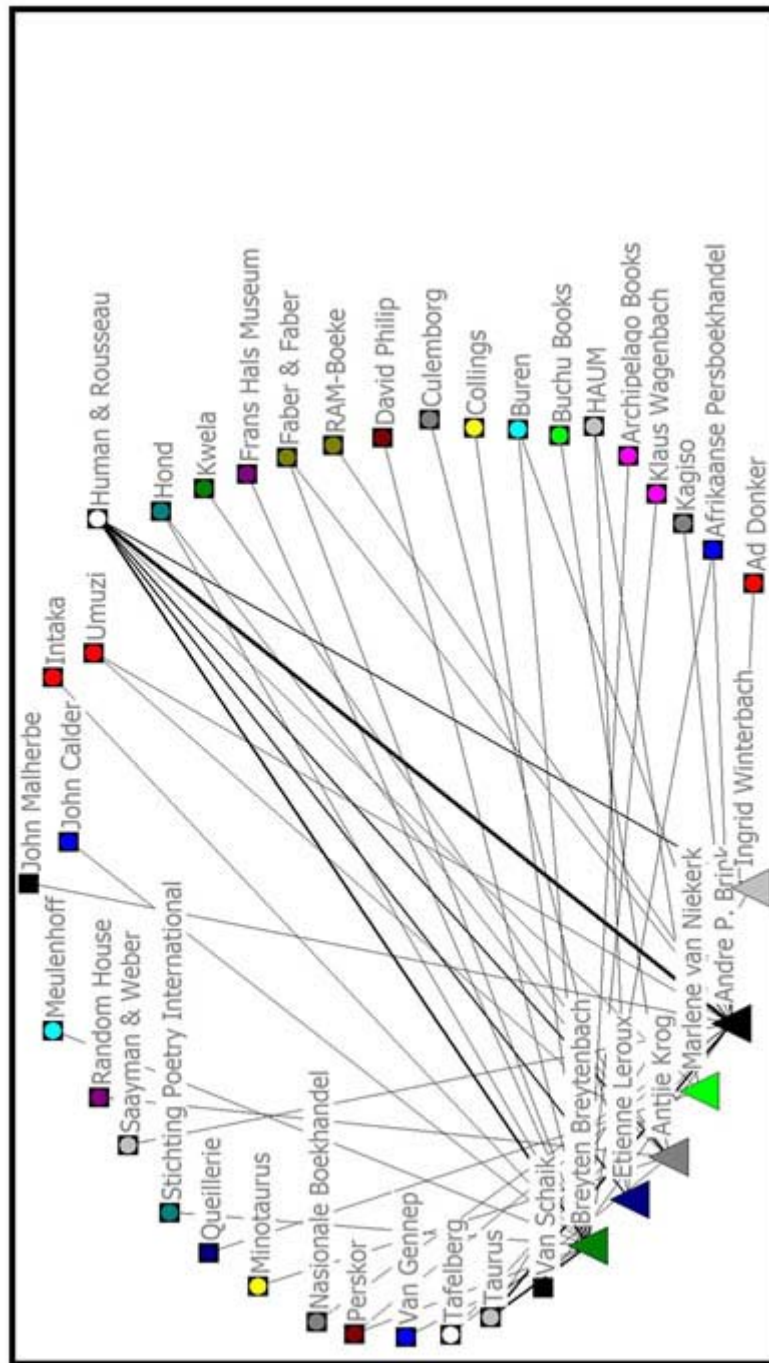
navorsing doen oor migranteliteratuur, is dit duidelik uit figuur 5 dat *Stilet* 22(1) 'n goeie beginpunt sou wees.

Sosiale-netwerk-analise (Social Network Analysis – SNA) is veral geskik om die verhoudings tussen entiteite binne 'n sisteem visueel voor te stel (Hanneman en Riddle 2005), en is konseptueel identies aan sisteemteorie, aangesien SNA die studie van verhoudings binne entiteite behels eerder as die entiteite self. Aangesien hierdie programmatuur wyd aangewend word in veral teenterrorisme militêre intelligensie, is heelwat programme reeds ontwikkel, byvoorbeeld:

- Future Point Systems se Starlight VIS (<http://www.futurepointsystems.com>)
- Pacific Northwest National Laboratory se In-Spire (<http://in-spire.pnnl.gov>)
- Visual Analytics se i2 Analyst's Notebook (<http://www.i2group.com/us>)
- FMS se Sentinel Visualizer (<http://www.fmsasg.com/products/SentinelVisualizer>).

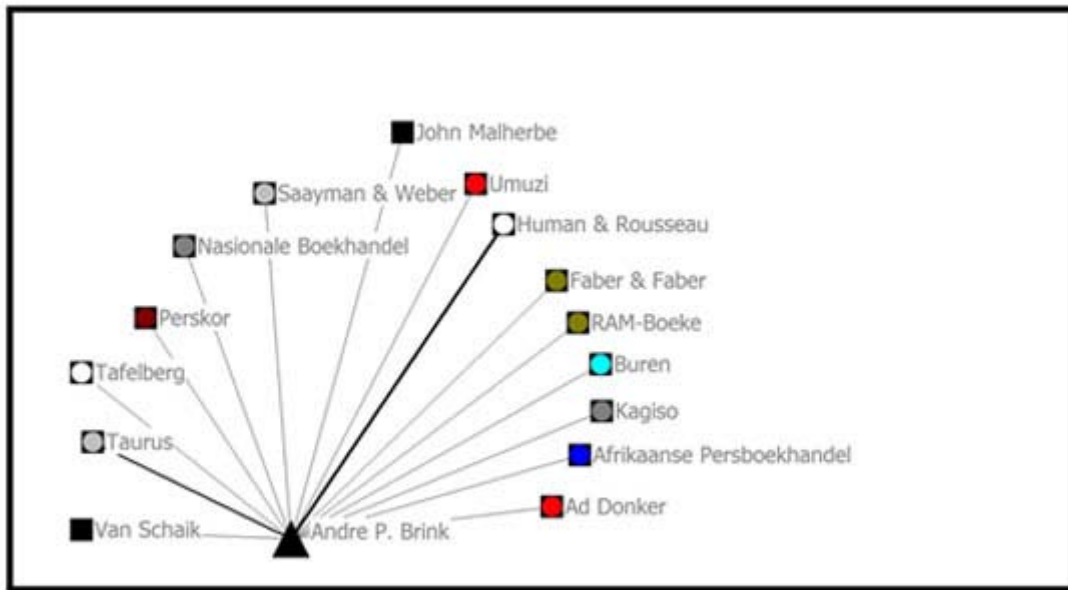
Ongelukkig vir die akademiese navorser is hierdie programme baie duurder as wat departementele begrotings toelaat: In-Spire kos byvoorbeeld nagenoeg R200 000, en Starlight VIS verg 'n minimum van vyf lisensies, wat beteken dat die program uiteindelik R450 000 kos. Sentinel Visualizer kos minder as 'n tiende van In-Spire, maar anders as Starlight en In-Spire kan dit slegs met gestruktureerde data werk, met ander woorde .xml, Microsoft Access of Excel, en nie met PDF- of Microsoft Word-dokumente nie. Dit gebeur egter selde dat Afrikaanse letterkunde datastelle in 'n gestruktureerde formaat teëkom, wat die gebruik van so 'n program beperk. Nietemin, as die navorser bereid is om die tyd te bestee en data in 'n gestruktureerde formaat te kry, kan SNA-programmatuur wat slegs met gestruktureerde data werk, met vrag gebruik word, soos geïllustreer deur die gebruik van Netdraw.

Netdraw (<http://www.analytictech.com/Netdraw/netdraw.htm>) is 'n gratis program wat ontwikkel is deur Steve Borgatti. Netdraw spesialiseer in SNA, en alhoewel data voorberei moet word (in die regte formaat moet wees in 'n teksdokument (.txt)), skep dit die geleentheid om verhoudings tussen entiteite binne die Afrikaanse literêre sisteem grafies voor te stel en verder te ontleed. Ter illustrasie is 'n aantal bekende Afrikaanse skrywers, in hierdie geval Etienne Leroux, Breyten Breytenbach, André P. Brink, Marlene van Niekerk, Ingrid Winterbach en Antjie Krog, se publikasies uitgebeeld in 'n poging om hul “uitgeweryprofiel” (Venter 2006:470-1) duideliker te kan ontleed. Van Coller (2004:1) noem dat skrywers wat die etiket “legitiem” dra én gekanoniseer is, onder andere by bepaalde uitgewerye publiseer, en die vraag is waar hierdie “legitieme” skrywers publiseer. In die volgende figuur word die betrokke skrywers met driehoeke aangedui, en hul uitgewerye met vierkante:

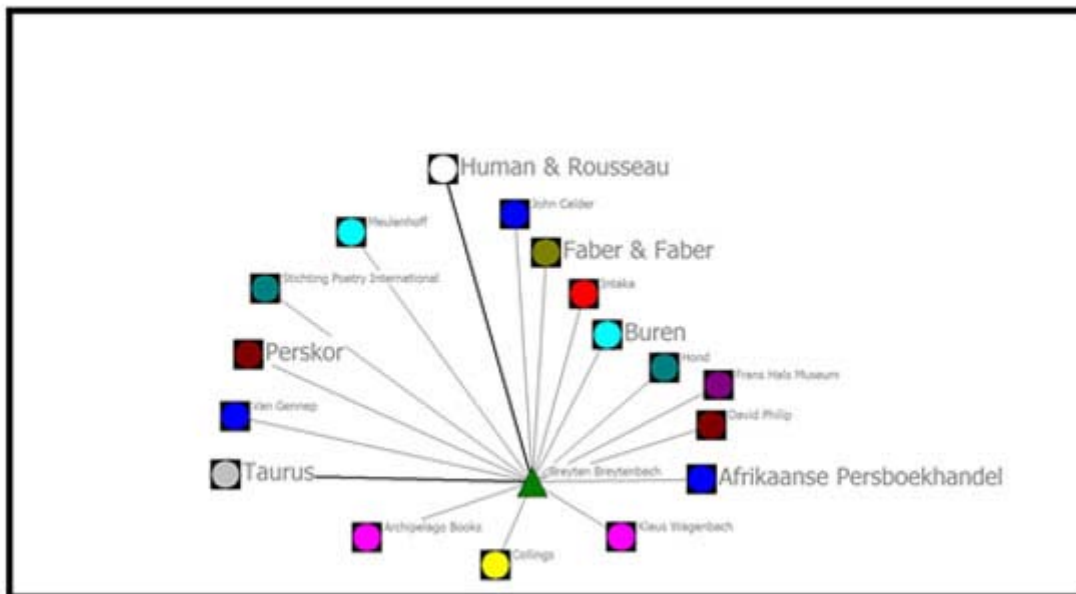


Figuur 6. Uitgeweryprofile van skrywers

Die dikte van verbindingslyne dui op hoeveel boeke 'n skrywer by 'n uitgewery gepubliseer het: dikker lyne verteenwoordig dus 'n sterker verbintenis. In hierdie visuele voorstelling is dit duidelik dat Human & Rousseau die sentrale uitgewery in hierdie korpus is – 'n sleutelentiteit – omdat hierdie skrywers in die eerste plek hul sterkste verbintenis (dikste lyn) met Human & Rousseau het, en in die tweede plek omdat dit die enigste uitgewery is wat met al ses skrywers verbind word. Skrywers se verbintnisse met uitgewerye oor die algemeen kan ook afsonderlik vergelyk word, byvoorbeeld Brink en Breytenbach se profile:



Figuur 7. Uitgeweryprofiel van Breyten Breytenbach

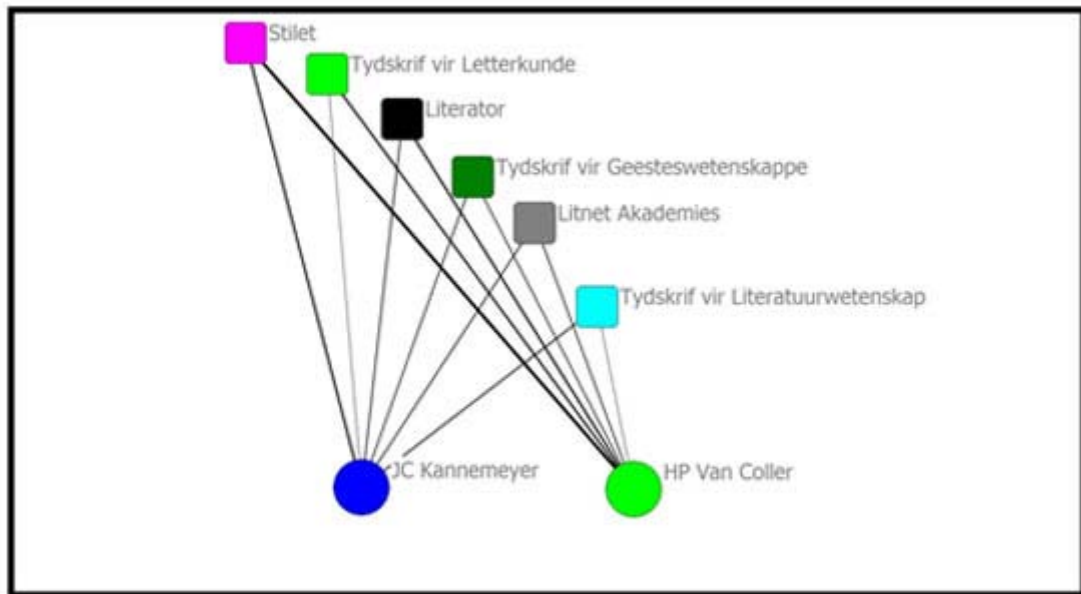


Figuur 8. Uitgeweryprofiel van André P. Brink

Beide skrywers se sterkste verbintenis blyk met Human & Rousseau te wees. Soos voorheen wil ek nie hier tot wetenskaplike gevolgtrekkings kom nie: ek wil bloot illustreer hoe tegnologie aangewend kán word. Ook is daar heelwat meer wat met so 'n program ontleed kan word; hierdie is bloot een manier waarop dit aangewend kan word.

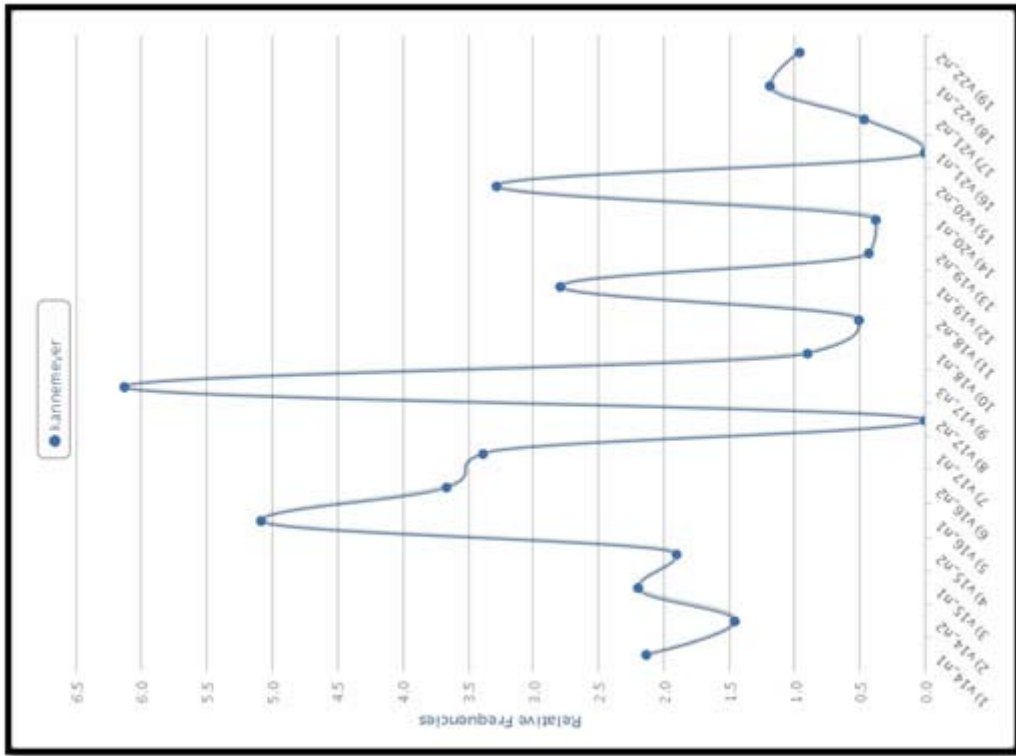
'n Ander, meer kontroversiële, manier om inligtingstegnologie aan te wend, is om letterkundiges te vergelyk. As vermeldings van J.C. Kannemeyer en H.P. van Coller se name

in *Stilet*, *Literator*, *Tydskrif vir Geesteswetenskappe*, *Tydskrif vir Letterkunde* en *Tydskrif vir Literatuurwetenskap* met behulp van Adobe Acrobat X Pro vergelyk word en die data vervolgens verwerk word in 'n vorm wat in Netdraw gebruik kan word, is dit duidelik dat albei die sigbaarste is in *Stilet*, aangesien die dikste swart lyne van beide letterkundiges na *Stilet* loop:

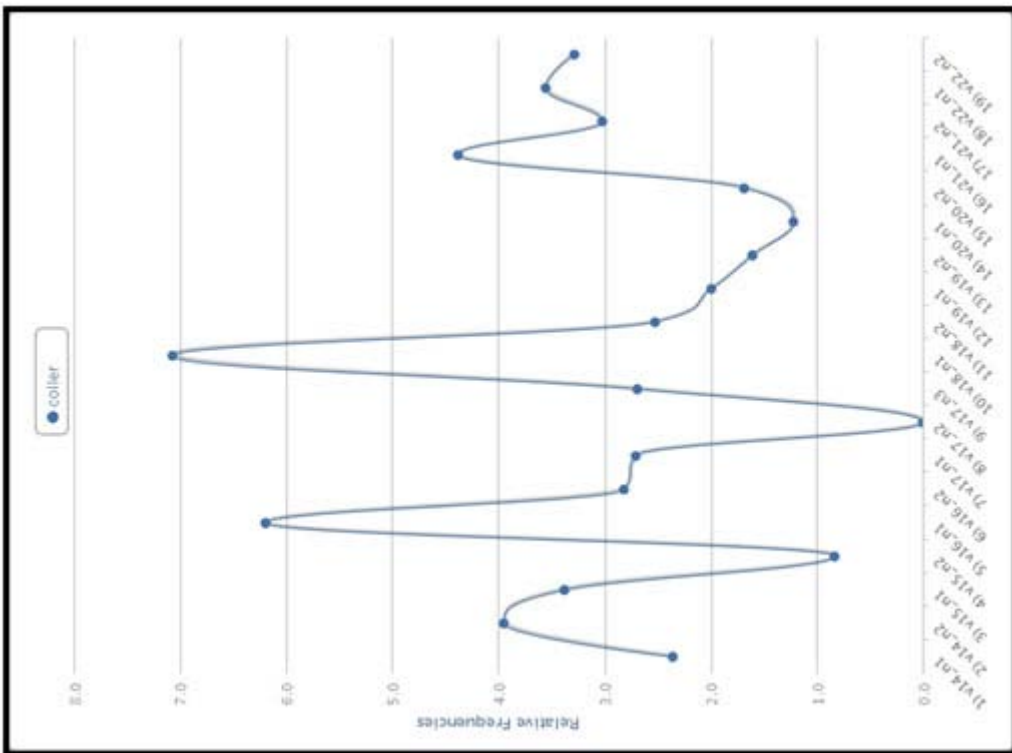


Figuur 9. Vermeldings van Kannemeyer en Van Coller

Die vraag ontstaan dus hoe hul vermeldings in *Stilet* lyk met behulp van Voyeur se grafieke:



Figuur 10. Frekwensie van vermeldings in *Stilet*: Kannemeyer



Figuur 11. Frekwensie van vermeldings in *Stilet*: Van Coller

Visualisasie skep die geleentheid om groot hoeveelhede data met 'n enkele oogopslag te bekyk, waarna afleidings gemaak kan word of tendense indringender ondersoek en verklaar kan word. Dit kan tendense en patrone oopvlek wat verdere navorsing prikkel wanneer vrae uitgewys word wat voorheen nie eens gevra is nie – Rumsfeld se “unknown unknowns”. Ook lei die integrasie van visualiseringshulpmiddele binne die sinmakingskringloop van die intelligensieproses tot die terugkeer na bronnemateriaal: Hoekom die skerp styging in vermeldings van Leroux in *Stilet* 16(1), of die migranteliteratuur in 22(1)? Alhoewel visualisasie dus op kwantitatiewe metodes berus, bespoedig dit ook kwalitatiewe navorsing, aangesien minder tyd spandeer kan word om tendense te identifiseer, wat meer tyd beskikbaar stel om tendense te ontleed en te verklaar. In toekomstige navorsing kan literatuurgeskiedenis vir hierdie doeleindes gedigitaliseer word, sodat skrywers se literêre prestige oor die jare wetenskaplik objektief, verifieerbaar, en herhaalbaar bestudeer kan word – 'n “harde”, kwantitatiewe benadering tot geesteswetenskaplike navorsing.

6. Verspreiding

Verspreiding behels die finalisering van die sogenaamde intelligensieprodukt (gewoonlik 'n verslag) en die aflewering daarvan by die “kliënt”. Binne 'n akademiese opset beteken *verspreiding* natuurlik die publikasie en disseminasie van navorsing, maar aangesien die klem in hierdie artikel op versameling en verwerking val, en aangesien e-pos en woordverwerkingsprogrammatuur alombekend is, word hier volstaan met om te noem dat inligtingstegnologie in hierdie fase reeds in 'n groot mate in algemene gebruik is.

7. Ten slotte

Treverton, Jones, Boraz en Lipsy (2006:25) voer aan: “If it is to stay relevant, intelligence needs to be like human vision – binocular, providing depth to what we can be seen in front; and peripheral, which provides warning.” Die skrywers voer dus aan dat inligting dieper en wyer ontleed behoort te word binne 'n militêre-intelligensie-konteks, maar dieselfde geld vir akademiese navorsing en ook die studie van die letterkunde. Veral die sisteemteorie poog om breër na die letterkunde te kyk, wat noodwendig beteken dat groter datastelle meer oorsigtelik beskou moet word, en veral hiervoor is inligtingstegnologie bowenal geskik, spesifiek programmatuur binne SNA.

Hierdie artikel het idees vanuit militêre intelligensie, veral die intelligensieproses/-siklus en data-ontginning en -visualisasie benut ten einde aan te toon hoe navorsing met inligtingstegnologie geïntegreer kan word. Veral Pirolli en Card (1999) se siening van die navorser/ontleder as 'n “inligtingsroofdier” is 'n bruikbare manier om na navorsing te kyk binne die huidige universitêre opset, waar begrotings en tyd al hoe meer beperk word: meer uitsette kan gelewer word met 'n kleiner inset, mits inligtingstegnologie aangewend word.

Ek hoop dat hierdie artikel verdere aanwending van inligtingstegnologie sal aanwakker sodat navorsingsuitsette van Afrikaanse akademici sal toeneem, en ook dat die gebruik van inligtingstegnologie tot nuwe insigte binne die Afrikaanse akademiese studie in die geesteswetenskappe sal lei.¹¹

Bibliografie

- Aldrich, R.J. 2009. Beyond the vigilant state: Globalisation and intelligence. *Review of International Studies*, 35:889–902.
- Athenikos, S.J. 2009. Interactive visualization and exploration of information on philosophers (and artists, scholars and scientists) in an e-learning portal for Digital Humanities. Symposium on Interactive Visual Information Collections and Activity (IVICA). Austin, 19 Junie. http://research.cis.drexel.edu:8080/sofia/pub/IVICA2009_Paper.pdf (27 Februarie 2012 geraadpleeg).
- Besser, H. 2004. The past, present, and future of digital libraries. In Schreibman e.a. (reds.) 2004.
- Biermann, J., R. Korsnes, J. Rohmer en C. Ünderger. 2004. *From unstructured to structured information in military intelligence – some steps to improve information fusion*. Ongepubliseerde referaat gelewer by RTO SCI Symposium on Systems, Concepts and Integration (SCI) Methods and Technologies for Defence Against Terrorism. RTO-MP-SCI-158.
- Bingham, A. 2010. The digitization of newspaper archives: Opportunities and challenges for historians. *Twentieth Century British History*, 21(2):225–31.
- Black, E.W. 2008. Wikipedia and academic peer review: Wikipedia as a recognised medium for scholarly publication? *Online Information Review*, 32(1):73–88.
- Block, M. 2002. My rules of information, *ExLibris*, 10(1). <http://infotoday.com/searcher/jan02/block.htm> (4 Januarie 2012 geraadpleeg).
- Borgman, C.L. 2010. The digital future is now: A call to action for the humanities. *Digital Humanities Quarterly*, 3(4):1–30.
- Bose, R. 2008. Competitive intelligence process and tools for intelligence analysis. *Industrial Management & Data Systems*, 108(4):510–28.
- Brei, W.S. 1996. *Getting intelligence right: The power of logical procedure*. Geleentheidspublikasie Nommer Twee. Washington: Joint Military Intelligence College.
- Ceruti, M.G. en S.J. McCarthy. 2000. Establishing a data-mining environment for wartime event prediction with an object-oriented command and control database. *Derde IEEE*

Internasionale Simposium oor Objekgeoriënteerde Sinchroniese verspreidingsprogrammering (Object-Oriented Real-time distributed Computing), Newport Beach, 15–17 Maart.

Darvill, D. 2011. Visual analytics: Visually exploring masses of data. *Newsletter of the Association of Canadian Ergonomists*, Lente-uitgawe, ble. 5–7.

Dawson, A. 2010. Academic freedom and the digital revolution. *AAUP Journal of Academic Freedom*, 1:1–19.

Even-Zohar, I. 1990. Polysystem Studies, *Poetics Today*, 11(1):1–94.

Geyer, F. (red.). 1996. *Alienation, ethnicity, and postmodernism*. Londen: Greenwood.

Google search basics: Basic search help 2010.

<http://www.google.com/support/websearch/bin/answer.py?hl=en&answer=134479> (13 Mei 2010 geraadpleeg).

Graham, J.M. en K.M. Carley. 2008. Intelligence database creation and analysis: Network-based text analysis versus human cognition. 41ste Hawaise Internasionale Kongres oor Sisteemwetenskap, Waikoloa, 7-10 Januarie.

Hanneman, R.A. en M. Riddle. 2005. *Introduction to social network methods*. <http://faculty.ucr.edu/~hanneman/nettext> (15 November 2011 geraadpleeg).

Head, A.J. en M.B. Eisenberg. 2009. *Finding context: What today's college students say about conducting research in the digital age*. Projek Inligtingsgeletterdheid Vorderingsverslag. Washington: Universiteit van Washington.

Hossain, M.S., C. Andrews, N. Ramakrishnan en C. North. 2011. Helping intelligence analysts make connections, scalable integration of analytics and visualization. Referate van die 2011 AAI Werkswikel, Kalifornië, Augustus 2011.

Jamali, H.R. 2008. What is not available online is not worth reading? *Webology*, 5(4). <http://www.webology.org/2008/v5n4/a63.html> (4 Januarie 2012 geraadpleeg).

Johnsson, B. 1981. The frame of reference: Poe, Lacan, Derrida. In Young (red.) 1981.

Kirschenbaum, M.G. 2007. The remaking of reading: Data mining and the digital humanities, NGDM 07. Nasionale Wetenskapstigting-simposium oor volgende generasie van data-ontginning en kuberontdekking vir innovasie. Baltimore, Maryland. Datum onbekend. <http://www.csee.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf>. (27 Februarie 2012 geraadpleeg).

Krizan, L. 1999. *Intelligence essentials for everyone*. Geleentheidspublikasie Nommer Ses. Washington: Joint Military Intelligence College.

- McKee, S., L. Koltutsky en M. Vaska. 2009. Introducing RefAware: A unique current awareness product. *Library Hi Tech News*, 26(9):1–6.
- Moayeri, M. 2007. Lost in cyberspace: Where to go? What to believe? *Webology*, 4(4). <http://www.webology.org/2007/v4n4/a47.html> (27 Februarie 2012 geraadpleeg).
- Moisil, I. 2009. Advanced methods for text retrieval. Verrigtinge van die 8ste WSEAS Internasionale Konferensie oor Kunsmatige Intelligensie, Kennis-ingenieurswese en Databasisse, Cambridge.
- Mouton, J. en H. Marais. 1990. *Basiese begrippe: Metodologie van die geesteswetenskappe*. Pretoria: RGN-Uitgewers.
- Neri, F. en M. Pettoni. 2009. Stalker, a multilingual text mining search engine for open source intelligence. *Advances in Soft Computing*, 53:35–42.
- Nicholas, D. 2008. The information-seeking behaviour of the virtual scholar: From use to users, *Serials*, 21(2):89–92.
- Noruzi, A. 2005. Google Scholar: The new generation of citation indexes. *Libri*, 55:170–80.
- NVAC (National Visualization and Analytics Center). 2005. *Illuminating the path: The research and development agenda for visual analytics*. Richland, Washington: National Visualization and Analytics Center.
- Patterson, E.S., E.M. Roth en D.D. Woods. 2001. Predicting vulnerabilities in computer-supported inferential analysis under data overload. *Cognition Technology and Work*, 3:224–37.
- Pirolli, P. en S.K. Card. 1999. Information Foraging. *Psychological Review*, 106(4):643–75.
- . 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. Verrigtinge van die Internasionale Konferensie oor Intelligensie-ontleding, McLean, Virginia.
- Ripple, A.S. 2006. Expert Googling: Best Practices and advanced strategies for using google in health sciences libraries. *Medical Reference Services Quarterly*, 25(2):97–107.
- Roberts, N.C. 2011. Tracking and disrupting dark networks: Challenges of data collection and analysis. *Information Systems Frontiers*, 13:5–19.
- Ryan, M. 2004. Multivariant narratives. In Schreibman e.a. (reds.) 2004.
- Schreibman, S., R. Siemens en J. Unsworth (reds.). 2004. *A companion to digital humanities*. Oxford: Blackwell.

- Schreibman, S., R. Siemens en J. Unsworth. 2004. The digital humanities and humanities computing: An introduction. In Schreibman e.a. (reds.) 2004.
- Seeman, M. 1959. On the meaning of alienation. *American Sociological Review*, 24(6):783–91.
- Senekal, B.A. 2011. Die digitalisering van NALN se knipselversameling: Die bemiddeling van 21ste-eeuse navorsing in die Afrikaanse letterkunde. *LitNet Akademies*, 8(2):46–65. <http://www.litnet.co.za/Article/die-digitalisering-van-naln-se-knipselversameling-die-bemiddeling-van-21ste-eeuse>
- Senekal, J.H. 1987. *Literatuuroppvatting: "Wese" en "waarhede" van 'n nuwe literêre teorie*. Bloemfontein: Universiteit van die Oranje-Vrystaat.
- Splivalo, D. 2009. *Chiliad selected for "KMWorld 100 companies that matter in knowledge management" award*. http://www.chiliad.com/2009_02-26.php (27 Januarie 2011 geraadpleeg).
- Taboada, M., M.A. Gillies, P. McFetridge en R. Outtrim. 2008. Tracking literary reputation with text analysis tools. Vergadering van die Vereniging vir Digitale Geesteswetenskappe, Vancouver. Junie.
- Tardy, C.M. 2010. Writing for the world: Wikipedia as an introduction to academic writing. *English Teaching Forum*, 48(1):12–27.
- Thomas, J.J. en K.A. Cook. 2006. A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1):10–3.
- Treverton, G.F. en C.B. Gabbard. 2008. *Assessing the tradecraft of intelligence analysis*. Santa Monica: RAND Corporation.
- Treverton, G.F., S.G. Jones, S. Boraz en L. Lipsky. 2006. *Toward a theory of intelligence*. Santa Monica: RAND Corporation.
- Van Coller, H.P. 2004. Is Leroux 'n vergete skrywer? *Stilet*, 16(1):1–31.
- Van Coller, H. en B. Odendaal. 2008. Die meer "beskeie" opsies van 'n "buitestander": M. Nienaber-Luitingh in die Afrikaanse literêre sisteem, *LitNet Akademies*, 5(3), Desember, 33–50. http://www.oulitnet.co.za/newlitnet/pdf/1a/LA_5_3_vancoller_odendaal.pdf
- Venter, C. 2002. 'n Sisteemteoretiese perspektief op die vertaling van Suid-Afrikaanse literatuur in Nederlands. Ongepubliseerde PhD-proefskrif, Universiteit van die Vrystaat.
- Venter, R. 2006. Die materiële produksie van Afrikaanse fiksie (1990-2005): 'n Empiriese ondersoek na die produksieprofiel en uitgeweryprofiel binne die uitgeesisteem. Ongepubliseerde PhD-proefskrif, Universiteit van Pretoria.

Vermeulen, W., O.L. Lategan en R. Litheko, R. 2011. *The research process*. Bloemfontein: Sun Media.

Warner, M. 2002. Wanted: A definition of intelligence. *Studies in Intelligence*, 46(3):15–22.

Warwick, C., M. Terras, I. Galina, P. Huntington en N. Pappa. 2008. Library and information resources and users of digital resources in the humanities. *Program: Electronic Library and Information Systems*, 42(1):5–27.

Wilden, A. 1980. *System and structure: Essays in communication and exchange*. New York: Tavistock.

Young, R. (red.) 1981. *Untying the text: A post-structuralist reader*. Boston: Routledge & Kegan.

Žižek, S. 2004. *What Rumsfeld doesn't know that he knows about Abu Ghraib*. 21 Mei. <http://www.lacan.com/zizekrumsfeld.htm> (1 November 2011 geraadpleeg).

Eindnotas

¹ *User interface*.

² Vir meer inligting oor die literêre sisteem, sien Senekal (1987), Even-Zohar (1990), Venter (2002), Venter (2006) en Van Coller en Odendaal (2008).

³ Volgens Warner (2002:21–2) lê die onderskeid tussen intelligensie en inligting eerder in die aspek van geheimhouding wat 'n integrale deel van eersgenoemde is. Treverton e.a. (2006:7) noem egter dat geheimhouding in die hedendaagse opset nie meer so sterk beklemtoon word nie. 'n Beduidende hoeveelheid data binne militêre intelligensie kom ook van sogenaamde oop bronne – inligting wat in die openbare domein beskikbaar is.

⁴ 'n Verdere belangrike ooreenkoms tussen akademiese navorsing en intelligensie is dat 'n uiteindelijke verbruiker (kliënt) die *intelligensieprodukt* sal gebruik, maar dat *intelligensie* (soos akademiese navorsing) streef na objektiwiteit: “The realm of intelligence is that of ‘fact’, considered judgment, and probability, but not prescription” (Krizan 1999:17). Intelligensie, soos akademiese navorsing, is dus nie veronderstel om in diens te staan van politieke oorwegings nie, maar is veronderstel om na die objektiewe oorweging van inligting te streef. Uiteraard word die verwerkingsfase direk beïnvloed deur die navorser se persoonlike voorkeure, verwysingsraamwerk en wetenskaplike paradigma – kortweg: sy menslikheid (Mouton en Marais 1990:10–2). Krizan (1999:36–8) gee 'n uiteensetting van verskillende kategorieë van vooroordele en wanopvattinge wat die eindresultaat beïnvloed, maar hier word daarmee volstaan om te noem dat intelligensie, soos akademiese navorsing, streef na objektiwiteit, maar dat die menslike komponent altyd steeds relevant bly.

⁵ Hierdie diagram is aangepas vanuit Pirolli en Card (2005:2) en beskrywings verwys dus na hulle gebruik van terminologie.

⁶ Pirolli en Card (2005:5) verwys na die tendens om meer ruis (minder-relevante inligting) te verdra ter wille van sagter “seine” (relevante inligting wat andersins gemis kon wees).

⁷ Sien Van Coller (2004:1–6) vir meer inligting oor kanons en kanonisering.

⁸ Tot en met einde 2010.

⁹ Indien elke artikel van hierdie tydskrifte reeds beskikbaar was, sou meer wetenskaplike gevolgtrekkings bereik kon word. *Tydskrif vir Nederlands en Afrikaans* is weggelaat omdat dit nie digitaal beskikbaar nie.

¹⁰ Aangesien Voyeur slegs aanlyn beskikbaar is, is hierdie soektog met Adobe Acrobat X Pro gedoen, maar die resultate behoort identies te wees, aangesien digitaal-gegenereerde PDF dokumente nie karakterherkenning behoort te beïnvloed nie.

¹¹ My dank aan die Erfenisstigting se Navorsingstrust wat hierdie studie moontlik gemaak het.