





Sagen op de internationale kaart in ISEBEL

THEO MEDER

Volkverhaalonderzoekers willen graag meer weten over de geschiedenis, verspreiding en betekenis van bepaalde sagen en hun motieven, bijvoorbeeld over trollen of kabouters, maar hoe krijg je dat internationale sagenmateriaal, in al die verschillende talen, bij elkaar voor comparatief onderzoek? ISEBEL kan uitkomst bieden: het acroniem staat voor *Intelligent Search Engine for Belief Legends*.

Op zoek naar een harvester

Het begon allemaal met een idee op een conferentie van de International Society for Folk Narrative Research in juni 2009 in Athene, Griekenland. Theo Meder presenteerde een paper met de titel *From a Dutch Folktale Database towards an International Folktale Database*. Het idee was om een reeds functionerende volksverhalendatabase te vullen met extra volksverhalen uit meerdere talen. Om verschillende volksverhalenbanken, die al in verschillende landen draaiden, samen te voegen tot één grote, centrale superserver. Tijdens de discussie vroeg Tim Tangherlini zich af of al deze regionale en nationale instellingen wel enthousiast al hun verhalen en metadata zouden willen overdragen aan een centrale database waarover ze weinig controle hadden. Tangherlini suggereerde dat het technisch handiger zou zijn om een harvester of 'oogstmachine' te ontwikkelen, die af en toe verhalen en metadata uit bestaande volksverhalendatabases zou ophalen via een online verbinding, zodat deze verhalen vervolgens centraal konden worden doorzocht. Dit idee rijpte een aantal jaren. Vervolgens deed zich in 2017 de mogelijkheid voor om een samenwerkingsproject gefinancierd te krijgen via het Transatlantic Digging into Dataprogramma van NWO. De drie samenwerkende partijen bestonden uit:

- Meertens Instituut, Amsterdam: Theo Meder en Vic Ding
- UCLA (nu UC Berkeley): Tim Tangherlini en Peter Broadwell
- Universiteit van Rostock / Wossidlo Instituut: Christoph Schmitt, Petra-Himstedt-Vaid, Holger Meyer en Alf-Christian Schering

Het doel van het project zou zijn om een internationale oogstmachine te ontwikkelen voor sagen onder de naam ISEBEL. De aanvraag werd toegekend voor een periode van drie jaar: 2017-2020. Gedurende deze periode zou ISEBEL volksverhaalmateriaal uit drie databases samenbrengen die beheerd worden door de deelnemers aan het project, namelijk:

- De Nederlandse Volksverhalenbank (Meertens Instituut, Amsterdam)
- De Deense volksverhalendatabase (UC Berkeley, VS)
- Wossidlo (Rostock, Duitsland)

Digitale databanken

Hoewel elke van de drie databanken volksverhalen bevatte, waren er duidelijke verschillen in reikwijdte en achterliggende structuur. De Nederlandse Volksverhalenbank bestrijkt een periode van de middeleeuwen (ca. 1200) tot heden en bevat verschillende collecties, met een overheersende focus op collecties uit de 19e en vooral de 20e eeuw. De talen variëren van Middelnederlands en 17e-eeuws Nederlands tot modern Nederlands, Fries en diverse streektaalen. De genres die in het systeem zijn gecatalogiseerd, zijn ook divers: sprookjes, traditionele en hedendaagse sagen, legenden en moppen. In totaal bevat de Nederlandse Volksverhalenbank meer dan 48.000 volksverhalen.

De Deense Volksverhalendatabase bestaat uit één grote collectie: het folkloristische levenswerk van Evald Tang Kristensen (1843-1929). Rond de eeuwwisseling legde hij een omvangrijke verzameling volksliederen, sprookjes, moppen, sagen, raadsels en spreekwoorden aan, grotendeels verzameld op het Deense schiereiland Jutland. De Deense database brengt vooral zijn sagen samen: alles bij elkaar meer dan 31.000 verhalen.

De Noordoost-Duitse volksverhalendatabase bevat ook het levenswerk van één verzamelaar: Richard Wossidlo (1859-1939). Hij verzamelde talloze volksverhalen en tradities in Mecklenburg. Al zijn materiaal werd uiteindelijk bewaard in het Wossidlo-archief in Rostock, dat naar schatting 2 miljoen papieren documenten bevat. Het materiaal is geordend in een uitgebreid en kruisverwijzend kaartstelsel. Voor ISEBEL is momenteel een selectie van ruim 14.000 Duitse verhalen gedigitaliseerd.

ISEBEL

Het leek verstandig om traditionele sagen en de bijbehorende metadata te selecteren voor ISEBEL, aangezien elk van de databanken dergelijke records in overvloed had en een dergelijke selectie vergelijkend onderzoek mogelijk maakt. Een bijkomend voordeel van het selecteren van deze subsets is dat een groot aantal van deze geloofssagen nooit is gepubliceerd. ISEBEL zou de toegankelijkheid van deze collecties vergroten en doorzoekbaar maken. De sagen kunnen worden gekarakteriseerd als *belief legends*, die over een breed



Jager en drie kabouters, Mira, ca. 1850-ca. 1910
Rijksmuseum, Amsterdam

scala aan bovennatuurlijke fenomenen gaan, zoals spoken, geesten, kabouters, trollen, duivels, heksen, tovenaars, magie, weerwolven, meerminnen, nachtmerries en reuzen.

Het ISEBEL-systeem kreeg toegang tot meer dan 31.000 Deense sagen, 26.000 Nederlandse sagen en 14.000 Mecklenburgse sagen, in totaal meer dan 71.000 sagen. Hoewel dat een aanzienlijke verzameling vertegenwoordigt, is een belangrijk voorbehoud dat niet alle verhalen gelijktijdig zijn verzameld. Het Duitse en Deense materiaal dateert van enkele decennia voor en na 1900. In die periode werd er ook in Nederland verzameld, maar niet zo intensief, en de oogst was toen veel kleiner: hooguit 700 sagen. De grote verzamelactie in Nederland vond pas in de jaren zestig en zeventig plaats. Zo verzamelde de Friese verzamelaar Dam Jaarsma (1914-1991) in die periode ruim 16.000 volksverhalen en 23 andere verzamelaars ongeveer evenveel. De implicatie is dat toen iemand als Dam Jaarsma sagen verzamelde, de vertellers er vaak al niet meer in geloofden: het waren de verhalen die ouders of grootouders voor waar hielden. Veel van Tang Kristensens en Wossidlo's informanten geloofden rond 1900 nog wel in de bovennatuurlijke verschijnselen en gebeurtenissen uit hun verhalen.

Het was technisch nog niet eenvoudig om al het materiaal uit de drie databases adequaat te koppelen. De Deense en Nederlandse databases zijn tamelijk eenvoudige relationele databanken. De Duitse database moest echter de structuur van het systeem in het Wossidlo-archief representeren met alle kruisverwijzingen, en daarvoor werd een *hypergraafdatabase* ontwikkeld met talloze onderling verbonden knopen.

ISEBEL oogstte vervolgens alle complete sagen. Aan elk verhaal werd de volgende metadata gekoppeld: Wie

is de verteller en wat is zijn geslacht? Wanneer en waar werd het verhaal verteld? Met welke trefwoorden is het verhaal gekarakteriseerd? Welke locaties komen voor in het verhaal? Wie is de verzamelaar? Welke motieven en verhaaltypen uit de internationale catalogi kunnen (indien mogelijk) aan het verhaal worden toegekend? Logo en website (isebel.eu) van ISEBEL waren snel ontwikkeld. De zoekmachine met geografische verspreidingskaart en metadata vergde veel meer ontwikkeltijd: search.isebel.eu.

Zoeken in het Engels

De reden waarom ISEBEL 'intelligent' wordt genoemd, is de doorzoekbaarheid van de gegevens. Onderzoekers kunnen zoeken in een van de doeltalen. *Varulv* levert Deense sagen over weerwolven op. Maar alles kan ook in het Engels gezocht worden. *Werewolf* levert alle Deense, Nederlandse, Friese en Duitse treffers op. Aanvullende analyses, zoals de visualisatie van de geografische spreiding, vergroten de output van zo'n brede zoekopdracht.

Alles handmatig in het Engels vertalen was gezien de hoeveelheid sagen ondenkbaar. Bovendien, als er later meer databanken bij ISEBEL zouden aansluiten, zou automatische vertaling de enige werkbare oplossing zijn. Het probleem zat dus vooral in het automatisch en met een acceptabele nauwkeurigheid naar het Engels vertalen van alle teksten: met Standaarddeens, -duits en -nederlands ging dat redelijk, maar met bijvoorbeeld Nederduits en Fries werd het lastiger omdat nauwkeurige taalmodellen hier ontbraken.

Google Vertalen leverde goede resultaten op voor Deens, Duits en Nederlands in de standaardspelling, maar voor het tegen betaling vertalen van het hele corpus was er geen budget. ISEBEL moest dus opgeleid

worden om zelfstandig te vertalen. Peter Broadwell en Tim Tangherlini namen de verantwoordelijkheid voor het automatische vertaalproces, waarbij zij erin slaagden zeer acceptabele vertalingen te realiseren door middel van machinelearningtechnieken. De vertalingen leverden zogenaamde *dirty translations* op, die onzichtbaar zouden blijven voor de eindgebruiker; lastig leesbaar, maar afdoende om mee te zoeken.

Er was echter nog een hobbel te nemen: in volksverhalen worden soms specifieke begrippen gebruikt waar het vertaalprogramma op stukloopt. Termen als *duivelsdrek* 'asafoetida' (als afweer tegen hekserij), *stalkaars* 'dwaallicht', *tsjoenster* 'heks', *karismatiker* 'gebedsgenezer', *gammen* 'griffioen', *Wesselbalg* 'wisselkind', *Freikugel* 'magische kogel' of *Hexenproben* 'heksentest' worden vaak niet goed vertaald. Om problemen met zulke woorden uit het domein van het volksgeloof aan te pakken, is een *domeinspecifieke woordenlijst* gemaakt voor woorden waarvoor automatische machinevertaling dreigt te mislukken. Deze tweetrapsraket van *vuile vertaling* aangevuld met een *domeinspecifieke woordenlijst* leverde betere zoekresultaten op, maar er waren nog steeds problemen. Ten eerste gooiden spellingsvarianten roet in het eten: zo werd *duivelsdrek* niet herkend als *duivelsdrek*. Ten tweede verwezen verhalen vaak naar fenomenen die niet expliciet werden genoemd, zoals een vrouw met heksachtig gedrag die in het verhaal nooit heks wordt genoemd, of iemand die zichzelf van het leven berooft, terwijl het woord *dood* niet voorkomt.

Uiteindelijk zijn we bij een drietrapsraket uitgekomen: we besloten de trefwoorden die bij de verhalen als metadata waren toegevoegd, ook mee te vertalen. Aangezien het trefwoorden in de standaardtaal waren, stuitte dat nauwelijks op vertaalfouten.

Als eenvoudige test werd er gezocht op de term *witch*: zonder vertaling van de trefwoorden leverde dat ongeveer 700 verhalen op, terwijl dat met de vertaalde trefwoorden zo'n 3000 verhalen opleverde. Kortom, de gezochte verhalen worden nu gevonden, en een groot deel is niet langer *lost in translation*.

Comparatief onderzoek en Europese ambities

De technische ontwikkeling van de database, de vertaaltechnologie en de geografische visualisatie hebben uiteindelijk relatief veel tijd en budget in beslag genomen, wat ten koste gegaan is van vergelijkend onderzoek, waar ISEBEL uiteindelijk voor gebouwd is.

Er is een bescheiden begin gemaakt met onderzoek naar sagen over meerminnen. De geografische verspreidingskaart laat zien, zeker bij verder inzoomen, dat verhalen over meerminnen – niet geheel verrassend – vooral gevonden worden in de kuststreken. Naarmate men verder landinwaarts kijkt, neemt de frequentie van meerminnsagen snel af.

Vergelijkend onderzoek naar de distributie van weerwolf-sagen stelt ons voor problemen bij een verklaring. In Noordoost-Duitsland komen de verhalen over weerwolven vooral voor in bosrijke gebieden. In Nederland daarentegen zien we een duidelijk onderscheid tussen het katholieke zuiden en het protestantse noorden. In het zuiden worden

veel meer sagen over weerwolven gevonden. Mogelijk is er een connectie met de langduriger vervolging van weerwolven (en heksen) in de katholieke gebieden.

Momenteel wordt ook onderzoek gedaan naar sagen over personen die op magische wijze de wind kunnen draaien. Zij worden te hulp geroepen als er brand is: wanneer er een boerderij in vlammen dreigt op te gaan, zijn bepaalde *tovenaars* in staat om de wind te draaien, zodat het vuur een andere kant op gaat. Opvallend genoeg blijken het in ISEBEL regelmatig katholieke priesters te zijn, die hiertoe in staat zijn, maar protestantse dominees nooit.

Graag willen we met behulp van ISEBEL nog veel meer onderzoeksvragen stellen. Het is daarbij van belang dat we het sagenmateriaal met meer talen uitbreiden. Binnenkort gaat er een ISEBEL-plusproject van start, waarin onderhoud, verfijning en uitbreiding van de database gerealiseerd kunnen worden, dankzij financiering vanuit het Meertens Instituut. Inmiddels is al begonnen met de invoer van 6000 IJslandse sagen. Opvallend genoeg vinden we in de IJslandse sagen geen enkel weerwolfverhaal. Er kwamen in IJsland ooit wel wolven voor, maar die zijn in een zeer vroeg stadium al uitgeroeid door boeren die hun vee beschermden. Kennelijk heeft dat repercussies gehad voor de vertelcultuur en een gebrek aan angst voor weerwolven.

Voor opname in ISEBEL staan weldra Noorse en Zweedse sagen op het programma. Er zijn toezeggingen om ook Ierse en Schotse sagen te gaan aanleveren. Graag zouden we ook het verhaalmateriaal uit de Vlaamse Volksverhalenbank aan ISEBEL toevoegen. De digitale sagenverzamelingen uit Portugal en Catalonië zijn eveneens geschikte kandidaten. Onderzoek naar repertoires van mannelijke en vrouwelijke vertellers staat op het verlanglijstje. We zouden, indien mogelijk, graag meer specifieke zoekfuncties en filters toevoegen, zodat er kant-en-klare modules klaarstaan om onderzoek te doen naar bijvoorbeeld spookhuizen, heksenkransen of duivelspacten. Op zo'n manier kunnen we ISEBEL langzaam maar zeker uitbreiden tot een Europees onderzoeksinstrument, met een rijke hoeveelheid materiaal aan volksverhalen, volksgeloof en immaterieel erfgoed.

Literatuur

- Meder, T. (2010). *From a Dutch Folktale Database towards an International Folktale Database*. *Fabula*, 51(1-2), 6-22.
- Meder, T. (2018). ISEBEL: Intelligent Search Engine for Belief Legends. *Volkskunde*, 119(1), 87-89.

Theo Meder is als senior volksverhaalonderzoeker verbonden aan het Meertens Instituut in Amsterdam, en als bijzonder hoogleraar Volksverhalen en Vertelcultuur verbonden aan de Rijksuniversiteit Groningen. Samen met RUG-studenten David de Vries en Eline van Ekert publiceerde hij onlangs het boek *Trijntje en Duivelsdrek: 17 oude Groninger sprookjes van Trijntje Soldaats en haar toehoorders*.

Contact: theo.meder@meertens.knaw.nl

Neerlandia 01

Nederlands-Vlaams tijdschrift voor taal, cultuur en maatschappij

JAARGANG 127 2023

Cultuur & maatschappij

[Nederland, Vlaanderen, begrijpen we elkaar?] Begin niet te snel te jijen en jounen. Begin niet meteen over geld. Het contact is in België net zo belangrijk als het latere contract.

Taal

[Aan de haal met taal] Levende talen kunnen nooit helemaal in een standaardvorm gecodeerd worden. Een taalgebruik voor alle toekomstige gebruikers vastleggen is een illusie.

Kubrieken

[Stage] Al snel kwam ik era chter dat de internationale neerlandici misschien nog wel enthousia ster waren over de studie dan de studenten die ik in Groningen kende.



Ze wordt bijzonder gewaardeerd omdat ze een jong meisje was. En tegelijkertijd wordt ze juist daarom als literator ondergewaardeerd