Alle zeventiende-eeuwse kranten in Delpher getranscribeerd

Oud nieuws voor nieuw onderzoek

De transcriptie van een grote hoeveelheid kranten maakt het mogelijk om taal- en cultuurhistorische veranderingen in de Gouden Eeuw grootschalig te onderzoeken.

Nicoline van der Sijs

Naar de ontwikkelingen van de Nederlandse Gouden Eeuw wordt veel onderzoek gedaan. Drukwerk uit de zeventiende eeuw vormt daarbij een belangrijke bron. En alhoewel er veel drukwerk is gepubliceerd en bewaard, bestond er tot nu toe geen aaneengesloten digitaal tekstcorpus waarmee taal- en cultuurhistorische veranderingen systematisch konden worden onderzocht.

20 miljoen woorden

Maar er is goed nieuws. Vrijwilligers hebben alle 17e-eeuwse kranten op Delpher - lopend van 1618 tot 1700 - getranscribeerd. Hiermee is het beschikbare digitale onderzoekcorpus van de zeventiende eeuw in één klap bijna verdubbeld. Het krantencorpus bestaat uit 6.184 verschillende kranten die samen een kleine 20 miljoen woorden bevatten. Ter vergelijking: de DBNL-teksten voor deze eeuw bestaan uit circa 24 miljoen woorden. Door het transcriberen kunnen krantenteksten voor het eerst systematisch met de computer worden onderzocht. Tot nu toe was dat niet mogelijk omdat de optische tekenherkenning waarmee de teksten op Delpher waren gelezen, niet overweg kon met het gotische schrift en Oudnederlands. Medio 2020 komt het getranscribeerde krantencorpus beschikbaar via Delpher.

Verrijking van data

Het Meertens Instituut werkt aan het verder verrijken van de digitale tekstbestanden. Zo worden de metadata opgeschoond en uitgebreid en worden afzonderlijke artikelen semi-automatisch afgesplitst en voorzien van informatie over de tekstsoort (zoals advertentie, binnenlands nieuws, officiële mededeling). Ook de geografische namen die in de krantenkoppen voorkomen, worden verrijkt en benut: aan iedere naam wordt de moderne spelling toegevoegd. Die moderne schrijfwijzen kunnen vervolgens worden ingevoerd in een kaartprogramma, dat week voor week visualiseert waar het nieuws binnen en buiten Europa vandaan kwam, en hoe de geografische focus in de loop van de eeuw veranderde.

Lacunes in kennis

De opgeschoonde en verrijkte krantenteksten komen in 2020 ook beschikbaar via een aparte interface. Dan kan iedereen zijn eigen onderzoeksvragen stellen, bijvoorbeeld naar maatschappelijke veranderingen of veranderingen in het taalgebruik. De teksten kunnen allerlei lacunes in kennis en gegevensbronnen aanvullen: zo ontdekte het Meertens Instituut al dat kranten een groot aantal woorden en spellingen bevatten die ontbreken in de bestaande historische lexica van het Nederlands. Het krantencorpus kan een proeftuin worden voor het testen van tools en modellen, zoals semantische vectoren en *topic modelling*. En de liefhebber kan natuurlijk ook gewoon het laatste nieuws van een bepaalde datum lezen.

meertens.knaw.nl

Nicoline van der Sijs is projectleider bij het Meertens Instituut. Heeft u vragen of suggesties, of wilt u meewerken aan dit project? Neem dan contact op: post@nicolinevdsijs.nl.

1684. N° 10

Amsterdamsche Donderdaghse Courant

PORTUGAL.

Lisbon den 7 February. Gedurende de indispositie van den Koninck ontstaen over de groote droefheyt wegens 't overlyden van de Koninginne, soo hebben aen dit Hof alle hooge tiecken stil gestaen, doch nu is syne Majesteyt wederom volkomen gereconvalesceert, soo dat nu weer aen alle de vreemde Ministers en andere grooten audientie geeft, ende dier van wyder ook noch de complimenten van condoleantien heeft ontfangen, ende soo sullen dienvolgens alle affaires die tot noch toe stil gestaen hebben, nu ook weer herstat en levendig gemaeckt werden; en daer onder mede die wegens het huwelyck van de Infante van Portugael: men inclineert hier noch even seer tot de continuatie van de vrede; alsoo daer uyt meer voordeels als uyt den oorlog wort verwacht.

SPANJEN.

Madrit den 17 February. Aen het Hof ende in den Raet van de Monarchie wort met allen yver voortgevaren in de besorginge van al het geene tot eenen vigoureusen oorlog tegen Vranckryck wordt vereyst, ende om de Nederlanden te stellen in eene poincture van wapenen als hesby de Geallieerdens van syne Majesteyt wordt vereyst ende ook conform is met de alliantie, soo syn boven de ordinaire maendelycke subsidien, die volgens gewoonte syn overgesonden, noch andere extraordinaire remisen derwaerts overgemaekt, de waerde van twelf hondert duysent ryckdaelders, verdeelt door drie bysondere wegen: ende men meent dat de twee duysent soldaten die derwaerts gedestineert syn, om voor af te trecken, nu al sullen geembarqueert wesen om naer Oostende overgescheept te worden. Niet minder vlyt wort er gedaen om alle vordere Regimenten en Compagnien, die daer toe in de respective Rycken en Provincien vaerdig syn, op het spoedigste naer Catalonien te doen marcheren. Den Keyserlyken Ambassadeur den Grave van Mansfelt heeft dagelycks eenige conferentie met de voornaemste Grandes van de Monarchie over de besorginge van de assistentien voor den Keyser tegen den Turck, maer alsoo den noot nu soo groot niet en is als doen Weenen belegert was, is voor een derde verminderinge daer in gedaen, soo wel in de geestelycke als wereltlycke contributien die als nog daer toe geschieden, om deselve in den particulieren oorlog volgens de ordre des Koninghs te gebruycken: ende daer enboven is ook een vierde part van alle de vreemde effecten, boven de Fransche effecten die geheel geconfisqueert syn, uyt de vloot van nieuw Spanien genomen, om hetselve aen de geinteresseerden tegens 8 pet cento tot Cartagena te doen betalen. Daer marcheren noch al gedurig vele troupen van alle kanten naer Catalonien; alwaer met de verstercking van Fuicerda seer sterck wordt voort gegaen. Van Lisbon is naericht dat den Koning van Portugael wederom in voorige gelomheyt was hersteit, en dat er van daer in korte eene extraordinaire Ambassade herwaerts stonde afgesonden te worden, om de alliantie met dese Monarchie te renoveren.

ITALIEN.

Romen den 12 February. Syne Heyligheyt ende het geheele Hof hebben eene groote estime voor den Senator Landi van Venetien, en heeft den Paus seer voordeelig van hem met den Cardinael Cibo gesproocken. Den Cardinael Pio heeft selfs geseyt dat hy in twee jaeren by den Paus soo veel niet hadde kennen uytrichten, als gemelten Heer Landi in twee audientien. De meeste Cardinalen gaen hem besoecken; en den Connestabile Colonna is er ook in persoon geweest, en heeft hem getracteert als of hy den carsster van Ambassadeur hadde. Den tweeden deses heeft men het *Te Deum* gesongen wegens de victorie die de Cosacken op de Turcken ende Tartaren hebben bevochten, gelyck men daer van den 30 passato door eenen extraordinairissen Courier de tydinge heeft gekregen: Door de doodt van den Cardinael Rospigliosi syn eenige considerable sommen in den kerckelycken schat gebracht, gelyck ook het overlyden van den Cardinael Boncompagno, dewelcke alle tot den oorlog tegen de Turcken sullen worden gemployeert. Syne Heyligheyt tot noch toe continuerende groote remisen aen den Keyser ende den Koning van Poolen te senden, soecke nu alle middelen om desgelycks de Republyck van Venetien met gelt by te staen. Alle uren verwacht men hier de volmachten van den Keyser

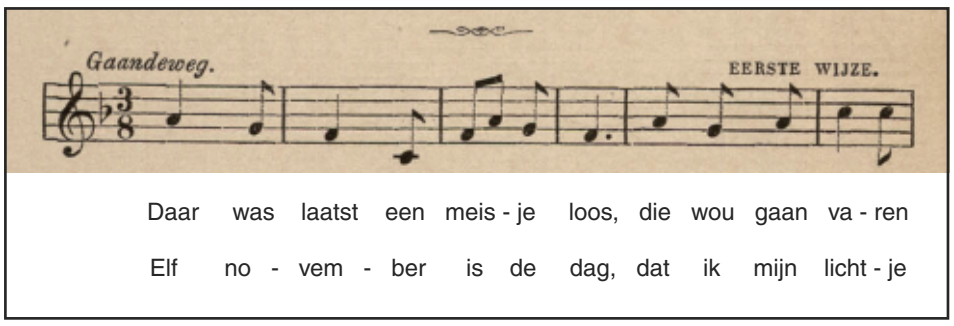
In plaats van woord voor woord lezen, kunnen onderzoekers de computer de getranscribeerde krantenteksten laten doorzoeken. *Credits: Detail uit Amsterdamse Courant, 1684, via delpher.nl*

Artificiële Intelligentie achter de Liederenbank

Melodiegelijkenissen opsporen met algoritme

In de Nederlandse Liederenbank, waarin ruim 175.000 Nederlandse liederen zijn ontsloten, kun je zoeken naar melodiegelijkenissen. Peter van Kranenburg ontwikkelde het algoritme achter deze functie.

Mathilde Jansen



In de Nederlandse Liederenbank is het mogelijk om naar melodiegelijkenissen te zoeken. Een van de zoekresultaten bij 'Elf november is de dag' is 'Daar was laatst een meisje loos'. *Credits: Meertens Instituut*

Wie in de Nederlandse Liederenbank 'Elf november is de dag' intypt, komt via 'vergelijkbare melodieën' terecht bij 'Daar was laatst een meisje loos'. Die mogelijkheid om naar melodiegelijkenissen te zoeken, is de verdienste van Peter van Kranenburg. Hij is computationeel musicoloog aan het Meertens Instituut en onderzoekt muziek aan de hand van computermodellen.

In 2010 promoveerde hij op een uitlijningsalgoritme. "Dat schrijft de melodieën zo onder elkaar, dat de overeenkomende noten precies onder elkaar komen te staan", legt Van Kranenburg uit. "Het algoritme zoekt uit op welke plekken ruimte toegevoegd moet worden, zodat de corresponderende delen onder elkaar staan. Hoe meer ruimte, hoe slechter de gelijkenis. Grof gezegd."

Tune families

Als je een uitlijning maakt van een querymelodie met alle melodieën uit de Liederenbank, en die sorteert, dan komen de meest gelijkende melodieën bovenaan. "Net als Google-resultaten", verduidelijkt Van Kranenburg. Zo zie je welke melodieën varianten zijn van elkaar, en kun je ze onderverdelen in families, ook wel 'tune families' genoemd.

Tegenwoordig is veel kunstmatige intelligentie gebaseerd op neurale netwerken. Daarom onderzocht Van Kranenburg samen met collega's van het Meertens Instituut en de Universiteit Antwerpen of dit ook werkte bij melodiegelijkenissen. "Om het neurale netwerk te trainen, werden steeds twee melodieën aangeboden die wel op elkaar lijken en twee die niet op elkaar lijken. Als je dat lang genoeg doet, met heel veel verschillende melodieën, in ons geval zo'n zesduizend, dan hoop je dat zo'n netwerk op een gegeven moment leert wat het betekent dat twee melodieën op elkaar lijken." En dat lukte. Het model vond melodiegelijkenissen met een betrouwbaarheid van 70 tot 80 procent. Iets beter dan het uitlijningsalgoritme. "Nog geen grote verbetering, maar het laat wel zien dat het model werkt. En dat biedt perspectief voor de toekomst en vormt nieuwe uitdagingen. Want het neurale netwerkmodel is misschien wel intelligent, voor mensen is het soms moeilijk te interpreteren wat het allemaal doet. Het begrijpelijk maken van die netwerken is een belangrijk onderzoeksgebied. Daar willen we in een volgende stap aan bijdragen door te onderzoeken wat ons netwerk geleerd heeft over melodische gelijkenis."

liederenbank.nl

JONG TALENT

'Een uitlijningsalgoritme schrijft de melodiën en overeenkomende noten precies onder elkaar'

Peter van Kranenburg

Van Kranenburg studeerde Musicologie aan de Universiteit Utrecht en Electrical Engineering aan de TU Delft. Hij promoveerde in 2010 aan de Universiteit Utrecht. Hij werkt als computationeel musicoloog bij het Meertens Instituut en de Universiteit Utrecht.