

# De robot is nog lang geen journalist

## Robotjournalistiek

Automatisering in de pers heeft nog veel weg van de mechanische turk: de schakende 'machine' uit de 18e eeuw waar een mens in bleek te zitten. Automatische schrijvers van nieuwsberichten zijn nog verre van (kunstmatig) intelligent.

Door onze medewerker  
**Amanda Verdonk**

'Automate everything', is een van de slogans op de wand bij de start-up Automated Insights in universiteitsstad Durham in de Amerikaanse staat North Carolina. De 55 medewerkers hoeven maar uit het raam te kijken om de daad bij het woord te voegen. Ze werken vanuit een kantoor in een honkbalstadion aan het automatiseren van sportwedstrijden.

„Op vrijdagmiddag kijken we graag met een biertje in de hand naar een wedstrijd”, vertelt directeur productmanagement Adam Long, wijzend uit het raam naar het speelveld. Hier spelen de Durham Bulls, een honkbalploeg uit de Minor League, de net-niet-topdivisie. Op speeldagen kunnen er tienduizend toeschouwers in het stadion; nu is het leeg want het seizoen is afgelopen.

Voor landelijke media is het ondoenlijk om naar alle tienduizend Minorleaguewedstrijden een verslaggever te sturen, maar Amerikanen zijn dol op sport en statistieken over hun favoriete spelers en ploegen. Dus als de Durham Bulls spelen, zitten er naast lokale verslaggevers ook mensen in dienst van statistiekbedrijven op de tribune. Die volgen elke actie van de spelers minutieus: wie er een run maakt, in welke inning en in welke minuut. Datatypisten voeren alles in, en met deze database gaat Automated Insights aan de slag. Met hun programma Wordsmith kun je een tekstsjabloon maken, synoniemen zoeken ('scoren', 'een punt maken', 'treffen') en regels instellen zoals: als er zeldzame gebeurtenissen zijn, zoals 'no hitters' of 'grand slams', meldt dat dan als eerste.

### Dertien competities

Sinds afgelopen zomer levert Automated Insights deze kant-en-klare teksten aan persbureau Associated

Press (AP), dat nu automatisch verslagen publiceert van 142 honkbalteams en dertien competities in de Minor League. Over die subtop werd eerder slechts mondjesmaat geschreven; de computers vervangen geen verslaggevers, verzekeren ze bij AP.

Uit verschillende wetenschappelijke onderzoeken blijkt dat mensen het verschil tussen computergegenerateerde en menselijk geschreven teksten niet kunnen zien. Ze vinden de computerteksten zelfs betrouwbaarder ogen, maar ook saai omdat ze veel cijfers bevatten en weinig creatief geschreven zijn. Automated Insights maakt hun berichten voor AP bewust saai en zakelijk, maar dat hoeft niet zo te zijn.

Wordsmith schreef bijvoorbeeld over de staatsverkiezingen in North Carolina: „Het was een tumultueuze nacht voor de senaat in North Carolina, omdat vijf districten van kleur veranderden en de Democraten hun meerderheid kwijt raakten aan de Republikeinen.” Hoe verzint een computer dat, van die tumultueuze nacht? „Dat is heel simpel”, legt Locy Stewart van het bedrijf uit. „Ik heb dat woord zelf bedacht. Het is gewoon een regel die je instelt: als meer dan vijf zetels veranderen, zal het wel tumultueus zijn, of een ander synoniem. En verkiezingen worden altijd 's nachts beslist.”

Stewart, die net als zijn collega's geen journalistieke achtergrond heeft, schreef ook een sjabloon voor een necrologie, en gebruikte dat voor computerpionier Marvin Minsky, gebaseerd op zijn Wikipediapagina. Het kostte hem drie kwartier om te maken. Ook de teksten die Automated Insights schrijft over American football voor Yahoo hebben een eigen toon: enigszins plagerig en doortrapt.

### Mensenwerk

Zo komt er anno 2016 nog heel wat mensenwerk kijken bij het proces van sportwedstrijd tot nieuwsartikel: mensen tikken de data in, mensen maken de sjablonen en mensen heb-



## Robotjournalist wint nog lang niet de Pulitzerprijs

**Adam Long,**  
Automated Insights

ben veel werk om te zorgen dat dat proces soepel verloopt. Bij AP hebben ze zelfs een 'news automation editor' aangenomen sinds ze hun verhalen automatiseren.

Sportwedstrijden zijn nog redelijk

eenvoudig te automatiseren: het is precies te voorspellen wanneer ze plaatsvinden en het gegevensbestand is betrouwbaar. AP legde de lat voor zichzelf ook een stuk hoger toen ze besloten in 2014 ook de financiële

kwartaalreportages van beursgenoteerde bedrijven te automatiseren. Hiervoor maken ze gebruik van gegevens van het bedrijf Zacks Investment Research. Hier halen menselijke analisten data over omzet en

### FACTCHECK

## Ook feitenchecker kan nog niet zonder de mens

**Een presidentskandidaat doet bij een live debat een feitelijke uitspraak** en direct verschijnt op tv een rode tekst in beeld: fout!

**Het is nu nog geen werkelijkheid**, maar volgens de Britse organisatie Full Fact is de komst van automati-

sche factcheckers nabij.

**Veel bewijs voor die stelling is er echter nog niet.** De Amerikaanse publieke omroep NPR experimenteerde tijdens het eerste verkiezingsdebat in september met een live fact-check. De omroep gebruikt

een online transcript, dat steeds werd aangevuld. Het duurde telkens 1 à 2 minuten voordat uitspraken van Clinton en Trump waren verwerkt, met dank aan menselijke typisten.

**Een team van twintig journalisten** ging daarna

aan de slag met het snel checken van de uitspraken. Software van de Universiteit van Texas, *Claimbuster*, kan op basis van geschreven tekst aangeven of passages 'check-waardig' zijn. Maar ook in dat geval moet de mens, vooralsnog, de werkelijke check doen.

winst uit pdf's van bedrijven en zetten dat in een database.

Het stelt AP in staat om veel meer kwartaalverslagen te publiceren dan voorheen. De verslaggevers beperkten zich tot de driehonderd grootste bedrijven, maar dankzij Wordsmith bericht AP nu over 3.700 bedrijven. Zonder menselijke tussenkomst komen de berichten terecht bij ruim 1.700 kranten en 5.000 radio- en tv-stations wereldwijd. Vooral door de snelheid en accuraatheid hoopt AP sneller dan de concurrentie te zijn en zo van waarde te zijn voor hun lezers; de berichten kunnen investeringsbeslissingen van miljoenen beïnvloeden.

### Uitglijder

Het kostte een jaar om het systeem op de rit te krijgen, vertelt Lisa Gibbs, bij AP verantwoordelijk voor de fi-

nanciële berichtgeving. „Bedrijven zijn irritant creatief in de manier waarop ze hun kwartaalberichten schrijven. Ze bedenken verschillende, soms verhullende termen om hun winst of hun omzet te omschrijven. De analisten bij Zacks pikken soms de verkeerde cijfers eruit. Fouten gebeuren als er mensen bij betrokken zijn.”

Vorig jaar maakte AP een uitglijder, toen de computer abusievelijk schreef dat de winstcijfers van Netflix tegenvielen. Maar door een recente aandelensplitsing had de computer het cijfer verkeerd geïnterpreteerd; de winst was juist boven verwachting. Ondertussen hadden onder meer de *Los Angeles Times* en CNBC het bericht al overgenomen.

Zo lang mensen nog nauw betrokken moeten zijn bij het maken en onderhouden van de robotschrijvers is

### OVER DIT ARTIKEL

Dit artikel is het eerste in een tweeluik over robotjournalistiek. Volgende week verschijnt deel 2.

Deze publicatie is tot stand gekomen met steun van het Fond Bijzondere Journalistieke Projecten, in het kader van het project Nieuwe Journalistiek: [www.nieuwejournalistiek.nl](http://www.nieuwejournalistiek.nl).

van kunstmatige intelligentie nog niet echt sprake. Bij Automated Insights - niet toevallig staat de afkorting van hun naam ook voor *artificial intelligence* - denken ze dat dat snel gaat veranderen. Long: „Kunstmatige intelligentie is nu nog op het niveau van het uitvoeren van programmeerregels en patroonherkenning. We zijn nog niet zover dat computers zelf kunnen nadenken.”

### Hij speelt vals

De volgende stap is dat de computer zelf de tekstsjablonen kan schrijven. Maar daarvoor moet het de vocabulaire van elke individuele sport en marktsector goed kennen. Hij kan al wel vrij accuraat aangeven of een woord een werkwoord of een zelfstandig naamwoord is.

Wordsmith wordt nu uitgerust met een optie om samenvattingen te maken van teksten. Daarvoor moet hij teksten begrijpen, zou je denken. Long: „Hij speelt vals. Hij knipt er alleen belangrijke woorden uit. We leren hem termen te herkennen, bij een weerbericht zijn dat bijvoorbeeld: zonnig, 77 graden en Durham - vaak zijn dat de woorden die uit de database komen.” De software kan zo bijvoorbeeld een blog samenvatten in een tweet. Deze techniek, *natural language processing*, is een voorzichtige vorm van *machine learning*. Wordsmith kan ook zelfstandig Wikipedia en nieuwsites afstruinen op zoek naar synoniemen. Maar pas als de computer teksten begrijpt kan hij ook zelf leren hoe zinnen worden opgebouwd en inschatten wanneer hij het moet hebben over een homerun of een driepunter. „Ultieme machinelearning, zodat de klant alleen maar data hoeft aan te leveren, is een paar jaar verwijderd”, aldus Long.

Ligt een Pulitzerprijs voor een robot dan binnen handbereik, zoals Kris Hammond, de CTO van concurrerend bedrijf Narrative Science voorspelde? „Dat slaat nergens op”, zegt Long. „Die uitspraak was alleen maar bedoeld om krantenkoppen mee te halen. Coproductie van mens en robot is het beste. Dan kun je de snelheid en schaalgrootte van robots combineren met dat waar mensen fantastisch in zijn, zoals complexe analyse en nuance.”

Hammond kreeg sowieso ongelijk. Volgens zijn voorspelling zou een computer dit jaar al een Pulitzerprijs zou winnen. Maar met de huidige stand van zaken is dat toekomstbeeld nog mijlenver verwijderd.