

]

‘Een boek is voor ons de ultieme *big data*’

Interview met Bookarang, zes maanden na het winnen van de internationale GAU competitie *Renew the Book*

Boeken zijn net mensen. Uiteenlopend van stijl, sfeer, niveau, complexiteit enzovoort. En lezers zijn ook net mensen: met ervaringen, stemmingen en voorkeuren. Hoe zorg je ervoor, in deze dubbelcomplexe werkelijkheid, dat een lezer een boek vindt dat hem of haar raakt? Bookarang, een jong ICT-bedrijf in Amsterdam, vindt het antwoord op die vraag bij kunstmatige intelligentie. Het won er de door de GAU uitgeschreven competitie *Renew the Book* mee.

Bookarang is een klein bedrijfje in Amsterdam dat zich niet eenvoudig laat categoriseren. Zelf noemt Victor Bergen Henegouwen, een van de oprichters, het bedrijf “de perfecte combinatie van alfa’s en bèta’s. Alfa’s zijn mensen die schrijven en lezen en daarover nadenken, en bèta’s zijn de mensen die denken in nullen en enen. Bookarang brengt die twee samen. We hebben hier experts op het gebied van kunstmatige intelligentie en redacteurs die boeken analyseren.”

In december 2015 won Bookarang de competitie *Renew the Book* van de Groep Algemene Uitgevers (GAU). In deze internationale competitie voor *start-ups*, gericht op innovatie van het boekenvak, waren vijf vernieuwende bedrijfjes uit verschillende landen genomineerd. De jury koos Bookarang als meest innovatieve.

Aanbeveling op inhoud

Bookarang ontwikkelt technologie om literatuur te analyseren en op basis daarvan inhoudelijke boekaanbevelingen te doen. Maar wie bij ‘boekaanbevelingen’ denkt aan het bekende rubriekje bij online boekwinkels ‘Anderen kochten ook...’ zit er in het geval van Bookarang totaal naast. Bergen Henegouwen: “Iedereen weet dat die manier van aanbevelen alleen een beetje werkt als je Amazon heet. Bij de meeste andere webwinkels zijn de aantallen *views* per boek vaak veel te laag. Bovendien is de data vervuild doordat mensen cadeaus voor anderen kopen.” Medeoprichter Frederique Schut van Bookarang: “Dat ‘anderen kochten ook...’ gaat over het gedrag van anderen. Dat is wat men steeds doet: wat lezen je vrienden? Wat heeft de redactie voor je op een rijtje gezet? Wat onze technologie onderscheidt is dat wij uitgaan van het boek zelf en dat we de boek-kenmerken leggen naast de persoonlijke voorkeur van de consument. Ons advies is dus uitsluitend gebaseerd op de inhoud van het boek, niet op verkoopcijfers.”

Low level features

Het idee van Bookarang begon overigens niet bij geschreven tekst, maar bij muziek. Bergen Henegouwen is al langer betrokken bij een bedrijf dat audio analyseert. “We kijken hoe muziek klinkt en of het lijkt op iets anders. Dat doen we voornamelijk

door het meten van zogeheten 'low level features': ritme, tempo, toonsoort, timbre en allerlei andere dingen die je kunt meten. Vervolgens kijken we naar de perceptie van mensen: dit stukje klinkt als dat stukje. En dat vergelijken we met die *low level features*. Dan zie je patronen en overeenkomsten. Op een gegeven moment stop je een nieuw stuk muziek in die computer, en die kan dan bepalen waar het op lijkt. We gebruiken dus de menselijke input om de algoritmes te trainen."

"Toen ik Frederique hierover vertelde, zag ze meteen de potentie. Zij komt uit de boekenwereld en haar eerste reactie was: zoiets zou je voor boeken moeten hebben. Voor ons was het eigenlijk meteen duidelijk dat dat technisch haalbaar was."

Het boek als *big data*

De literaire nuances van een boek onderbrengen in een systematiek, dat lijkt onbegonnen werk. Om nog maar te zwijgen over de perceptie en de smaakvoorkeuren van een lezer. Maar volgens Bookarang kan het best. Bergen Henegouwen: "Het is misschien een beetje pijnlijk voor de alfa's onder ons, die meestal vinden dat literatuur kunst is en kunst kun je niet in patronen vangen. Wij laten zien dat het wel kan."

Hoe het werkt? Schut: "We meten heel concrete zaken in een boek: het aantal punten, de gemiddelde zinslengte, de uniciteit van woorden. Enkele tientallen van die *low level features*. Dat kun je geautomatiseerd meten. En dan blijkt dat er clusters zijn in de data, patronen die wij zelf helemaal niet ervaren als we lezen. Daarnaast kijkt een redacteur naar allerlei andere parameters. Hoe intelligent is een boek, hoe literair, hoe grappig enzovoort. Die analyse is verregaand gestructureerd en genormaliseerd. Wij kijken dus alleen maar naar het de inhoud van het boek. De volledige tekst is voor ons de ultieme *big data*."

Machine learning

Wanneer een nieuw boek uitkomt bij een van de uitgevers waarmee Bookarang samenwerkt (dat zijn er inmiddels tientallen), wordt het automatisch geanalyseerd met eigen algoritmes en daarnaast systematisch gelezen door de redactie. Die twee analyses komen samen en zo ontstaat per boek een set objectieve data. En dan? Schut: "De data gebruiken we om toepassingen te ontwikkelen voor retailers en bibliotheken. Daarbij kun je denken aan online boekentippers en alternatieven voor dat 'Andere mensen kochten'. Maar we hebben ook schermen ontwikkeld voor fysieke boekwinkels, waar je als consument persoonlijk boekadvies krijgt. In alle gevallen gaat het erom de lezer te helpen bij het vinden en ontdekken van het juiste boek."

Bookarang werkt daarnaast aan het volledig automatiseren van het analysesysteem. Bergen Henegouwen: "We gebruiken de handgeannoteerde data om de computer te trainen in het voorspellen van uitkomsten. Dat is letterlijk *machine learning*. Buiten het boekenvak zie je dat ook; Google gebruikt een soortgelijke benadering bij fotoherkenning en ook bij de zelfrijdende auto. Dit is de techniek van nu; iedereen is ermee bezig."

Renew the book

Voor Bookarang kwam de nominatie voor *Renew The Book* min of meer als geroepen. Niet dat het bedrijf noodlijdend was; het had amper twee jaar na de start al omzet, producten en klanten (waaronder CPNB, AKO en ECI). Maar het ging allemaal veel langzamer dan Schut en Bergen Henegouwen en hun algoritme-expert Niels Bogaards hadden gedacht. “Een klassieke fout voor *start-ups* geloof ik”, lacht Schut. “We zijn hier ontzettend *technology-driven*. Het werkt, dus waarom staan mensen niet in de rij? We zijn soms heel ongeduldig.”

Bij de competitie hoorde een *bootcamp* van enkele weken. In die periode werden de genomineerde bedrijven intensief gecoacht door Rockstart (gespecialiseerd in de begeleiding van innovatieve *start-ups*) en voerden ze gesprekken met mentoren uit het werkveld. Bookarang ontdekte zo zijn eigen blinde vlek. “Onze oplossing is wel in orde. Maar we moeten het eerst met de klant hebben over zijn vraag. Pas als je die scherp hebt, kun je over oplossingen gaan praten.”

Geen limiet

Renew the Book heeft voor Bookarang verschil gemaakt, zegt Bergen Henegouwen. “Iedereen in de boekenbranche kent ons nu als een serieuze partij. We ontvingen de prijs in december en sindsdien zijn processen behoorlijk versneld. Potentiële klanten weten ons te vinden en ook uitgevers zien de potentie: onze technologie kan hen helpen om ook de backlist gericht onder de aandacht te brengen in plaats van alleen bestsellers en nieuwe titels. Dus onze positie is steviger geworden. Dat is belangrijk, want we willen onze technologie blijven ontwikkelen en blijven investeren in het automatiseringsproces. Als we onze analyses helemaal geautomatiseerd hebben, dan is er geen limiet meer. Dan lonkt het buitenland.”

Victor Bergen Henegouwen: “Mijn oudere broer studeerde Nederlands en gaf mij als tiener boeken waarvan hij dacht dat ik ze *cool* vond. Dat heeft mij tot lezen gebracht. Dat eenvoudige gebaar, een boek aanbevelen waardoor iemand de liefde voor lezen ontdekt of herontdekt, dat kan onze technologie ook. Technologie kan mensen dus echt aan het lezen krijgen.”

Frederique Schut: “Sinds we met deze technologie bezig zijn heb ik geen boek meer gelezen dat tegenviel. Als ik dat zeg, zijn mensen verbaasd. Ze zijn eraan gewend dat boeken vaak een beetje tegenvallen. Maar dat hoeft helemaal niet. Je kunt boeken ontdekken die je helemaal niet kent en die toch bij je voorkeur passen. Onze technologie helpt daarbij.”