

# Tekstmining succesvol ingezet bij palliatieve database

10 mei 2023



Beide onderzoekers waren benieuwd wat natural language processing ofwel tekstmining als techniek voor de zorg te bieden heeft. Deze techniek is namelijk krachtig en veelbelovend, maar ook nog relatief nieuw. Daardoor weet nog lang niet iedereen wat je er in de zorg eventueel allemaal mee kan. De database PRADO was ideaal voor hun onderzoek omdat deze gestructureerde en ongestructureerde info bevat. De ongestructureerde informatie staat in vrije tekstvelden, die lastig was te analyseren. Maar dankzij tekstmining wordt dit nu dus steeds eenvoudiger.

Wikipedia geeft een heldere definitie van tekstmining : ‘Het betreft een [proces](#) om met ICT-technieken waardevolle informatie te halen uit grote hoeveelheden tekstmateriaal. Concreet gaat men teksten softwarematig structureren en ontleden, transformeren, inbrengen in databanken en uiteindelijk interpreteren.’

## **Exploratief onderzoek met tekstmining**

Het is zonder meer handig voor zorgprofessionals als ze met speciale tools, zoals tekstmining, ongestructureerde data uit vrije tekst kunnen destilleren en structureren om er echt [inzichten](#) uit te halen. Juist in die ongestructureerde data staat namelijk enorm veel gegevens die nodig zijn om inzichten te krijgen in bijvoorbeeld behandelingen. Die analyses kunnen bovendien gekoppeld worden aan de gestructureerde data waar men over beschikt.

Op de [website](#) van IKNL, waar beide onderzoekers werken, vertelt Grootendorst:

“Tekstmining, ook wel natural language processing, is een techniek waarmee je ongestructureerde data, zoals grote hoeveelheden tekst, om probeert te zetten naar gestructureerde bruikbare data. Het maakt niet uit of dat nou mdo’s, tweets of PRADO-verslagen zijn.’

Bijzonder was dat de wetenschappers niet eerst een onderzoeksvraag hadden of een hypothese formuleerden. In plaats van traditioneel gingen ze exploratief te werk en lieten ‘de data spreken’. Met behulp van topic modelling lukte het om een overkoepelend beeld te krijgen van wat er in al die 90.000 verslagen in PRADO stond. Met al deze onderwerpen konden Grootendorst en Brom bijvoorbeeld de trends in frequentie van onderwerpen analyseren.

## **Schat aan info verborgen in PRADO**

Met behulp van classificatie deden Grootendorst en Brom een eerste verkenning.

Grootendorst: “We hebben 90.000 consultverslagen tot onze beschikking, dus het was belangrijk eerst een overkoepelend beeld te krijgen van wat er in die verslagen staat. Dat hebben we gedaan door middel van topic modeling. Zo konden we de, zonder daarbij zelf alle verslagen door hoefden te lezen, een goed overzicht krijgen van de inhoud. Mooi is dat we ook al die gestructureerde data erbij hebben.

“We weten zo wanneer consulten plaatsvonden en welke problemen en symptomen aan de vraag gekoppeld zijn”, vertelt Grootendorst. “Dus we kunnen zien of een onderwerp als mobiliteit vaker terugkomt over tijd, of dat palliatieve chemotherapie minder vaak terugkomt. Of we zien onderwerpen die we juist helemaal niet verwacht hadden”.

## **Intelligente zoekmachine & algoritme**

De ontdekte onderwerpen geven een goed beeld van de data die in de verslagen verscholen ligt. Vervolgens wilden Grootendorst en Brom nog specifiek zoeken met behulp van een intelligente zoekmachine. Met die specifieke techniek werd het mogelijk gemaakt om van ongestructureerde data, zoals tekst, naar gestructureerde data te gaan. Vervolgens konden Brom en Grootendorst bijvoorbeeld zien hoe vaak een bepaald concept, zoals het medicijn gabapentine, voorkwam bij welke problemen en symptomen. De laatste techniek die de onderzoekers inzetten voor de data-analyse was ‘named entity recognition.’ Op die manier ging een speciaal getraind algoritme op zoek naar concepten die niet zo duidelijk beschreven zijn, waardoor ze niet goed te vinden waren door de slimme zoekmachine.

## **Grote mogelijkheden tekstmining**

Met de informatie die aan de hand van topic modeling, de zoekmachine en named entity recognition werd verzameld, maakten de onderzoekers visualisaties die inzicht bieden in wat de database met palliatieve data allemaal te bieden heeft. Bekeken werd of de inzichten relevant waren voor consulenten in de palliatieve zorg en of de inzichten kunnen helpen bij het verder vormgeven van de palliatieve consultatie in Nederland.

Het interessante onderzoek liet goed de mogelijkheden zien die een uitgebreide tekstmining kan hebben bij het ontsluiten van (ongestructureerde) data. Onderzoeker Brom noemt echter ook een nadeel van dit onderzoek, te weten dat het urgentie miste. Dat kwam vooral omdat het onderzoek niet werd gedaan op basis van een concrete vraag uit het veld. Toch zijn de

uitkomsten veelbelovend en denken beiden dat tekstmining meer toepassingsmogelijkheden heeft in de zorg. “Je kan door middel van tekstmining bijvoorbeeld ook op een schaalbare manier gestructureerde informatie uit mdo-verslagen halen, of dat nou gaat over medicatiegebruik, symptomen of bijwerkingen.”