

# Onderzoeksvoorstel Bachelor Informatiekunde 2016

Niek de Visscher (10667474)

## **Inleiding:**

Ik heb van Maarten Marx de mogelijkheid aangeboden gekregen om een eerder onderzoek deels te repliceren op een nieuwe dataset, namelijk de Nederlandse troonredes van de afgelopen 200 jaar. Het onderzoek in kwestie is een onderzoek naar de "State of the Union" speeches uit Amerika. In dat onderzoek werd gekeken naar wat er veranderde binnen de speeches sinds ze voor het eerst werden gehouden en hoe de speeches er hedendaags uitzien. Er werd gekeken naar de verschuiving in taalgebruik en inhoud in de speeches. Dit werd gedaan om een beeld te krijgen wat de belangrijke onderwerpen in een speech waren en welke thema's behandeld werden over de jaren. De State of the Union is niet de enige speech die al lange tijd wordt gehouden. Zo ook de troonredes van ons eigen Nederland. Deze troonredes worden sinds 1818 jaarlijks door de koning/koningin gegeven. De troonredes kunnen gezien worden als een Nederlandse versie van de State of the Union. De troonredes bevatten echter vaak reflecties op het afgelopen jaar in plaats van een update over de staat van het land. De staat van het land wordt ook wel besproken, maar meer in de context van de wereld, namelijk een kijk naar de positie van Nederland in de wereld. Dit is echter niet altijd het geval geweest. Zo was de inhoud vroeger anders dan nu, net als het taalgebruik. Tijdens het onderzoek zal er gekeken worden naar de veranderingen in de troonredes over de jaren. Het gaat hier vooral om de onderwerpen en thema's die behandeld worden. Er zal allereerst niet specifiek worden gekeken naar de verandering in taalgebruik en grammatica, maar dit kan wel gebruikt worden om een beeld te vormen over de veranderingen van de troonredes.

## **Onderwerp beschrijving:**

In mijn onderzoek wil ik een deel van het Amerikaanse onderzoek repliceren gericht op de Nederlandse troonredes. Het gaat hier om de troonredes sinds 1818, toen de eerste troonrede werd gehouden. Ik zal hierbij veel gebruik maken van het onderzoek dat is gedaan voor de Amerikaanse State of the Union en kijken naar andere mogelijkheden om de dataset te analyseren. Om te kijken of de analyse automatisch gedaan kan worden via computers wil ik gebruik maken van de programmeertaal Python. Als het lukt zou hiermee de inhoud van en verandering van inhoud van teksten die over vele jaren zijn geschreven automatisch kunnen worden getoond.

## **(Voorlopige) Onderzoeksvraag:**

Hoe kan de verschuiving in onderwerpen/thema's over 200 jaar troonredes worden weergegeven doormiddel van automatische tekstanalyse?

## **Onderzoeksmethode/aanpak:**

Voor het onderzoek zal ik me de eerste maand vooral focussen op het onderzoeken van de verschillende mogelijkheden om de dataset en dus de teksten te kunnen analyseren. Dit zal zich ook vooral richten op de mogelijkheden die Python biedt om dit te automatiseren en vereenvoudigen. Dit automatiseren en vereenvoudigen zal worden geholpen door gebruik te maken van bestaande algoritmes/analyse technieken en Python-libraries die kunnen helpen bij het verwerken van de dataset.

## **Planning:**

Week 14 – Uitwerken onderzoeksvraag (Het gaat er hier ook om een duidelijk beeld te krijgen wat haalbaar is in het repliceren van het eerdere onderzoek)

Week 15 – Theoretisch onderzoek (vinden van manieren om de dataset te analyseren) /literatuur vinden

Week 16 – Theoretisch onderzoek / literatuur vinden

Week 17 – Theoretisch onderzoek / begin met data analyse doormiddel van python scripts

Week 18 – voortzetting data analyse

Week 19 – voortzetting data analyse

Week 20 – voortzetting data analyse

Week 21 – voortzetting data analyse

Week 22 – Resultaten analyseren / beginnen met het schrijven van de scriptie

Week 23 – schrijven van de scriptie

Week 24 – Conclusie en discussie

Week 25 – Afronding afstudeerproject

Week 26 – Afronding afstudeerproject

## **(Voorlopige) Referenties:**

Alix Rule, Jean-Philippe Cointet, and Peter S. Bearman. (2014)  
Lexical shifts, substantive changes, and continuity in State of the Union discourse, 1790–2014. *PNAS* 2015 112: 10837-10844.

Tausczik, Yla R., and James W. Pennebaker. "The psychological meaning of words: LIWC and computerized text analysis methods." *Journal of language and social psychology* 29.1 (2010): 24-54.

Salton, Gerard, and Christopher Buckley. "Term-weighting approaches in automatic text retrieval." *Information processing & management* 24.5 (1988): 513-523.