

Debat zoekmachine op Europese schaal

Aantal woorden: **980** | Ingeleverd op: **31-03-2023**

Asher de Jong
asher.de.jong@student.uva.nl
Universiteit van Amsterdam
Amsterdam, Nederland

Maarten Marx
M.J.Marx@uva.nl
Universiteit van Amsterdam
Amsterdam, Nederland

1 INTRODUCTION

Elke dag worden er weer debatten gevoerd in de politiek. Als iemand een debat wil volgen kan bij veel landen de (live) opname worden bekeken. Echter vergt het veel tijd om een debat te volgen en in dit geval wordt er nog van uitgegaan dat de kijker ook de gesproken taal spreekt. In dit onderzoek zal de focus liggen op een zoekmachine die de taalbarrière zou kunnen verlagen en waarbij makkelijk door duizenden debatten kan worden gefilterd op basis van trefwoorden.

Momenteel heeft het Verenigd Koninkrijk al een website waar kan worden gekeken naar de standpunten van politici (<https://www.theyworkforyou.com/>). Voor dit project is het idee om dit concept uit te breiden naar een website waarmee door debatten kan worden gezocht, die zijn gehouden in Europese landen verspreid over de afgelopen jaren. Inzage van de debatten zal mogelijk worden gemaakt door een zoekmachine die functioneel is op de achtergrond in combinatie met de ParlaMint dataset (ParlaMint: CLARIN ERIC, z.d.). Dit is een uitgebreide dataset met debatten die zijn gehouden in 17 verschillende Europese landen (ParlaMint: CLARIN ERIC, z.d.). Deze data is beschikbaar en overzichtelijk gemaakt door het ParlaMint project, dat gefinancierd is door Clarin European Research Infrastructure Consortium (ParlaMint: CLARIN ERIC, z.d.).

Als toevoeging zal er ook een mogelijkheid zijn om dingen met elkaar te kunnen vergelijken. Er zou bijvoorbeeld een

vergelijking gemaakt kunnen worden tussen partijen, landen en personen. Deze vergelijkingen zullen op basis van trefwoorden zijn. Ook zal er bij deze vergelijkingen een Ngram overzicht worden neergezet om een beter beeld te creëren bij de relevantie van de trefwoorden bij een bepaalde groep. Een voorbeeld van Ngrams is Google Ngram Viewer voor boeken, waaruit kan worden opgemaakt hoe vaak iets voorkomt (*y-as*) in boeken op een datum (*x-as*) (Google Books Ngram Viewer, z.d.).

2 ONDERZOEKSDOEL & RELEVANTIE

Dit project is gericht op het versimpelen van onderzoek door debatten van Europese landen, hierbij is de zoekmachine gericht op zowel geïnteresseerden als onderzoekers die op een snelle manier grote hoeveelheden aan debatten willen doorzoeken. De zoekmachine zou van simplistische aard moeten zijn om te vermijden dat er voorkennis nodig is voor het gebruik. Bij de realisatie van deze zoekmachine komt de mogelijkheid voor Europese burgers om zelf beter een mening te kunnen vormen door bijvoorbeeld te kijken naar wat politici te zeggen hebben, om een indruk te krijgen van waar een politicus daadwerkelijk over praat en of hij/ zij het daarmee eens is. Voor onderzoekers zal het makkelijker worden om analyses te doen door middel van de vergelijkingfunctie die de zoekmachine aanbiedt. Zo zou een onderzoeker nu met een simpele query en wat filters een Ngram kunnen bekijken over de

relevantie van een trefwoord, zonder eerst alle ParlaMint data zelf te hoeven verwerken (wat ook nog enigszins programmeerkennis vereist).

Bij de genoemde doelstelling en de daaropvolgende sociale en wetenschappelijke relevantie hoort de volgende onderzoeksvraag van dit project:

“Tot in hoeverre is het mogelijk om een zoekmachine te maken die zoekt door debatten op Europese schaal om standpunten te bekijken en vergelijken, waarbij geen voorkennis van de zoekmachine nodig is voor de gebruiker?”

Bij deze hoofdvraag horen de volgende subvragen om de hoofdvraag te kunnen beantwoorden:

- (1) In hoeverre geeft een Ngram een duidelijke visualisatie van trefwoord relevantie aan een onervaren gebruiker?
- (2) Waar moet rekening mee worden gehouden tijdens het maken van een zoekinterface?
- (3) In hoeverre is het mogelijk om door niet Engelse talen te zoeken met een andere ingevoerde taal dan de taal waarin wordt gezocht?

3 LITERATUURLIJST

Allereerst zal het boek *Search User Interfaces* worden gebruikt om te functioneren als inspiratie voor de zoekinterface die tijdens dit project ontworpen gaat worden (Hearst, 2009). Daarnaast zal worden gekeken naar al bestaande websites met uitgebreide search interfaces.

Verder zal er worden georiënteerd op hoe Ngrams kunnen worden toegevoegd aan een website (Nithyashree, 2022). Vervolgens wanneer de site functioneel is zal er onderzoek worden uitgevoerd om te kijken wat de beste

manier van toetsing is om de eerste subvraag te kunnen beantwoorden.

Verder zullen veel kleine problemen tijdens het programmeren moeten worden opgezocht op internet, waarbij veel van deze problemen een duidelijke oplossing op Stackoverflow hebben (Stack Overflow, z.d.).

4 ONDERZOEKSPLAN

Aangezien dit geen standaard onderzoek is, zal de aanpak van dit project ook anders zijn samengesteld. Om te beginnen is zal eerst de zoekmachine zelf werkend gemaakt moeten worden, door middel van preprocessing van data en door het opstellen van filters. Vervolgens moet de zoekmachine op een server worden gezet om het zoekplatform zelf niet te vertragen.

Wanneer de zoekmachine online staat en volledig functioneel is, dan is het tijd voor het maken van een *minimal viable product* uitwerking van de website die gebruik gaat maken van de zoekmachine. Tijdens het opbouwen van de website zal een klein veldonderzoek plaatsvinden om meer te weten te komen over gebruikelijke kenmerken van een goede zoek interface. Het veldonderzoek zal worden uitgevoerd door op websites te kijken naar hoe zoekfilters worden gebruikt en het boek van Hearst (2009) zal ook bij deze stap een grote rol spelen.

Als er een *minimal viable product* is, kan er worden gekeken naar hoe de vergelijkfunctie toegevoegd kan worden. Bij deze vergelijkfunctie zal ook het Ngram overzicht worden geïmplementeerd. De filters voor vergelijking zullen tijdens deze stap al zijn vastgesteld, aangezien dit in overweging genomen zal worden tijdens het ontwerpen van de zoekmachine zelf.

Nu alle functionaliteiten aanwezig zijn, kan er kwalitatief onderzoek worden uitgevoerd met de doelgroep. Hiervoor zullen mensen worden

gezocht zonder enige achterliggende ervaring en dit zal gebeuren door middel van convenience sampling. Deze mensen zullen een zoekopdracht krijgen die zij uit moeten voeren en achteraf zullen er nog een aantal vragen worden gesteld over de zoekervaring.

Bij dit onderzoek wordt er vanuit gegaan dat onderzoekers (die ook onderdeel zijn van de doelgroep) zich over het algemeen net zo goed kunnen redden met de zoekmachine, gezien hun achtergrondkennis.

REFERENTIES

Google Books Ngram Viewer – Google Product. (z.d.). <https://books.google.com/ngrams/>

Hearst, M. A. (2009). *Search User Interfaces*. CAMBRIDGE UNIVERSITY PRESS.

<https://searchuserinterfaces.com/>

Nithyashree, V. (2022, 1 december). *What Are n-grams and How to Implement Them in Python?*

Analytics Vidhya.

<https://www.analyticsvidhya.com/blog/2021/09/what-are-n-grams-and-how-to-implement-the-m-in-python/>

ParlaMint: CLARIN ERIC. (z.d.). <https://www.clarin.eu/parlamint>

Stack Overflow (z.d.). Stack Overflow. <https://stackoverflow.com>