WELKOM!
INSPIRATIESESSIE DEEL 1

HET VOORBEWERKEN EN ANALYSEREN VAN OPEN ANTWOORDEN UIT DE NSE

FEBRUARI 2024

avans hogeschool

Even voorstellen

Rebecca Hubers-Hamers

Data Scientist / Buitenpromovendus

r.hamers1@avans.nl

linkedin.com/in/rebeccahamers/

Arash Yadegari Ghahderijani

Data Engineer / Promovendus

a.yadegarighahderijani@avans.nl

linkedin.com/in/arash-yadegari/

De verwerking van ‘open antwoorden’

- Wat zijn open vragen?
- Wat is het doel van open vragen in onderzoek?
- Hoe verwerken wij de open antwoorden op dit moment en hoe zou dat beter kunnen?

Hoe doen we het nu?

Op dit moment leveren wij de open antwoorden onbewerkt
in een Excel file aan onze academies.

Dit is niet optimaal, gezien dit de academies **veel tijd** kost om doorheen te lezen.
Verder is het lastig voor ze om al die losse opmerkingen om te zetten naar **daadwerkelijke bruikbare actiepunten/inzichten** om hun onderwijs te verbeteren.

Dit proces willen we graag optimaliseren.

Avans NSE vorig jaar: 14.123 respondenten x (7+3) open vragen = 141.230 open antwoorden

Voorbewerking van antwoorden

Om de open antwoorden te kunnen presenteren in bijvoorbeeld een dashboard, zonder in problemen te komen met AVG-gerelateerde regels, moeten ze eerst **voorbewerkt** worden zodat ze geen persoonsgegevens of andere aanstootgevende content meer bevatten.
We noemen dit ook wel **pre-processing**.

Dit doen we aan de hand van **code** (dit kan met verschillende programmeertalen).

avans hogeschool

Technologieën

databricks

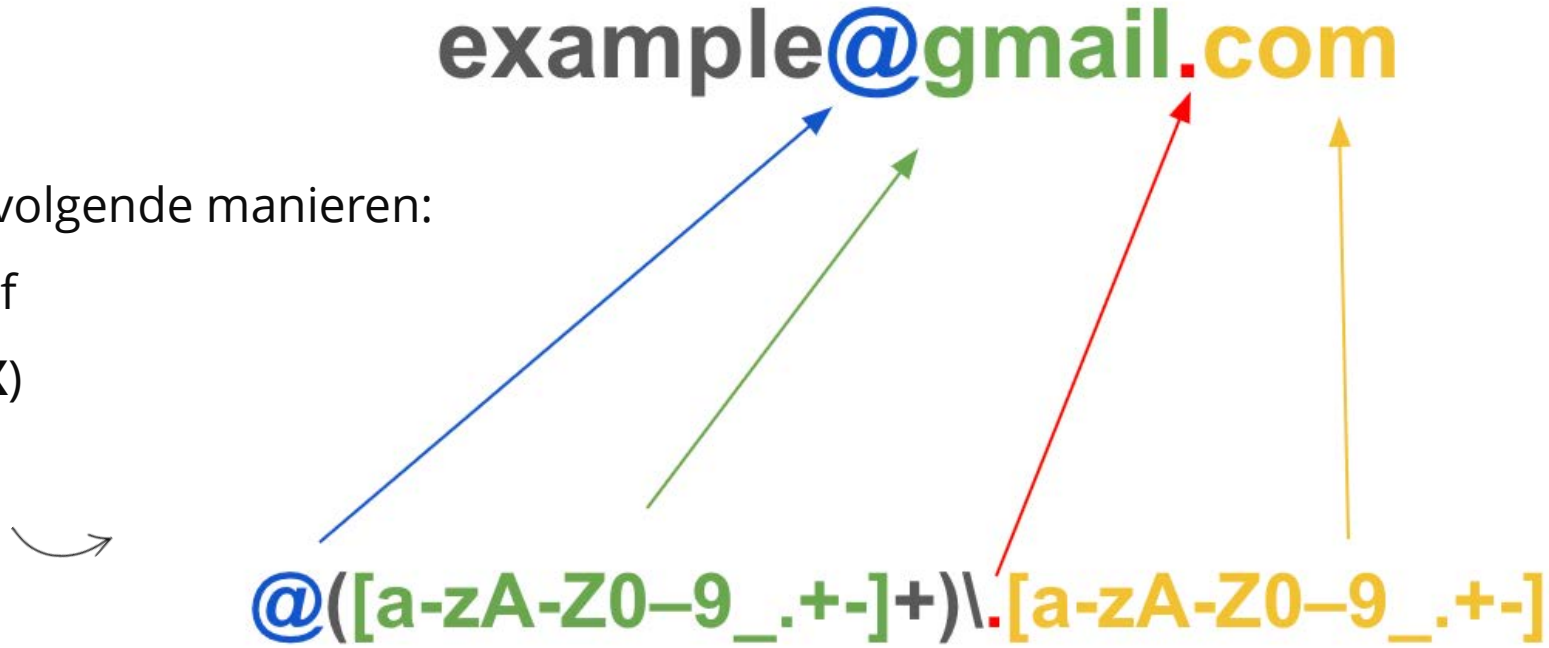
Waar zoeken we naar?

- **Persoonsgegevens** (voornamen, achternamen, adressen, telefoonnummers, e-mails)
- **Overig** (scheldwoorden, externe links)

We zoeken naar deze data op de volgende manieren:

- Filteren op **zoekwoorden** en/of
- Via regular expressions (**REGEX**)

  REGEX is een reeks speciale tekens en symbolen waarmee je complexe zoekpatronen kunt definiëren.



Kwaliteitscontroles

Maar hoe weet je zeker dat alle gevoelige informatie uit de open antwoorden gehaald zijn?

Dit doe je aan de hand van een laatste **menselijke check**.

Een computer werkt heel zwart/wit en voert dus alleen uit wat je van hem vraagt.

Echter zijn de open antwoorden zowel context- als spelfoutgevoelig.

Answer	KeywordsFound
Voor sommige opdrachten of projecten moet je zelf van tevoren al een bedrijf regelen. Dit vind ik altijd best wel vervelend, omdat je van tevoren zelf nog niet heel veel weet over de inhoud van de opdracht, dus dan kom je bij zo'n bedrijf eigenlijk ook met een best vaag verhaal aan waarom je dat bedrijf nodig hebt. Daarnaast vind ik het al helemaal vervelend om een bedrijf te moeten benaderen terwijl andere studenten dat ook doen, om vervolgens maar 1 bedrijf te kiezen. Het is best lullig als er...	lul
In mijn eerste jaar had ik veel verschillende begeleiders, dit vond kk niet heel fijn. Ik had liever 1 vaste begeleider gehad. Gelukkig heb ik dit nu wel in jaar 2!	kk

Ook houden we een **'white list'** bij. Dus sommige woorden die er in eerste instantie uitgefilterd worden, mogen in sommige contexten alsnog getoond worden.

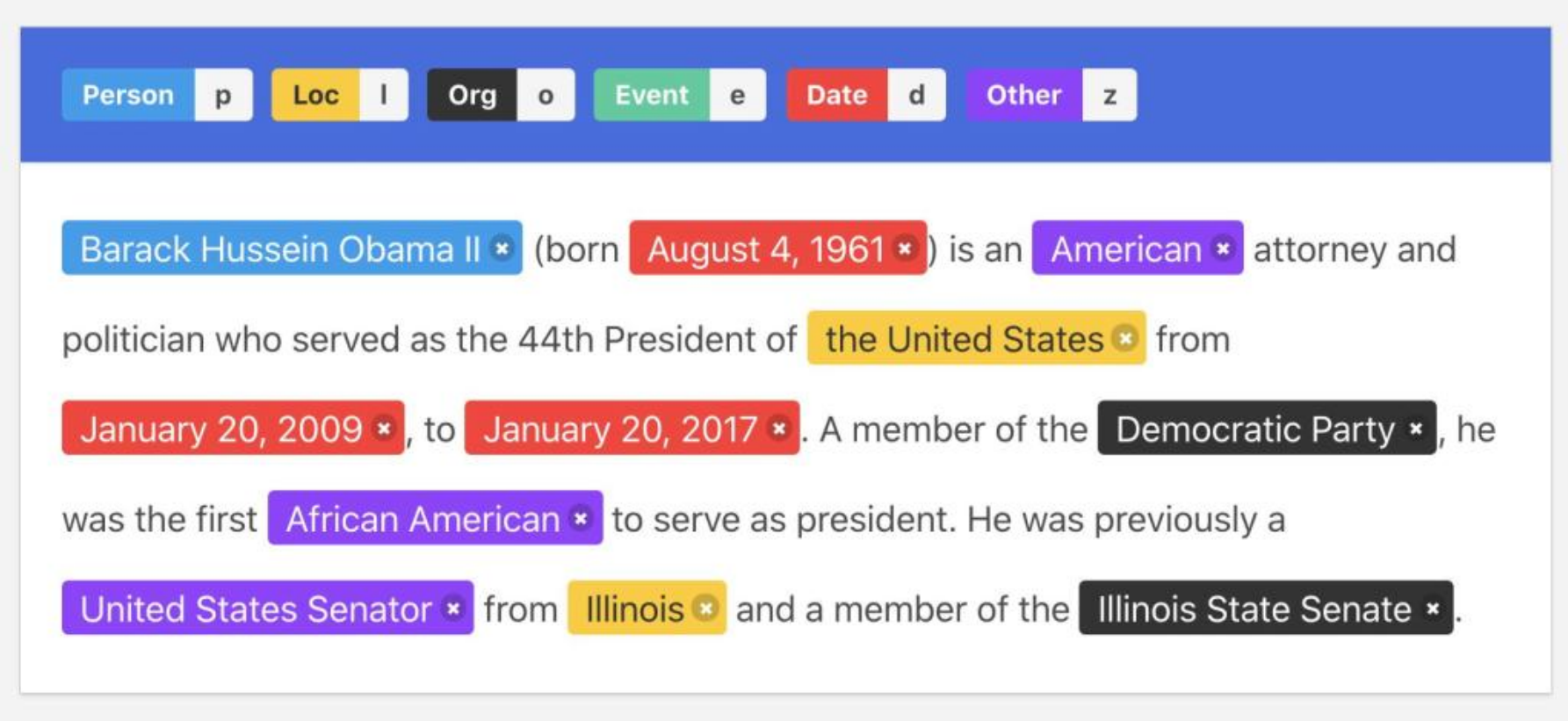
→ "Dan ben je de lul" is bijvoorbeeld anders dan "docent X is een lul"

Lessons learned so far

- Het ‘probleem’ van gevoelige (persoons)gegevens in de open antwoorden van de NSE is mogelijk een *kleiner* probleem dan we aan de voorkant inschatten;
- Deze automatische filteringen zo sluitend mogelijk maken klinkt makkelijker dan het in de praktijk is. Het aantal contextgevoelige situaties is groot en soms lastig om in een geautomatiseerd proces mee te nemen. De menselijke check is dus zeker nodig.
- Hoe definieer je de zoekwoorden? Hoe generiek/toegespitst maak je de REGEX?

Toekomst voor pre-processing

- Lijst van Avans medewerkers en studenten namen gebruiken voor naam-filter
- Aankomende experimenten voor pre-processing: **Named-entity recognition**



Next up...

UNIVERSITY
OF AMSTERDAM

Erik en Bart van de **Universiteit van Amsterdam** gaan jullie meer vertellen over verschillende analysetechnieken. O.a. sentiment analyse, topic modelling en het gebruik van AI.

VRAGEN?

Rebecca Hubers-Hamers

Data Scientist / Promovendus | r.hamers1@avans.nl | linkedin.com/in/rebeccahamers/

Arash Yadegari Ghahderijani

Data Engineer / Promovendus | a.yadegarighahderijani@avans.nl | linkedin.com/in/arash-yadegari/

Maatschappelijk betrokken. Ambitieus. Persoonlijk.

avans hogeschool