

18-12-2025



TECH REVIEW

Blok 2 - Team 2

Kaan Gögcay
Amber van Hassel
Amir Jacobs
Terence Terzol
Bibiëne Wüst

IN DEZE REVIEW

Huidige en nieuwe proces

Data

Datapreparatie

Modelkeuze

Modelarchitectuur

Evaluatie-metric

Resultaten tot nu toe

Conclusie tot nu toe

Volgende iteratie

HUIDIGE PROCES

- Twee-staps controle
- Herschrijven door middel van de OVER-schrijfwijzer (BI-niveau)
- Feedback van een andere medewerker

NIEUWE PROCES

- Schrijfwijzer assistent
- Onderbouwde suggesties op basis van de OVER-schrijfwijzer
- Medewerker controleert zelf

DATA (I)

OVER-schrijfwijzer

- Regels, gebaseerd op BI-niveau Nederlands
- PDF

OVER-gemeenten brieven

- Voorbeelden van niet herschreven brieven.
- De OVER-schrijfwijzer is nog niet toegepast
- Niet het gewenste eindresultaat

DATA (2)

De voorgaande data bevatte geen voorbeelden van teksten of woorden op BI-niveau Nederlands.

Externe data (nog niet gebruikt)

Dataset	Uitleg	Aantal woorden (synoniemen naar BI-niveau)
Moeilijke-Woordenboek van de Gemeente Amsterdam	Synoniemen lijst voor 'moeilijke'woorden	747
Juridisch woordenboek van de Gemeente Amsterdam	Lijst met alternatieve juridische termen op BI-niveau	134
Inclusieve woordenlijst van de Gemeente Amsterdam	Lijst met termen en hun bijbehorende inclusieve alternatieven	43
Synoniemen woordenlijst van BI-teksten.	lijst met versimpelde alternatieven	17

DATAPREPARATIE

OVER-schrijfwijzer

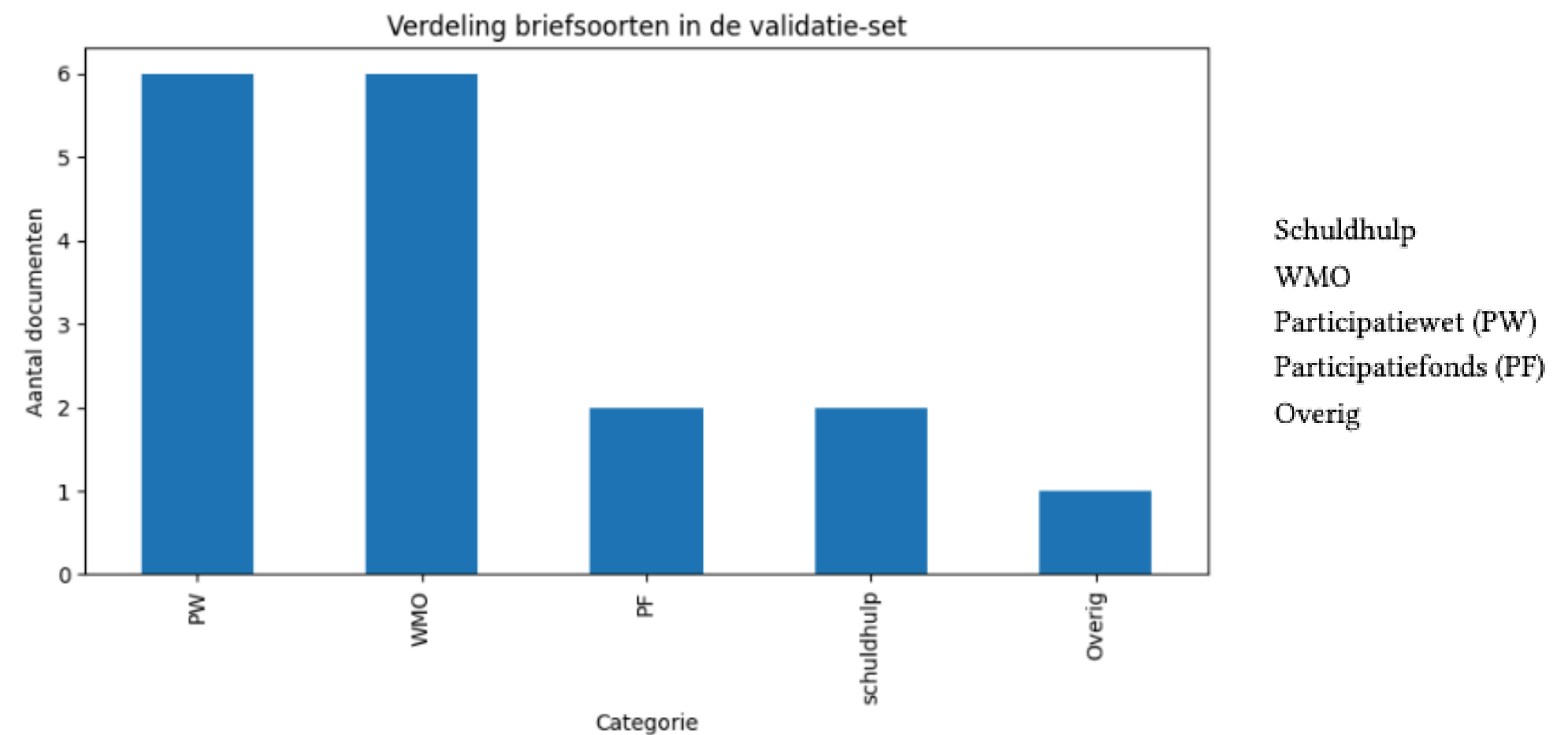
- Ingelezen (AMIR)

OVER-gemeenten brieven

- Verdeeld in vijf subgroepen
- Scheef verdeeld
- Generaliseerbaar?

Stratified-split

- Geen train-data (geen correcte voorbeelden)
- 50% val-, 50% test-split
- Brieven eerlijk verdeeld (stratified)



MODELKEUZE

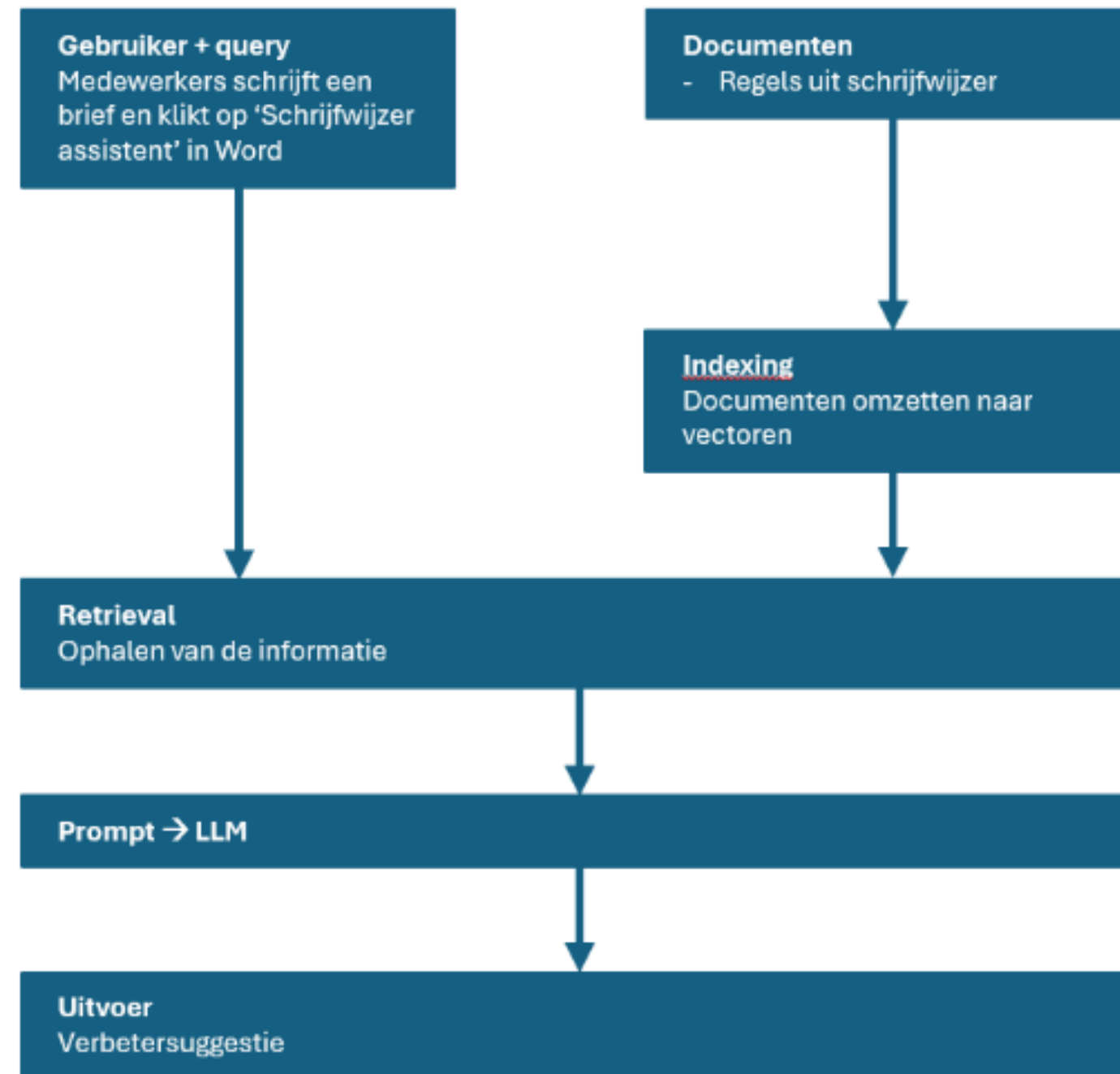
Naïve-rag

- LLM
- Granite-3.3-2b-model
- + OVER-schrijfwijzer

Baseline

- LLM
- Granite-3.3-2b-model
- Prompt

MODELARCHITECTUUR



EVALUATIE-METRIC

Metric	Uitleg	Voordeel	Nadeel
LiNT- -score	Speciaal ontwikkeld voor Nederlandstalige teksten. Gevalideerd met beoordelingen van docenten.	Bepaald de moeilijkheidsgraad van de tekst.	XXX
SARI	Bekijkt hoeveel een vereenvoudigde tekst afwijkt van het origineel.	Meet de kwaliteit van de vereenvoudiging.	Meet niet de feitelijke leesbaarheid voor eindgebruikers.
FKBLEU	Combinatie van parafrasekwaliteit en de leesbaarheidsindex.	Houdt met de twee kwaliteiten rekening.	Hangt minder samen met menselijke oordelen over eenvoud en is geneigd systemen te belonen die weinig wijzigingen aanbrengen in de oorspronkelijke tekst.

LiNT-II-score loopt van 0 - 100. Hoe lager de score, hoe dichter de tekst bij het BI-niveau ligt.

RESULTATEN TOT NU TOE (I)

- Geen, Baseline en naïve-rag model
- Validatie-set (brieven OVER-gemeenten)

Model	Gemiddelde LiNT-II-score	Gemiddelde verandering in LiNT-II-score
Geen model	47.61	-
Baseline model	45.83	1.78
naïve-RAG-model	45.75	1.86

- Laagste LiNT-II-score: naïve-RAG-model
- Verschil met baseline nihil

RESULTATEN TOT NU TOE (2)

- LiNT-II-scores per briefcategorie

Groep	Gem. LiNT-II-score originele brieven	Gem. LiNT-II-score herschreven brieven	Gem. verschil LiNT-II-scores
Participatiefonds (PF)	46.00	40.85	5.14
Participatiewet (PW)	45.49	41.91	4.12
Schuldhulp	57.17	50.02	7.09
WMO	47.27	45.72	1.69
Overig	43.85	39.16	4.68

- ‘Schuldhulp’ brieven behalen de hoogste LiNT-II-scores, maar daalt wel het meeste (niet-herschreven → herschreven)
- ‘Overige’ brieven behalen de laagste LiNT-II-scores

CONCLUSIE TOT NU TOE

Het toevoegen van de OVER-schrijfwijzer aan het naïve-rag-model voegt een nihile verbetering toe aan het model.

Het naïve-rag-model behaalt een gemiddelde LiNT-II-score van 45.75 op de herschreven brieven.

Het naïve-rag-model behaalt een kleine verbetering ten opzichte van het baseline model (0.08), wat verwaarloosbaar is.

Model presteert niet consistent over de verschillende briefsoorten. Schuldhelpbrieven tonen de grootste verbetering (7.09), terwijl WMO-brieven met slechts (1.69) verbeteren.



VOLGENDE ITERATIE

Het toevoegen van de externe data (synoniem lijsten) aan het model.

Onderzoeken op welke manier een advanced-rag-implementatie de resultaten verder kan verbeteren.

Implementatie van verwijzingen naar Oschrijfwijzerregels bij het genereren van suggesties.

Onderzoeken of het model als Word-extensie gebruikt kan worden.





VRAGEN?