



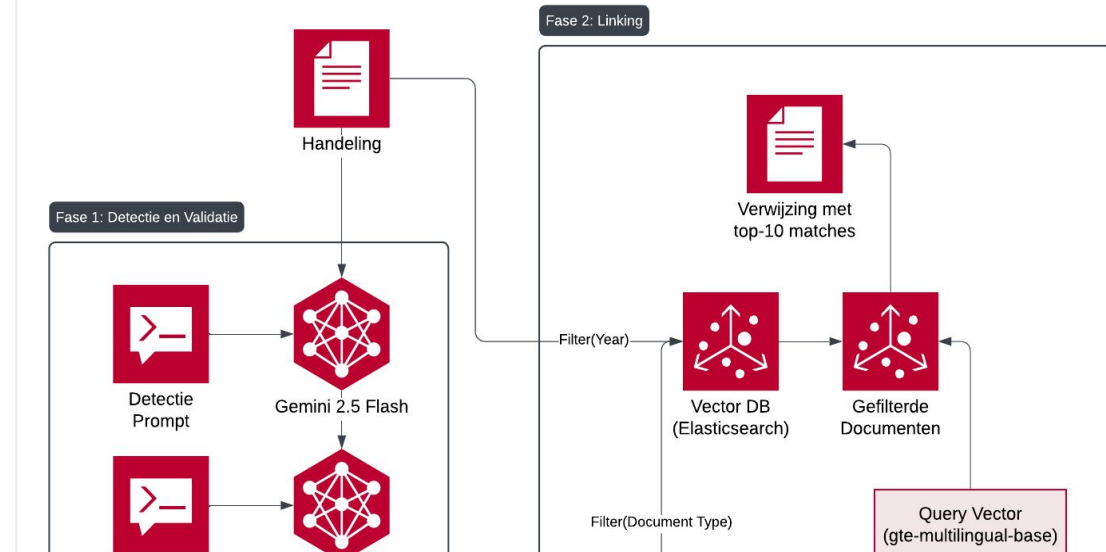
Een Masterthesis over het automatisch detecteren en linken van verwijzingen naar documenten in de Handelingen der Staten Generaal

De ‘Wikification’ van de Handelingen

Floris Bos

27 juni 2025

Master Computer Science (Joint Degree UvA/VU)



De voorzitter:

Ik constateer dat de aanwezige leden van de fracties van D66, de VVD en het lid Verdonk
amendement hebben gestemd en de aanwezige leden van de overige fracties ertegen, zo

Ik stel vast dat door de verworping van dit amendement het andere op stuk nr. 12 voork
als verworpen kan worden beschouwd.

In stemming komt het nader gewijzigde amendement-Hamer/Omtzigt (stuk nr. 33, I).

De voorzitter:

Ik constateer dat dit nader gewijzigde amendement met algemene stemmen is aangenom

Ik stel vast dat door de aanneming van dit nader gewijzigde amendement het andere op
voorkomende nader gewijzigde amendement als aangenomen kan worden beschouwd.

In stemming komt het amendement-Luijben (stuk nr. 20).

De voorzitter:

Ik constateer dat de aanwezige leden van de fracties van de SP en de PvdD voor dit ame
gestemd en de aanwezige leden van de overige fracties ertegen, zodat het is verworpen.

In stemming komt het amendement-Blok (stuk nr. 5).



Het Probleem (1/3)

Handeling: h-tk-20022003-3055-3080

De heer **Van der Staaij** (SGP):

In de brief stond dat dan in hun verblijf kan worden berust. Mijn vraag was gericht op opheldering. Betekent het inderdaad dat in die gevallen tot verlening van een verblijfsvergunning kan worden overgegaan?

Minister **Nawijn**:

Daarop kan ik volmondig "ja" antwoorden.

De motie-Vroonhoven-Kok zie ik als een ondersteuning van mijn beleid. Aan de hand van de criteria zal ik invulling geven aan de toepassing van de discretionaire bevoegdheid.

De heer De Wit vraagt in zijn motie om een specifieke regeling. Ik heb al gezegd dat ik niet voor een specifieke regeling ben. Ik moet dus aanvaarding van die motie ontraden.

Ook in de motie-Vos/Van der Laan wordt om een specifieke regeling gevraagd. Ook aanvaarding van deze motie moet ik ontraden.

Ten slotte moet ik als minister ook aanvaarding van de motie-Varela ontraden.

Mevrouw **Vos** (GroenLinks):

U zegt dat u als minister aanvaarding van de motie-Varela moet ontraden. Zou u daarover in een andere hoedanigheid nog een andere opvatting kunnen hebben?

Minister **Nawijn**:

Nee. Ik praat nu als minister. Ik moet aanvaarding van die motie ontraden, gelet op het kabinetsstandpunt.

Mevrouw **Vos** (GroenLinks):

Ik vraag het, omdat u dit zo nadrukkelijk zegt bij de motie-Varela. Bij het ontraden van aanvaarding van de andere moties deed u dit niet. Ik vroeg mij af wat de speciale betekenis van die toevoeging is. Ik moet er dus niet te veel waarde aan hechten?

Het Probleem (2/3)

Handeling: h-tk-20022003-3055-3080

De heer **Van der Staaij** (SGP):

In [de brief](#) stond dat dan in hun verblijf kan worden berust. Mijn vraag was gericht op op overgegaan?

Minister **Nawijn**:

Daarop kan ik volmondig "ja" antwoorden.

[De motie-Vroonhoven-Kok](#) zie ik als een ondersteuning van mijn beleid. Aan de hand van

De heer De Wit vraagt in [zijn motie](#) om een specifieke regeling. Ik heb al gezegd dat ik n

Ook in [de motie-Vos/Van der Laan](#) wordt om een specifieke regeling gevraagd. Ook aanv

Ten slotte moet ik als minister ook aanvaarding van [de motie-Varela](#) ontraden.

Mevrouw **Vos** (GroenLinks):

U zegt dat u als minister aanvaarding van [de motie-Varela](#) moet ontraden. Zou u daarover

Minister **Nawijn**:

Nee. Ik praat nu als minister. Ik moet aanvaarding van [die motie](#) ontraden, gelet op het ka

Mevrouw **Vos** (GroenLinks):

Ik vraag het, omdat u dit zo nadrukkelijk zegt bij [de motie-Varela](#). Bij het ontraden van a
Ik moet er dus niet te veel waarde aan hechten?

Kamerstuk: kst-19637-724

19 637

Vluchtelingenbeleid

nr. 724

MOTIE VAN HET LID VAN VROONHOVEN-KOK

Voorgesteld 18 februari 2003

De Kamer,

gehoord de beraadslaging,

overwegende, dat de minister bereid is zijn inherente afwijkingsbevoegdheid te gebruiken ten aanzien van een aantal
asielzoekers dat al langere tijd in Nederland verblijft;

van oordeel, dat voor deze asielzoekers zo spoedig mogelijk duidelijkheid moet worden geschapen;

verzoekt de minister voor schrijnende gevallen zijn inherente afwijkingsbevoegdheid te gebruiken op de door hem in
zijn brief van 17 februari 2003 aangeduide humane wijze;

verzoekt de minister tevens voor deze asielzoekers zo spoedig mogelijk duidelijkheid te verschaffen,

en gaat over tot de orde van de dag.

Van Vroonhoven-Kok

Het Probleem (3/3)

Expliciet (< 5%)^[1]



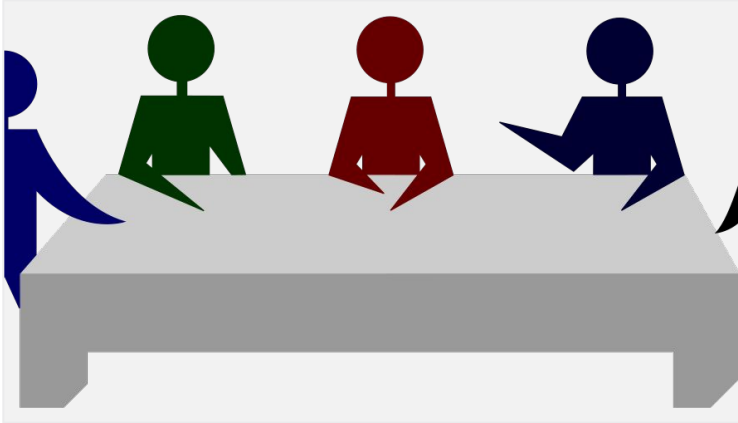
- “De motie-Moorlag (31532, nr. 248)”
- “de brief van de minister voor Vreemdelingenzaken en Integratie over het gebruik van zijn discretionaire bevoegdheid (19637, nr. 720).”

Impliciet



- “Hoe denkt mevrouw Hermann om te gaan met *de motie* zie zij in dit verband heeft ingediend?”
- “In november 2001 heeft de Kamer *een motie aanvaard* waarin de regering werd gevraagd haar een notitie toe te sturen.”
- “In *de brief* heeft u ook aangetroffen dat er gewerkt wordt aan een structurele aanpak”

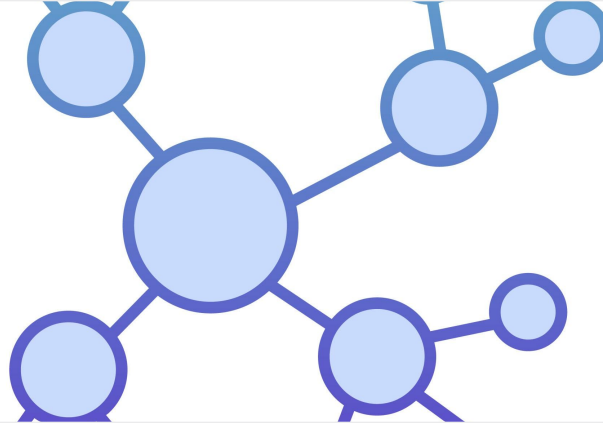
Maar waarom?



Maatschappelijk

Een overheid die transparanter en toegankelijker is

Bespaart journalisten, ambtenaren en burgers veel zoekwerk



Inzicht

Het creëert een knowledge graph van het parlementaire proces.

Analyseer welke moties invloedrijk zijn, welke onderwerpen 'hot' zijn, en hoe ideeën zich door de Kamer bewegen.



Uitdaging

Verwijzingen zijn vaag, dubbelzinnig en context-afhankelijk

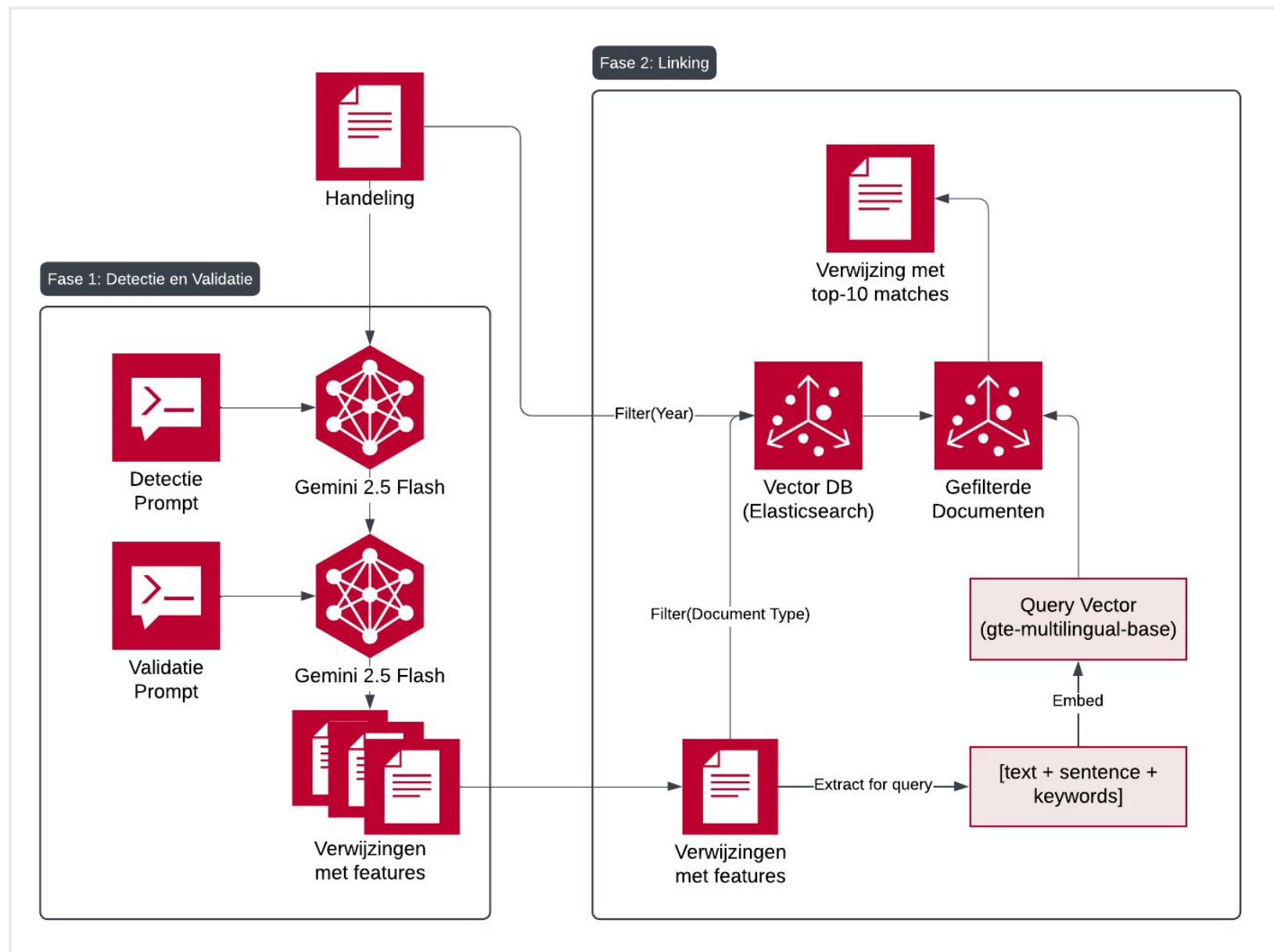
Het oplossen vereist een combinatie van NLP en IR

Uitleg

Twee phases

1. Detectie & Validatie

2. Linking

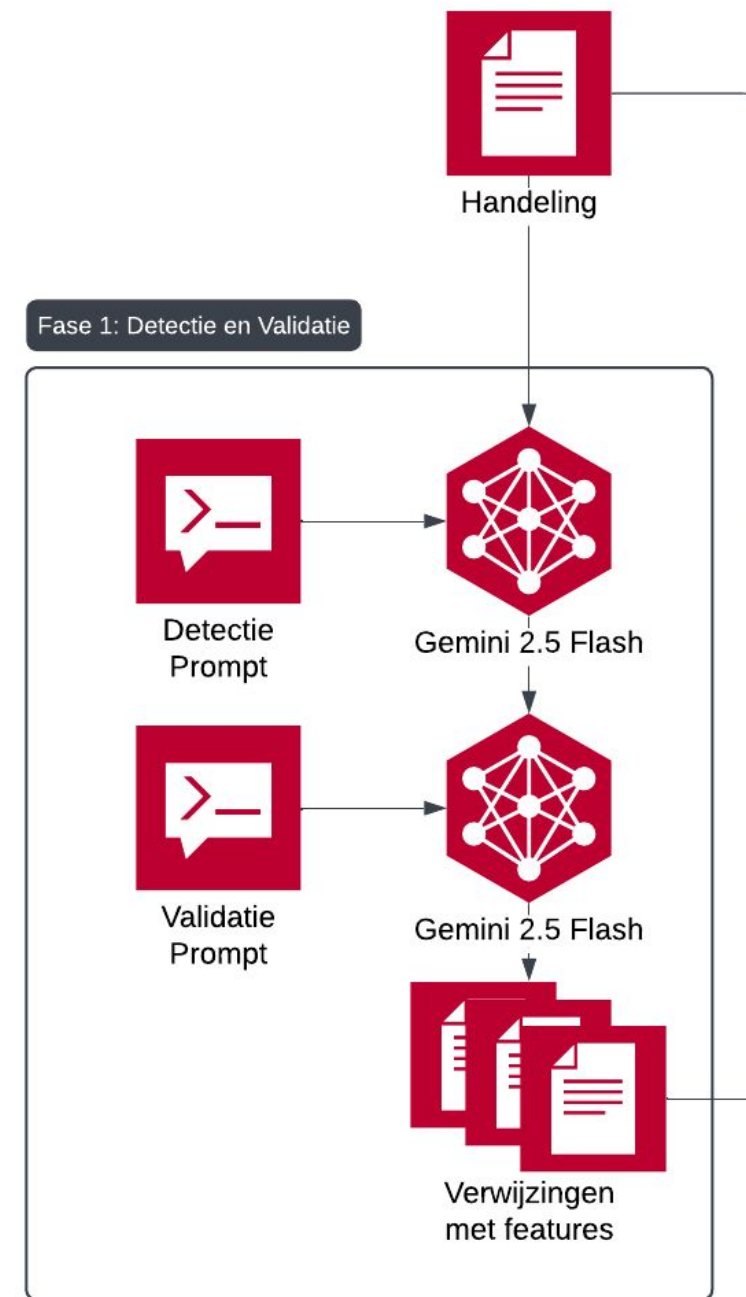


Uitleg

Detectie & Validatie

- Gebruiken een taalmodel om de verwijzingen te vinden
- Halen hier semantische features uit zoals de zin, referentietype, samenvatting, thematische keywords van TOOI waardenlijst^[5], en documenttype
- We proberen vier strategieën met twee assen:
 - Single vs. Two-pass
 - Zero-shot vs. Few-shot

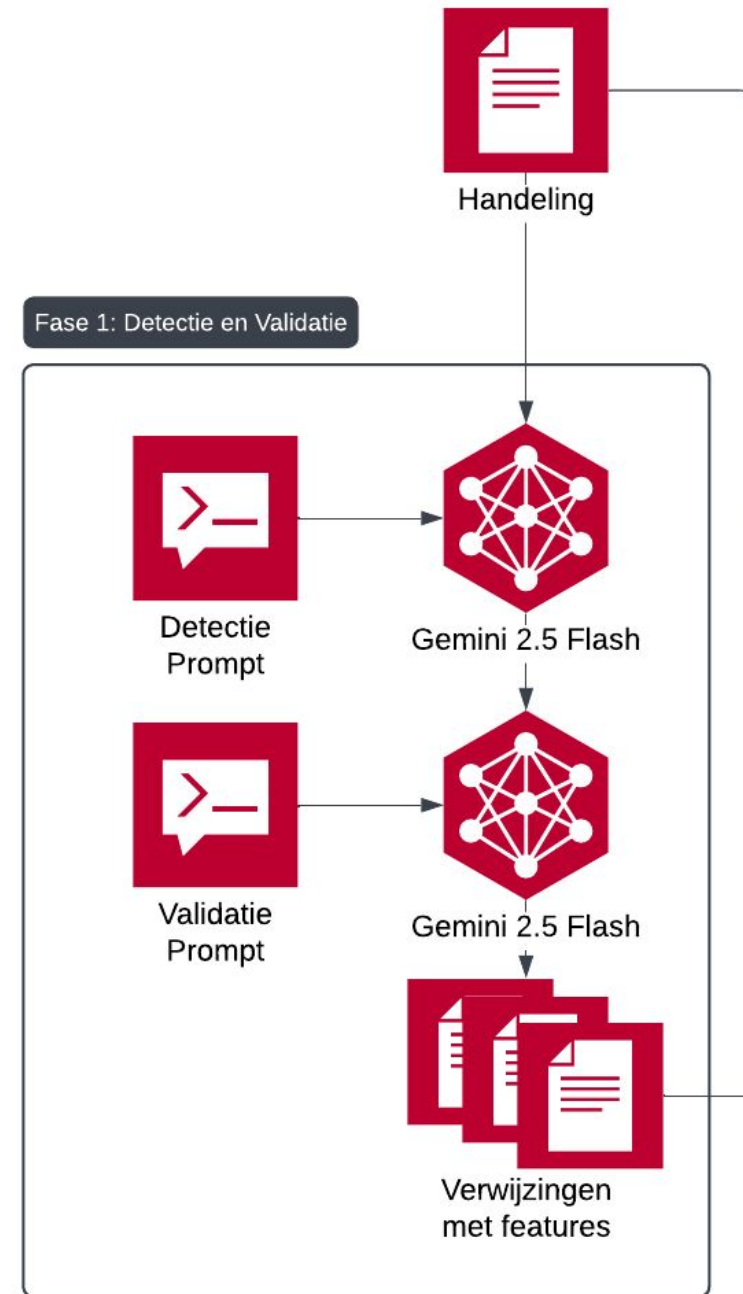
Techniek: Als taalmodel gebruiken wij *Gemini 2.5 Flash*^[2].



Experiment

Detectie & Validatie

- Als baseline een geannoteerde dataset van 5 handelingen
- Beste aanpak: Few-shot, Two-pass strategie
 - 40% van de gevonden verwijzingen is correct (Precision)
 - 47% van de verwijzingen wordt gevonden (Recall)
- Classificatie
 - Documenttype: 98%
 - Referentietype: 92%

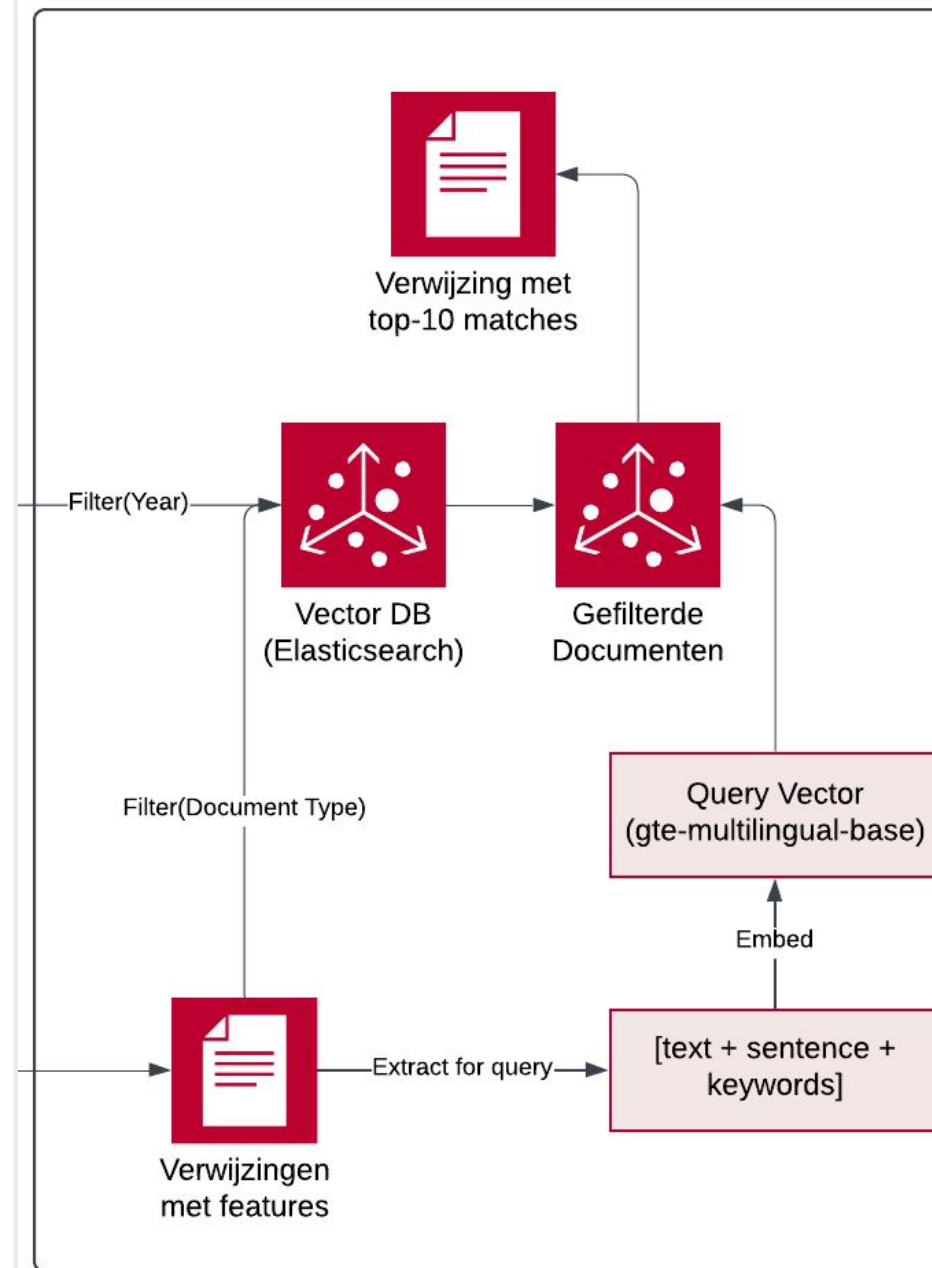


Wat is het?

Linking

- Vind voor een verwijzing het juiste document in een database met duizenden parlementaire stukken dmv. Vector Similarity Search
- Omzetten naar reeks getallen
 - Verwijzing (onze 'query')
 - Alle documenten in database
- Welke document-vector lijkt het meeste op onze query-vector?

Fase 2: Linking



Query Modulaire Bouwen

Linking



Documenttype	Motie
Tekst	de motie-Moorlag ()
Zin	constaterende dat naar aanleiding van de motie-Moorlag () er een landelijke, integrale aanpak voor problemen met schimmel en vocht [...]
Samenvatting	Verwijzing naar een eerdere motie van Moorlag c.s. over schimmel- en vochtproblemen.
Keywords	'woningmarkt', 'schimmel', 'vocht', volksgezondheid', 'gemeenten'

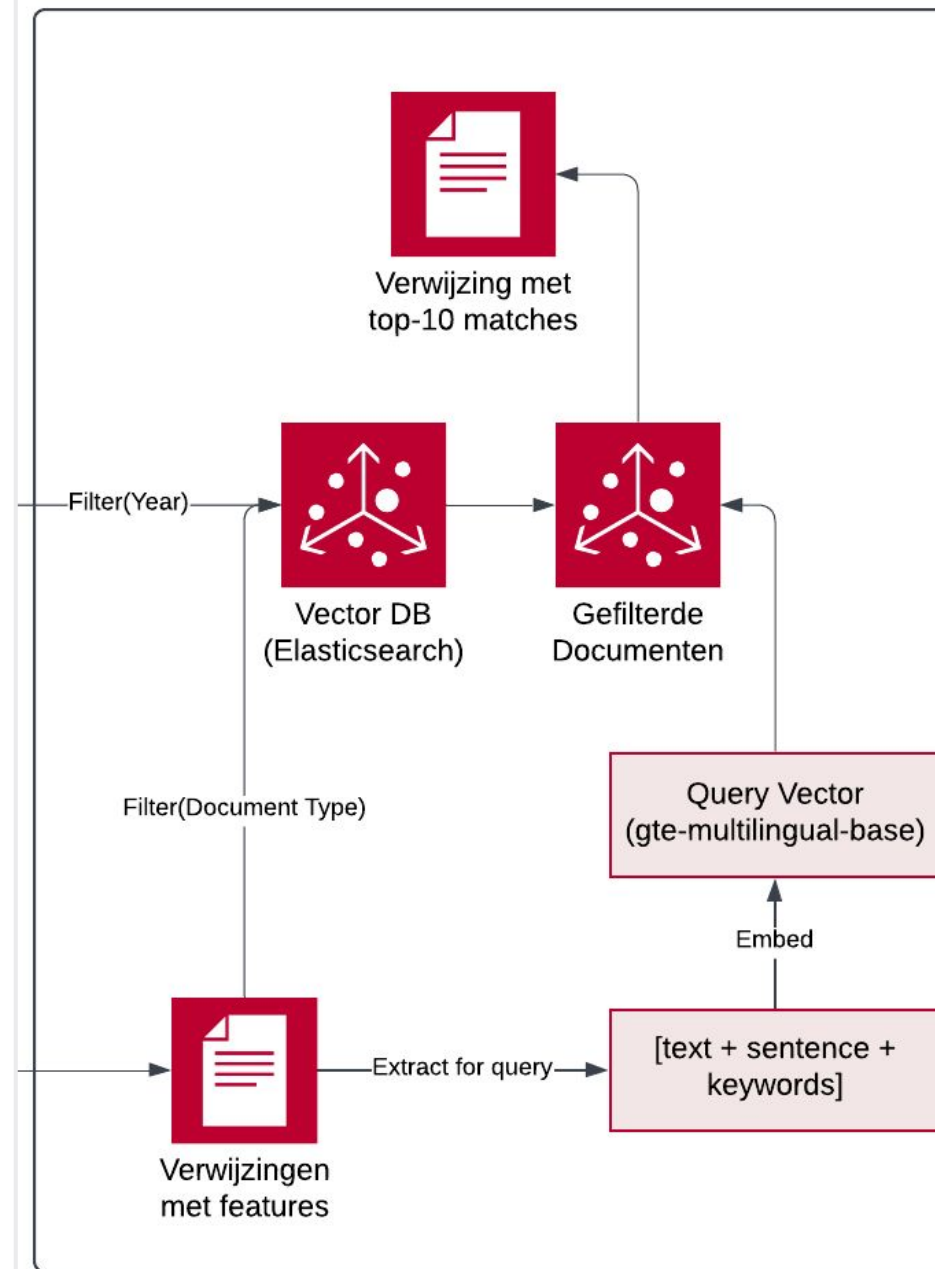
↓ Tekst + Zin + Keywords

Query: “de motie-Moorlag () constaterende dat naar aanleiding van de motie-Moorlag () er een landelijke, integrale aanpak voor problemen met schimmel en vocht [...] woningmarkt schimmel vocht volksgezondheid gemeenten”

↓ Embed

Vector: [-0.02, 0.05, -0.04, 0.01, ...]

Fase 2: Linking



Resultaten

Linking

Methode	Hit@1	Hit@10	Dossier Hit@1
Tekst en zin	3%	6%	13%
+ Samenvatting	14%	30%	27%
+ Keywords	26%	51%	43%
+ Keywords + Filters	35%	57%	50%

- Getest op 1.933 expliciete verwijzingen
- Keywords werken beter dan samenvatting
- Het filteren op jaar en documenttype heeft grote impact, vooral op Hit@10
- Combinatie van Keywords + Filters is de duidelijk de winnende strategie

Conclusie & Verder onderzoek

- Sterke baseline voor verder onderzoek
- De aanpak is taalafhankelijk en kan ook werken voor
 - Andere parlementen (bv. Europees Parlement)
 - Andere domeinen (juridische teksten, historische archieven)
- Verder onderzoek
 - Features: Meer features bedenken, metadata van kamerstukken gebruiken?
 - Re-ranking: Een tweede model dat de top 10 kandidaten slimmer rangschikt
 - Fine-tuning: Kleinere, efficiëntere modellen trainen voor betere detectie
 - Human-in-the-loop: Een feedback-systeem om het model continu te verbeteren



Vragen?

Kom vooral naar me toe in de pauze!

Bedankt!

Marc van Opijnen en Logius|Koop
Maarten Marx



Github: <https://thesis.florisbos.com>

Bronnen

- [1] Venema. P.: Improving Links to referenced Documents in Dutch Parliamentary Meeting Notes. Bachelor's thesis, University of Amsterdam, Amsterdam, The Netherlands (2024)
- [2] Google Cloud: Gemini 2.5 Flash | Generative AI on Vertex AI.
<https://cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/2-5-flash> (2025), accessed: 2025-05-27
- [3] Zhang, X., Zhang, Y., Long, D., Xie, W., et al.: mGTE: Generalized long-context text representation and reranking models for multilingual text retrieval. In: Proc. EMNLP (Industry Track) '24. pp. 1393–1412 (2024)
- [4] Elasticsearch: <https://www.elastic.co/elasticsearch> (2025), accessed: 2025-06-20
- [5] Thema-indeling voor Officiële Publicaties (TOP-lijst).
https://standaarden.overheid.nl/tooi/waardelijsten/work?work_uri=https%3A%2F%2Fidentifier.overheid.nl%2Ftooi%2Fset%2Fscw_toplijst, accessed: 2025-06-20