

# **Google Scholar zoekresultaten voor wetenschappelijke projecten: linked data & natural language processing**

**Bart De Paepe**  
**10 juni 2025**

**HO  
GENT**

# Inhoud.

Kader.

Methode.

Resultaten.

Discussie.

Conclusie.

**HO  
GENT**

**Kader.**

**HO  
GENT**

# Vlaams Instituut voor de Zee

Figuur: website url: <https://www.vliz.be>



**HO  
GENT**

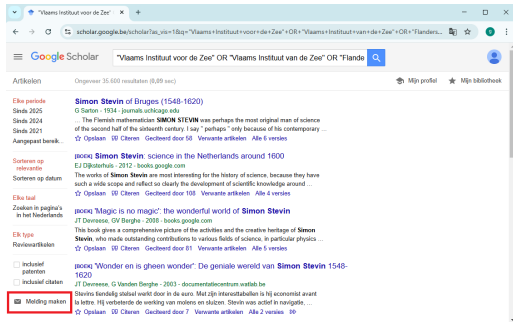
# Integrated Marine Information System

Raadpleeg IMIS

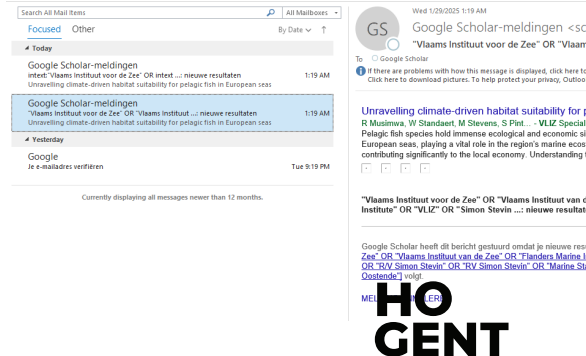
**HO  
GENT**

# Google Scholar

Figuur: zoekopdracht



Figuur: alert



# Het probleem.

D

Photo by Amin Zabardast on <https://unsplash.com/photos/a-street-sign-on-a-pole-next-to-a-building-Q57aftqNKiE>

**HO  
GENT**

# De onderzoeksvraag.

“Hoe kunnen de zoekresultaten van Google Scholar automatisch toegevoegd worden aan IMIS?”

- Hoe kunnen de uiteenlopende zoekresultaten van Google Scholar omgezet worden in gestructureerde data?
- Zijn alle zoekresultaten uniek identificeerbaar zodat er geen duplicaten opgeslagen worden?
- Hoe kunnen ook steeds nieuwe zoekresultaten van dezelfde zoekopdracht systematisch opgezocht worden?
- Kan er een score berekend worden hoe relevant elk zoekresultaat is ten opzichte van de zoekopdracht?
- Als er geen unieke identifier is, hoe wordt er dan gecontroleerd op duplicaten?
- Hoe blijft de tool overwegend onafhankelijk van third-party software?
- Hoe wordt er gegarandeerd dat de tool op de bestaande hardware kan draaien?



# Proof of Concept bouwen.

Geen grappige afbeelding met auteursrechten hier, wel [daar](#).

# Methode.

Google Scholar alert

✓ zoekopdracht  
instellen

Web scraping

✗ online LLM

✗ lokaal LLM

✓ Beautiful Soup

✗ Serpapi

Natural Language  
Processing

✓ relevantiescore

Linked data

✓ DOI in URL

✓ DOI in Crossref

✓ DOI op webpagina

Semantic search

✓ MongoDB Atlas Local

**HO  
GENT**

# Web scraping.

**HO  
GENT**

# Natural language processing.

**HO  
GENT**

# Linked data.

**HO  
GENT**

# Semantic search.

**HO  
GENT**

# Resultaten.

**HO  
GENT**

# Resultaten.

**HO  
GENT**



**Discussie.**

**HO  
GENT**

# Discussie.

**HO  
GENT**

**Conclusie.**

**HO  
GENT**

# Conclusie.

**HO  
GENT**