



GPT-3.5 metadata extraction

Metadata publicatie en extractie uit convenanten

Sander Oud
28 Juni 2024

Zoeken op de website

convenant

► Geavanceerd zoeken

Resultaat

 Raadsbrief Ondertekening [Convenant](#) Lokale & Duurzame Voedselketens provincie Utrecht Brieven

13 juni 2024 -- Raadsbrief Ondertekening [Convenant](#) Lokale & Duurzame Voedselketens provincie Utrecht

 [Convenant](#) Lokale & Duurzame Voedselketens Provincie Utrecht

13 juni 2024

Brieven : Raadsbrief Ondertekening [Convenant](#) Lokale & Duurzame Voedselketens provincie Utrecht

 Raadsbrief [Convenant](#) Lokale en Duurzame Voedselketens provincie Utrecht

13 juni 2024

Brieven : Raadsbrief Ondertekening [Convenant](#) Lokale & Duurzame Voedselketens provincie Utrecht

 2. Stand van zaken studentensport Utrecht

Raadsinformatiebijeenkomst - dinsdag 26 maart 2024 19:30

26 maart 2024 -- Tijd: 20...

Daarnaast is er in het uitvoeringsprogramma binnensport vastgesteld dat er een [convenant](#) komt voor de studentensport in Utre
De RIB heeft tot doel een helder beeld te geven van de actuele situatie van de studentensport in Utrecht, de uitdagingen te iden
samenwerking te verkennen...

Uitdagingen en mogelijke samenwerkingen: [Convenant](#) Studentensport (15 min)...

 2. WORDT VERPLAATST: Stand van zaken studentensport Utrecht

Raadsinformatiebijeenkomst - dinsdag 27 februari 2024 19:30

27 februari 2024 -- LET OP: Deze bijeenkomst wordt verplaatst naar een 26 maart 2024 van 20...

Daarnaast is er in het uitvoeringsprogramma binnensport vastgesteld dat er een [convenant](#) komt voor de studentensport in Utre
Uitdagingen en mogelijke samenwerkingen: [Convenant](#) Studentensport (15)...

Convenanten

- Afspraak tussen
 - Overheidspartijen
 - Particuliere partijen
- Publicatieplicht door WOO
- Semi-gestructureerde documenten
 - Tussen wie
 - Wanneer
 - Waarover
 - Doel

CONVENANT

Veilig varen doe je samen!



Provincie Zuid-Holland
Koninklijke Schippersvereniging Schuttevaer
Delftse Studenten Roeivereniging Proteus-Eretes
Delftsche Studenten Roeivereniging Laga
Roeivereniging De Delftsche Sport
Haagse Studenten Roeivereniging Pelargos
Roeivereniging De Laak
Roeivereniging Rijnland
Koninklijke Studenten Roeivereniging Njord
Leidsche Roei- en Zeilvereniging Die Leythe
Algemene Leidse Studenten Roeivereniging Asopos De Vliet
Roeivereniging Alphen
Koninklijke Nederlandsche Roeibond



provincie
ZUID HOLLAND

Onderzoeksvraag

In hoeverre is het haalbaar om semi-automatisch de convenanten van alle Nederlandse overheidsinstanties met bijbehorende metadata te verzamelen? En als de metadata niet beschikbaar is, hoe kunnen we deze dan effectief uit het document halen?

Deelvragen



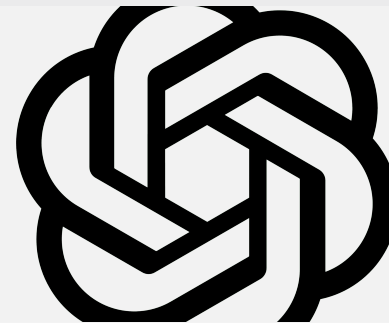
Ophalen van convenanten

Hoeveel convenanten zijn er en in hoeverre kunnen deze structureel van het web worden opgehaald om een dataset te vormen?

```
{  
  "doi": "nl.gm0867.Zi.2022.3",  
  "infobox": {  
    "aanbieder": "Gemeente Maalwijk",  
    "jaar": "2022",  
    "information category": "Zi",  
    "titel": "Opvanglocaties asielzoekers",  
    "beschrijving": "Woo verzoek gedaan op 14 september over de nieuwe  
    asielopvang van asielzoekers. Op 26 oktober is besloten  
    om gedeeltelijk te openbaren.",  
    "verzoeksdatum": "2022-09-14",  
    "besluitdatum": "2022-10-26",  
    "besluitcategorie": "Gedeeltelijk openbaar",  
    "Aantal vrijgegeven documenten": 16  
  }  
}
```

Beoordelen van Metadata

In welke mate worden convenanten gepubliceerd met de benodigde metadata?



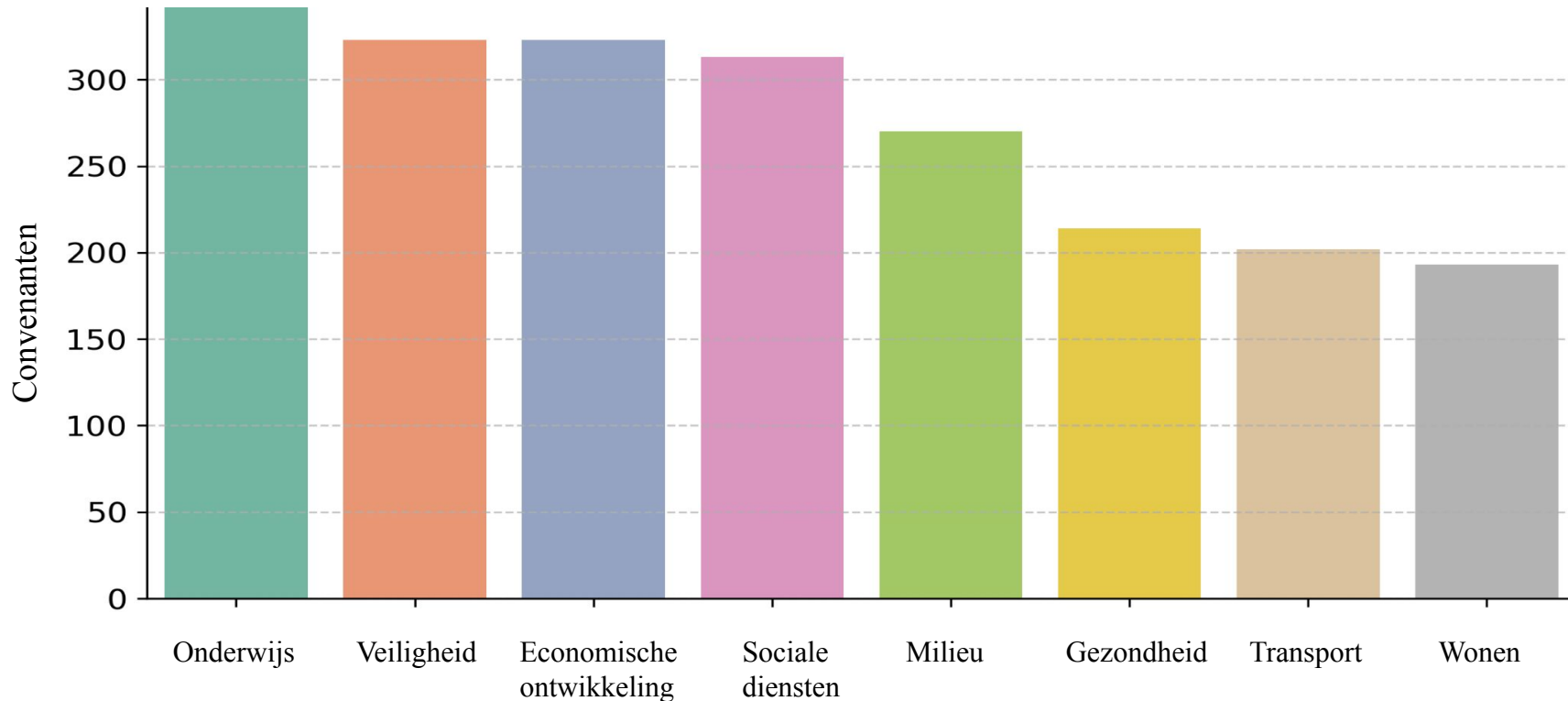
Extraheren van Metadata

Tot in hoeverre kunnen ontbrekende metadata uit de documenttekst worden gehaald?

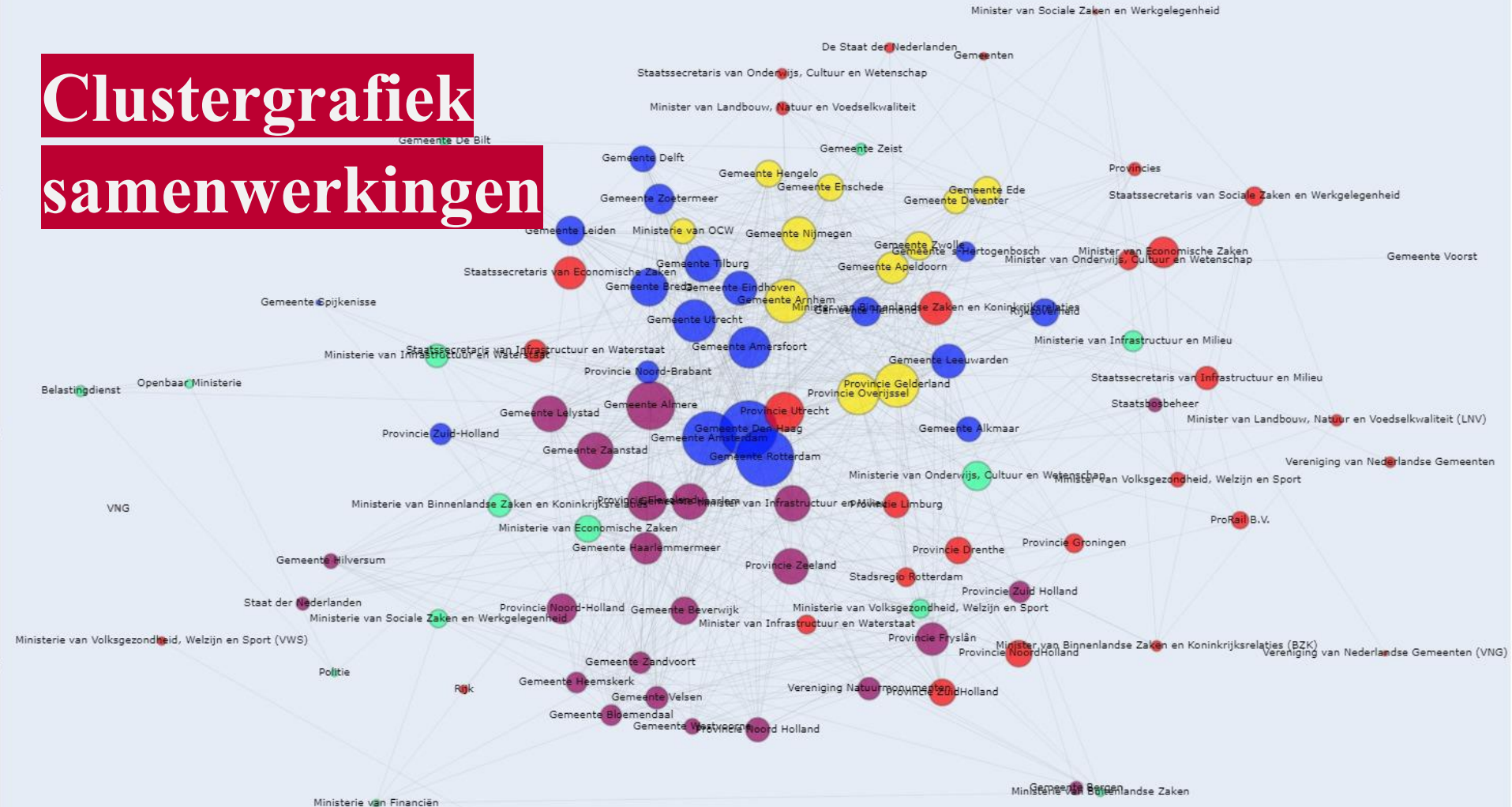
Waarde van het onderzoek

- Overheid controleren
- Test de capabiliteiten van GPT-3.5
 - potentieel voor het besparen van veel werk
- Dataset convenanten met metadata opent deuren voor verder onderzoek

Verdeling onderwerpen in convenanten



Clustergrafiek samenwerkingen





WOO-index

Locaties Woo-documenten

Informatiecategorie	Toelichting
Agenda's en besluitenlijsten bestuurscolleges 	De besluiten van Gedeputeerde Staten sinds 1 januari 2017, inclusief de toelichting en de bijlagen.
Agenda's en besluitenlijsten bestuurscolleges 	Agenda's en besluitenlijsten van Gedeputeerde Staten
Bereikbaarheidsgegevens 	
Beschikkingen 	Overzicht van alle (openbare) subsidies verstrekt door de provincie Zuid-Holland
Beschikkingen	
Bij vertegenwoordigende organen ingekomen stukken 	Bij Provinciale Staten ingekomen stukken
Jaarplannen en jaarverslagen 	Begroting, jaarrekening, voorjaarsnota & najaarsnota
Klachtoordelen 	Op dit moment publiceren wij het Jaarverslag klachtfunctionaris
Overige besluiten van algemene strekking	
Vergaderstukken decentrale overheden 	Vergaderstukken van Provinciale Staten
Wetten en algemeen verbindende voorschriften	
Woo-verzoeken en -besluiten 	

Convenanten ophalen

Bron		Convenanten	Titel	Onderwerp	Beschrijving	Publicatie datum	Partijen
Ministeries	Rijksoverheid	184					
	Belastingdienst	42					
Provinciën\	Notubiz	968					
Gemeenten	Bestuurlijke informatie	577					
Overig	Staatscourant	1119					
	Handmatig	38					
	Totaal	2928					

Metadata ophalen

Validatie

1) Beschrijving

2) Onderwerp

3) Ondertekeningsdatum

4) Startdatum

5) Betrokken partijen

6) Classificatie

You are a model that performs six tasks:

- | | |
|-----|---|
| (1) | <i>Provide a small two to three sentence description in Dutch about the full covenant.</i> |
| (2) | <i>Classify the topic of the covenant into one of the following categories: Environment, Housing, Transportation, Healthcare, Education, Economic development, Public Safety, Energy, Social Services, Agriculture, Technology and Innovation.</i> |
| (3) | <i>Extract the signing date of the covenant format dd-mm-yyyy. If it is unfindable, return None.</i> |
| (4) | <i>Extract the starting date of the covenant in the format dd-mm-yyyy. If it is unfindable, return None.</i> |
| (5) | <i>Extract all the involved organisations that have signed the covenant.</i> |
| (6) | <i>Classify whether this is a covenant or not (True or False). A covenant is an agreement by the government with one or more parties, aimed at achieving certain (policy) goals. A covenant includes written agreements on (the delivery of performance).</i> |

Your output should be a JSON object with six keys: 'description' (the short description), 'topic' (the classified topic), 'signingdate' (the signing date), 'startingdate' (the starting date), 'parties' (a list of involved parties), and 'covenant' (classification).

1) 0.86

2) 0.91

3) 0.09

4) 0.11

5) 9 uit 10 correct

1 op 4 gemist

6) 1 uit 16 correct

Conclusie

- Er kan geen concrete uitspraak worden gedaan over hoeveel convenanten er gepubliceerd zijn door een grote hoeveelheid die niet gestructureerd gepubliceerd zijn
- GPT-3.5 toont veel potentieel voor het verzamelen van informatie over partijen, onderwerpen, en beschrijvingen.
 - De limiterende factor zit in de maximum input

Aanbeveling

- Een grote stap naar het verbeteren van de vindbaarheid van convenanten is het classificeren als convenant en publicatie van metadata
 - Vooral gemeenten (via Notubiz en Bestuurlijke informatie) kunnen hier nog op verbeteren