

Extractie van metadata uit beschikkingen met AI



Datum	Ontvanger	Boete	Overtreden artikel	Type overtreding
04/12/2013	De heer [...]	20000	artikel 30h, eerste lid van de Wok	Het exploiteren van vijf, respectievelijk zeven kansspelautomaten op een voor het publiek toegankelijke plaats
23/03/2023	Imperial E-Club Limited	100000	artikel 1, eerste lid, aanhef en onder a, van de Wok	Online kansspelen aangeboden zonder vergunning op de Nederlandse markt
15/09/2020	Probe Investments Limited	15000	artikel 1, eerste lid, aanhef en onder a, van de Wok	Het geven van gelegenheid om mede te dingen naar prijzen of premies zonder de vereiste vergunning

MSc Scriptie Data Science
Project: CITaDOG
Harry Nan

Beschikkingen

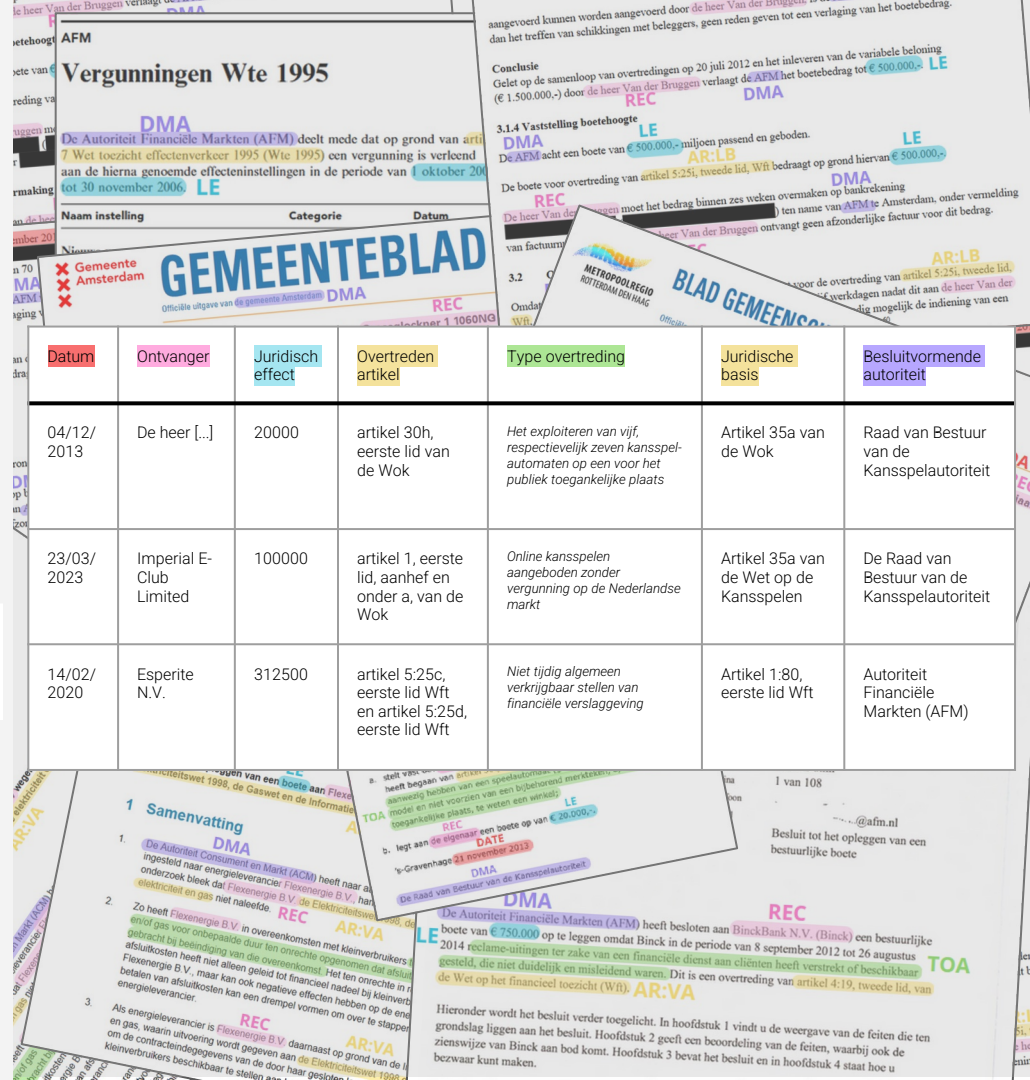
- Miljoenen beschikkingen
 - Honderden bestuursorganen
 - Vergunningen, sancties etc.
 - Kleine en grote gevallen



Zoekresultaten 4474072 resultaten

- Welke informatie kunnen we eruit halen?*
- Waar hebben we AI voor nodig?

* o.b.v. Woo 3.3a



Hoe is dit eerder gedaan?

- Automatische herkenning van patronen, structuren
- RegEx, NER, POS-tagging

24 Nov 2005
September 24, 2010
Jul 24, 2016
12/24/21

Patronen voor datum

Wet	Wet op de kansspelen, Artikel 35a
Subtype	Artikel
Instantie	Veiligheid en Justitie
Bron	Basis Wettenbestand
Datum	01-04-2021
	Toon meer >

Patronen voor wetsartikelen, uit LinkeXtractor

- Tekst moet uit herkenbare patronen bestaan

- Automatische extractie van informatie uit tekst, context bewuste extractie
- Large Language Models, zoals ChatGPT, voor verschillende taken (informatie extractie, samenvattingen etc.)



Informatie extractie m.b.v. ChatGPT via een (onvolledige) prompt

- Werkt slecht op lange stukken tekst



Hoe kunnen we effectief verschillende soorten informatie uit beschikkingen halen?

Subset: Sanctiebesluiten
Categorisatie van stukken informatie:

1. Homogene patronen, homogene context

- Bevatten duidelijke patronen
- Eenvoudig, weinig context nodig

Datum

2. Homogene patronen, heterogene context

- Bevatten duidelijke patronen
- Context is nodig voor juiste extractie

**Overtreden artikel,
Juridisch effect**

3. Heterogene patronen, heterogene context

- Hebben geen/weinig duidelijke patronen
- Context is nodig voor juiste extractie

**Ontvanger,
Type Overtreding (beschrijving)**

1. Homogene patronen en context

Datum

- Bevat duidelijke patronen

24 Nov 2005

September 24, 2010

Jul 24, 2016

12/24/21

- Uniform over alle beschikkingen
- Geen AI nodig voor goede extractie

model en niet voorzien van een bijbehorend merkteken, op een voor het publiek toegankelijke plaats, te weten een winkel;

b. legt aan de eigenaar een boete op van € 20.000,-.

Gravenhage 21 november 2013 DATE

CITY

De Raad van Bestuur van de Kansspelautoriteit

w.g.

J.J.H. Suyver

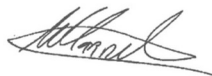
c) aan de vergunning de voorschriften, die zijn opgenomen in hoofdstuk 5, te verbinden.

4 Ondertekening

CITY

Delft 9 december 2016 DATE

namens Dijkgraaf en Hoogheemraden van Delfland,
de Teamleider Regulering en Planadvisering,



drs. W. van Harmelen

Datum in brief vorm, in combinatie met een stad

GEMEENTEBLAD Nr. 273515
Officiële uitgave van de gemeente Amsterdam 24 juni 2024 DATE

Besluit Bed & Breakfast vergunning Grossglockner 1 1060NG Amsterdam

Datum 2 december 2015 DATE

Ons kenmerk [REDACTED]

Pagina 59 van 70

evoerd door de heer Van der Bruggen, is de AFM van oordeel dat deze, anders

Datum 02 november 2019 DATE

Ons kenmerk 7971 / 00.019.052

1 Inleiding

1. De Kansspelautoriteit is belast met het toezicht op de naleving van het henealde

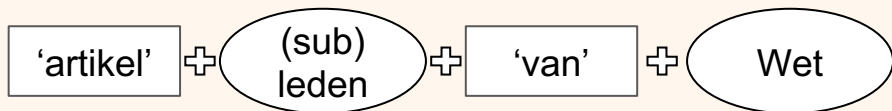
2 Gemeentebblad 2021 nr. 366504 21 oktober 2021 DATE

Datum in header of footer

2. Homogene patronen, heterogene context

Overtreden Artikel

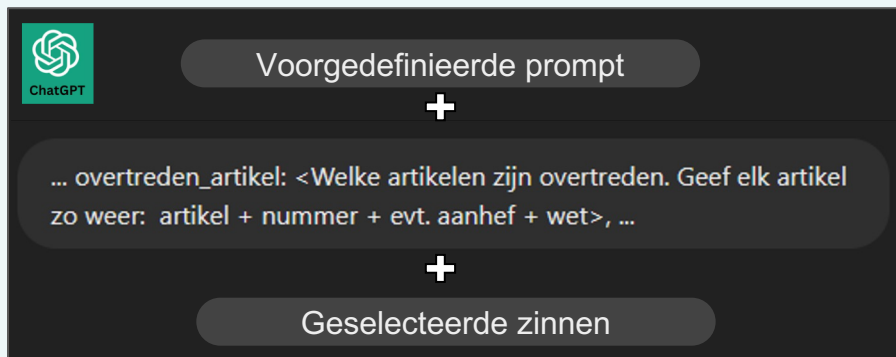
- Bevat duidelijke uniforme patronen



- Meer context is nodig om begrijpen of het artikel is overtreden
 - Gebruik van AI

De Kansspelautoriteit heeft vastgesteld dat Clipboard Publications B.V. reclame maakt voor kansspelen waarvoor geen vergunning op grond van de Wet op de Kansspelen (hierna: Wok) is verleend. Daarmee handelt Clipboard Publications B.V. in strijd met de Wok. De raad van bestuur van de Kansspelautoriteit (hierna: de raad van bestuur) draagt Clipboard Publications B.V. op het maken van reclame voor kansspelen waarvoor geen vergunning ingevolge de Wok is verleend (overtreding van artikel 1, eerste lid, onder b, van de Wok), te staken en gestaakt te houden door middel van het opleggen van een last onder dwangsom. context Overtreden artikel

- Extract alle zinnen in document waarin het patroon voor wetsartikelen voorkomen
- Vraag ChatGPT welk(e) artikel(en) word(en) overtreden:



3. Heterogene patronen en context

Type Overtreding (beschrijving)

- Geen duidelijke patronen
- Vaak besproken in context van Overtreden Artikel (vorige slide):

De Raad van Bestuur van de Kansspelautoriteit:

Overtreden Artikel

- a. stelt vast dat de eigenaar van Star SAT Electronica, de heer [...] een overtreding heeft begaan van artikel 30t, eerste lid, aanhef en onder c, van de Wok door het aanwezig hebben van een speelautomaat (gokzuil) van een niet toegelaten model en niet voorzien van een bijbehorend merkteken, op een voor het publiek toegankelijke plaats, te weten een winkel; **Type overtreding**

Type overtreding

De Autoriteit Financiële Markten (AFM) heeft besloten aan BinckBank N.V. (Binck) een bestuurlijke boete van € 750.000 op te leggen omdat Binck in de periode van 8 september 2012 tot 26 augustus 2014 reclame-uitingen ter zake van een financiële dienst aan cliënten heeft verstrekt of beschikbaar gesteld, die niet duidelijk en misleidend waren. Dit is een overtreding van artikel 4:19, tweede lid, van de Wet op het financieel toezicht (Wft).

Overtreden Artikel

Hieronder wordt het besluit verder toegelicht. In hoofdstuk 1 vindt u de weergave van de feiten die ten

- De informatie bevindt zich in de verkregen zinnen van de vorige slide.
- Vraag ChatGPT om het type overtreding te extraheren:



Voorgedefinieerde prompt



... 'type_of_activity': <Wat er is gebeurd waardoor de wet/het artikel is overtreden>...



Geselecteerde zinnen

Resultaten*

1. Homogene patronen, homogene context

	Precisie	Pakkans
Date	1.000	1.000

- Al niet nodig
- Generaliseerbaar over bestuursorganen / type beschikkingen

Datum

2. Homogene patronen, heterogene context

	Precisie	Pakkans
Overtr. Art.	1.000	0.950
Jur. Effect	0.975	1.000
Jur. Basis	0.611	0.611

- Effectief, maar patronen en prompt moeten correct zijn gedefinieerd

Overtreden Artikel, Juridisch Effect, Juridische Basis

3. Heterogene patronen, heterogene context

	Precisie	Pakkans
Type Over.	0.906	0.909
Ontvanger	0.900	0.917
Besl. Aut.	0.937	0.960

- Effectief, maar zin moet voldoende context bevatten en stuk informatie moet in zin staan

Type Overtreding, Ontvanger, Besluitvormende Autoriteit

Hoe nu verder?

- Toepassing op andere soorten beschikkingen
 - Aanpassingen van prompts, patronen herkenning etc.
- Van kwalitatieve analyse naar kwantitatieve analyse
- Volgend jaar: Verder met CITaDOG!



Datum	Ontvanger	Boete	Overtreden artikel	Type overtreding
04/12/2013	De heer [...]	20000	artikel 30h, eerste lid van de Wok	Het exploiteren van vijf, respectievelijk zeven kansspelautomaten op een voor het publiek toegankelijke plaats
23/03/2023	Imperial E-Club Limited	100000	artikel 1, eerste lid, aanhef en onder a, van de Wok	Online kansspelen aangeboden zonder vergunning op de Nederlandse markt
15/09/2020	Probe Investments Limited	15000	artikel 1, eerste lid, aanhef en onder a, van de Wok	Het geven van gelegenheid om mede te dingen naar prijzen of premies zonder de vereiste vergunning