

Note: The annotation guidelines are written in Dutch, since the administrative decisions are in Dutch.

## Annotatie Protocol voor Informatie Extractie uit Beschikkingen

### 1. Doel

Dit annotatie protocol richt zich op het systematisch extraheren van belangrijke juridische informatie uit bepaalde soorten beschikkingen, namelijk uit boetebesluiten en dwangsombesluiten. Dit protocol wordt gebruikt om machine geëxtraheerde informatie te analyseren en evalueren.

### 2. Documenten

De documenten waar informatie uit gehaald moet worden zijn twee soorten sanctiebesluiten van twee verschillende bestuursorganen. De documenten zijn publiekelijk beschikbaar op [woogle.wooverheid.nl](https://woogle.wooverheid.nl). De bestuursorganen die voor dit project gebruikt worden zijn:

- **Autoriteit Financiële Markten (AFM).** Dit bestuursorgaan handhaaft en houdt toezicht op de integriteit en transparantie van de financiële markten in Nederland.
- **Kansspelautoriteit (KSA).** Dit bestuursorgaan ziet toe op de naleving van wet- en regelgeving binnen de kanspelsector in Nederland.



De volgende soort sanctiebesluiten kun je verwachten:

- **Boetebesluit.** Dit is een besluit waarbij een bestuursorgaan als bestuurlijke sanctie een onvoorwaardelijke verplichting tot betaling van een geldsom oplegt aan een persoon of entiteit vanwege een overtreding van wettelijke voorschriften. Er kan ook besloten worden om een waarschuwing of een boete van 0 euro op te leggen.
- **Last onder dwangsom.** Dit is een besluit waarbij een bestuursorgaan als bestuurlijke sanctie een maatregel oplegt waarbij de overtreder de last krijgt om een overtreding binnen een gestelde termijn te beëindigen, bij gebreke waarvan hij een geldsom is verschuldigd.

### 3. Richtlijnen voor annotatie

De volgende richtlijnen voor annotatie moeten worden gevolgd:

- De informatie die wordt geëxtraheerd is direct gekopieerd vanuit het document en in één stuk aan elkaar te vinden.
  - o Stukken combineren uit verschillende stukken tekst kan dus niet.
  - o Spelfouten of dergelijke zitten dus ook in de extractie, als deze in het document voorkomen
  - o Bij een page break of dergelijke, moet de header, footer, paginanummer en dergelijke niet worden meegenomen.

- Voorbeeld van een **juiste** annotatie van Type Overtreding (en Overtreden Artikel):

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 **Type overtreding** 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. **Overtreden artikel**

- Voorbeeld van een **onjuiste** annotatie van Type Overtreding (en Overtreden Artikel):

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 **Type overtreding** 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. **Overtreden artikel**

- De geëxtraheerde informatie is opgehaald vanuit de juiste context.
  - Als de informatie op een ander stuk in het document vollediger is, maar niet in de context van de te annoteren informatie staat, kan deze niet worden gebruikt.
    - Voorbeeld: Als de wet van het overtreden artikel vollediger wordt weergegeven in een context waarbij niet duidelijk is dat het artikel wordt overtreden, kan deze niet worden gebruikt.
    - Voorbeeld: Het type overtreding wordt vollediger gemeld op een stuk waarbij niet duidelijk wordt vermeld dat deze handeling een overtreding is van de wet. De minder volledige versie in de juiste context wordt geëxtraheerd.

- Voorbeeld van een **juiste** annotatie van overtreden artikel, vanwege de aanwezigheid van de (juiste) context (in dit geval: in strijd met, overtreding). Hoewel het artikel van de wet 'Wok' eerder in het document vollediger vermeld is (Wet op de Kansspelen), wordt deze niet geëxtraheerd omdat de volledigheid niet in de juiste context is.

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**

- Elk uniek stuk informatie wordt eenmalig geëxtraheerd als stukken informatie meerdere malen in het document worden herhaald.

- Voorbeeld: Ondanks dat de unieke ontvanger 'Clipboard Publications B.V.' meerdere malen in het document wordt vermeld, wordt de ontvanger slechts eenmalig geëxtraheerd.

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.

**Extractie:**

**Ontvanger** Clipboard Publications B.V.

- Als er meerdere verschillende ontvangers, juridische effecten of artikelen worden benoemd, wordt elke unieke waarde geëxtraheerd.
  - o Wel: verschillende waardes, andere ontvangers etc.
  - o Niet: synoniemen, afkortingen etc.
- Het aantal boetes of dwangsommen dat wordt geëxtraheerd in een beschikking is altijd gelijk aan het aantal ontvangers.
  - o Voorbeeld: [boete 1, boete 2] , [ontvanger 1, ontvanger 2].

#### 4. Te Annoteren Informatie

De volgende informatie moet worden gelabeld en geëxtraheerd uit de beschikkingen, op basis van de richtlijnen gegeven in 3:

##### 1. Type sanctiebesluit

- o **Beschrijving:** Het type sanctiebesluit dat in het document aan de orde is. Dit is of 'Boetebesluit', of 'Dwangsom'.
- o **Annotatie:** Categoriseer het document als een 'Boetebesluit' of 'Dwangsom'. Zie sectie 2 voor extra informatie.

##### 2. Datum (Date)

- o **Beschrijving:** De datum waarop de beschikking is gegeven
- o **Annotatie:** Extraheer de datum waarop het besluit is genomen. Annoteer de datum in het formaat DD/MM/JJJJ.
  1. Voorbeeld: 25/01/2021
  2. Als de datum van het besluit ontbreekt, geef dit weer als 'UNKNOWN'.

##### 3. Ontvanger (Recipient)

- o **Beschrijving:** De persoon of entiteit die het sanctiebesluit ontvangt. Er kunnen meerdere ontvangers zijn.
- o **Annotatie:** Annoteer de naam van de ontvanger(s), zo volledig mogelijk.
  1. Als de ontvanger een persoon is, extract waar mogelijk ook de affiliatie of organisatie van deze persoon
    1. Voorbeeld: 'Jan Smit, eigenaar van FC Volendam'.
  2. Als de ontvanger een B.V. of bedrijf is, geef deze zo volledig mogelijk weer.
    1. Voorbeeld: 'Accountants Baat B.V.'
  3. Soms is de ontvanger geanonimiseerd. Extract dan de geanonimiseerde versie van de ontvanger
    1. Voorbeeld: 'De heer [...]'
  4. Als er synoniemen voor dezelfde ontvanger wordt gebruikt, extraheer dan enkel de meest volledige versie die in de juiste context te vinden is.
  5. Als er meerdere ontvangers zijn, gebruik dan een apart veld voor elke ontvanger in de volgende vorm: ['ontvanger 1', 'ontvanger 2'].

#### 4. Juridisch Effect (Legal Effect)

- **Beschrijving:** Het juridische effect (rechtsgevolg) van de beschikking, de juridische consequentie van het sanctiebesluit voor de ontvanger. Er kunnen meerdere juridische effecten zijn voor verschillende ontvangers.
- **Annotatie:** Extract het juridisch effect van de beschikking. Het juridisch effect verschilt voor een boetebesluit en een sanctiebesluit:
  1. *Boetebesluit.* Extraheer het getal dat wordt opgelegd als boete. (NB: op basis van onderdeel 1 is al duidelijk dat sprake is van een boete). Als er een waarschuwing wordt gegeven, of als er wordt besloten om geen boete op te leggen, geef dit dan weer als 'Waarschuwing', of '0'.
    1. Voorbeeld: 10000
  2. *Dwangsom.* Extraheer het getal dat wordt gegeven als de hoogte van de dwangsom, per eenheid die wordt gegeven tot een maximum. Geef dit weer in de volgende vorm: '*Hoogte per Eenheid tot Maximum*'. Deze stukken hoeven niet aan elkaar in het document voor te komen.
    1. Voorbeeld: 2500 per overtreding tot 25000

Hieronder staan de punten verder uitgewerkt:

2. *Hoogte:* Een getal van een bedrag dat moet worden bepaald per eenheid
  1. Voorbeeld: 2500
3. *Eenheid:* Een eenheid (zoals tijd, overtreding) waarbij de ontvanger de hoogte moet bepalen per strekking van de gegeven eenheid
  1. Voorbeeld: 'per dag', 'per overtreding'
4. *Maximum:* Het maximum aantal dat betaald moet worden als de overtreding niet wordt gestaakt.
  1. Voorbeeld: 25000

Als een onderdeel mist, vul deze dan niet in.

5. Voorbeeld: ['2500 per overtreding', '2500 tot 25000']

3. Als er meerdere juridische effecten zijn, geef dit dan weer in de volgende vorm: ['juridisch effect 1', 'juridisch effect 2']

#### 5. Overtreden Artikel (Violated Article)

- **Beschrijving:** Het artikel van de wet dat is overtreden. Er kunnen meerdere overtreden artikelen zijn.
- **Annotatie:** Annoteer het artikel op de volgende manier: 'artikel *Toevoeging van Wet*'.
  1. Voorbeeld van overtreden artikel: Artikel 33a van de Wet op de Kansspelen

Hieronder worden de schuingedrukte termen verder uitgelicht:

2. *Toevoeging*. Het nummer van het artikel waarnaar gerefereerd wordt. Dit is vaak een getal, soms gevolgd door een nummer of met een speciaal karakter (zoals :). Neem aanheffen etc. mee in de annotatie.
  1. Voorbeeld: 2:60, eerste lid, aanhef en onder c
3. *Wet*. Dit is de wet waar het artikel zich in bevind. Vaak wordt dit weergegeven na het woord 'van'.
  1. Voorbeeld: Wet op het financieel toezicht
4. Als er meerdere overtreden artikelen zijn, geef dit weer als:
 

[overtreden artikel 1', 'overtreden artikel 2']

  1. Als er gebruikt wordt gemaakt van 'Juncto' (in combinatie met), extract deze dan als één artikel
    - 1.

#### 6. **Besluitvormende Autoriteit (Decision Making Authority - DMA)**

- **Beschrijving:** De autoriteit die de boete of dwangsom oplegt.
- **Annotatie:** Annoteer de naam van de besluitvormende autoriteit of bestuursorgaan. Doe dit zo volledig mogelijk.
  1. Voorbeeld van besluitvormende autoriteit: raad van bestuur van de Kansspelautoriteit

#### 7. **Juridische Basis (Legal Basis)**

- **Beschrijving:** Het wetsartikel dat de besluitvormende autoriteit (DMA) de bevoegdheid geeft om het sanctiebesluit te nemen.
- **Annotatie:** Annoteer het artikel op eenzelfde manier als Overtreden Artikel (Violated Article)
  1. Voorbeeld van juridische basis: artikel 35a van de Wet op de Kansspelen

#### 8. **Type Overtreding (Type of Misconduct)**

- **Beschrijving:** Beschrijving van de handeling(en) die geleid hebben tot de overtreding van het artikel.
- **Annotatie:** Extraheer het stuk dat de feitelijke gedraging in detail beschrijft, wat er toe heeft geleid dat een wetsartikel is overtreden. Extraheer hiervoor de zin(nen) of gedeelte van de zin zo volledig mogelijk
  1. Voorbeeld van type overtredingen:
    1. Het aanbieden van gelegenheid tot gokken op sportwedstrijden op een niet toegelaten speelautomaat op een publiek toegankelijke plaats
    2. Het online aanbieden van kansspelen zonder vergunning
    3. Niet voldoen aan informatieverzoeken van de AFM

5. Tabel voor extractie

De volgende tabel wordt gebruikt om de geextraheerde informatie op te slaan:

Document	Type Sanctie	Datum	Ontvanger	Juridisch Effect	Overtreden Artikel	Besluitvormende Autoriteit	Juridische Basis	Type Overtreding
<a href="#">Doc 1</a>								
<a href="#">Doc 2</a>								
<a href="#">Doc 3</a>								
<a href="#">Doc 4</a>								
<a href="#">Doc 5</a>								
<a href="#">Doc 6</a>								
<a href="#">Doc 7</a>								
<a href="#">Doc 8</a>								
<a href="#">Doc 9</a>								
<a href="#">Doc 10</a>								