

Woogle en de toegankelijkheid van overheidsinformatie

Anne-Maartje Douque (AC IenO),
Maarten Marx (UvA)
VNG, Grip op Informatie
2023-04-17

Inhoud

1. Wat is dat, Woogle?
2. Hoe staat het met de vindbaarheid en toegankelijkheid van documenten vrijgegeven onder de Woo?

Samenvatting toegankelijkheid

Onze onderzoeksvraag:

Wat horen we, als we een openbaar gemaakt besluit op een Woo verzoek laten voorlezen door de computer?

- Van de meer dan 1 miljoen paginas op Woogle, hoor je op ruim 300.000 **helemaal niets**.
 - terwijl die pagina gewoon tekst bevat.
- Van de 3.400 besluiten genomen op een Woo verzoek door een ministerie (beschikbaar op overheid.nl) hoor je in 22% **helemaal niets**.
 - dit zijn gewoon brieven, waarschijnlijk gemaakt in Word (allemaal van na 2020)

Uitsplitsing naar bestuursorgaan

Toegankelijkheid

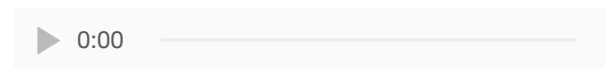
- gemiddeld is een op de 15 woorden een **niet bestaand woord**.

Dit zien U en ik

Geachte heer

In uw brief van 22 november 2019, gericht aan het ministerie van Landbouw, Natuur en Voedsel kwaliteit (hierna: LNV) heeft u met een beroep op de Wet openbaarheid van bestuur (hierna: Wob) informatie verzocht omdat u inzicht wil krijgen in de totstandkoming van het maatregelenpakket voor de stikstof- PFAS- problematiek van 13 november 2019.

Dit hoort iemand die het moet laten voorlezen:



Het brondocument

Wat is dat, Woogle?

Woogle is een onderzoeksproject van de Universiteit van Amsterdam

We wilden laten zien hoe PLOOI kan zijn als het gewoon het platform zou zijn zoals aangekondigd.

Heel erg vergelijkbaar met ORI (maar dan voor Woo dossiers)

De belofte van PLOOI

- Het is een **platform**, net als Marktplaats of Spotify,
- dat bestuursorganen en burgers bij elkaar brengt
- en het uitwisselen van stukken die vallen onder de Woo super makkelijk maakt.

Voor bestuursorganen

- "uploaden" van stukken gaat ongemerkt, net als vroeger bij ORI

Voor burgers

- alle stukken van alle bestuursorganen goed vindbaar en toegankelijk vanaf 1 plek
 - goede zoekfunctie, net als Google
 - startpagina (*verwijsindex*)

Toestand Woogle nu: functionaliteit

- 4 soorten Woo informatiecategorieën
 - 2d: agenda's/besluitenluisten
 - 2f: convenanten
 - 2i: Woo dossiers
 - 2k: beschikkingen
- 54 bestuursorganen
 - alle ministeries
 - veel provincies
 - flink wat gemeenten
 - van alles wat (UvA, NVWA, Kansspelautoriteit, Landelijke Politie,..)

Toestand Woogle nu

- Ongemerkt binnenstromen van data uit gemeenten via **openwebconcept** API van Common Grounds

- Zoekmachine van Google kwaliteit
- Elk bestuursorgaan, elke informatiecategorie, elk dossier, elk stuk, zelfs elke pagina, heeft een **unieke DOI, een digital object identifier**
 - de DOI **verwijst door**, als dat kan, naar de plek van het document **bij het bestuursorgaan**
 - Voorbeeld Hoeksche Waard: <https://doi.wooverheid.nl/?doi=nl.gm1963>

Toestand Woogle nu: stabiliteit

software

- Alles gratis, opensource, goed gedocumenteerd
 - MySQL, Python, Flask

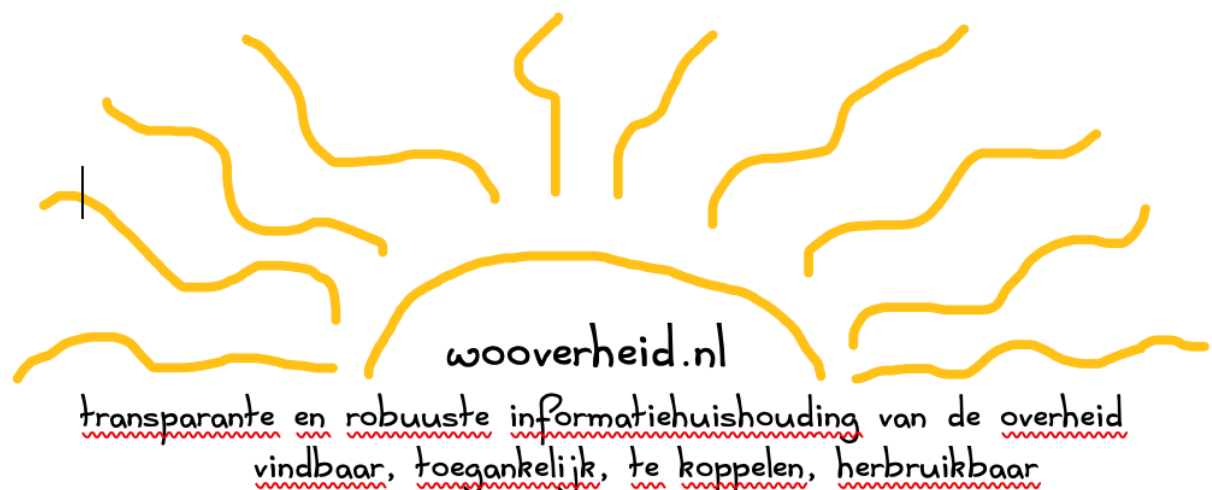
Financiering

- vergelijkbaar met ORI
 - erg lang erg onduidelijk, nu nog
 - veel eigen geld ingestopt
 - veel vrijwilligerswerk

Toekomst

- Wij hebben Woogle gratis aangeboden aan grote (markt)partijen
 - ~~nog geen positieve reactie~~
 - lijkt te gaan lukken!

Tsja



Nu het echte werk: onderzoek!

- Ik werk in een groep die onderzoek doet naar zoekmachines: [IRlab UvA](#).



Hoe maak je zoiets als Google?

Stel je wilt zoiets maken als PLOOI. Wat moet je doen?

1. De documenten (PDFs) die je vindbaar wilt maken allemaal ophalen.
 - Goed onthouden waar ze stonden (de URL ervan dus)
2. Dan alle woorden die daarin staan er allemaal uithalen.
 - Eigenlijk gewoon Control A, Control C, Control V
3. Dan maak je een index, net als achter in een boek
 - voor elk woord, onthou je in welk document, op welke bladzijde en op welke plek het staat



De zoekindex

- Voor elk woord in elke PDF, is er een lemma voor dat woord in de index
 - dat verwijst naar de (DOI code van de) PDF en de paginas in de PDF waar het voorkomt
- PFAS-problematiek: [nl.mnre1034.2i.2021.11.misc.1](https://doi.org/10.1034.2i.2021.11.misc.1) [1,3],
- hiermee kan men elk document terugvinden door **een woord dat in dat document staat** als zoekvraag te kiezen
 - gebruik meer termen om het document hoger in de lijst te krijgen.
- elke DOI code is gekoppeld aan de plek (de URL) waar het stuk staat.
- zo kan de zoekmachine dus verwijzen naar de juiste plek.
- Hups zo gevonden!

Klinkt makkelijk

- Maar waar denken jullie dat het misging?

1. In heel veel openbaargemaakte stukken **kan je geen control C doen**
2. Je kan de woorden er dus niet uithalen.
3. De tekst is niet **machine leesbaar**.

- En als het wel kan, staat het vaak vol met gekkigheid, zoals in

Geachte heer

In uw brief van 22 november 2019, gericht aan het ministerie van Landbouw, Natuur en Voedselkwaliteit (hierna: LNV) heeft u me aanbevolen om de openbaarheid van bestuur (hierna: Wob) informatie te zoeken omdat u inzicht wilt krijgen in de totstandkoming van het maatregelenpakket voor de stikstof- PFASproblematiek van 13 november 2019,

Met control C, control V gehaald uit [bron](#)

Waarom is dit zo?

- Dit weten we niet.
- Wel weten we dat het helemaal niet zo hoeft.
- En jullie weten vast wel welke gemeente ik nu als voorbeeld ga geven ;-)

Wat doet Woogie eraan?

- We proberen de stukken **toch vindbaar en toegankelijk te maken**.
- We **repareren de boel zo goed mogelijk** met de beste AI technieken.
 - Maar dat kan nooit foutloos
 - en kost erg veel tijd, rekenkracht en CO₂ uitstoot
- Maar veel liever willen we dat **het goed gaat aan de bron**.
- Daarom geef ik al die lezingen!

Mensen, doe het wel goed!

Twee grote problemen

1. Stukken zijn niet meer terug te vinden.
2. Stukken zijn niet, of niet goed, toegankelijk voor visueel gehandicapten.

Dat is toch hardstikke zonde!

- Jullie Wob-coördinatoren stoppen er zoveel tijd, geld en liefde in!
- Dit zal jullie gemeente toch niet overkomen?

Onze tip: kijk niet naar boven, kijk naar opzij

- daar zitten de best practices

Hartelijk dank!

Meer lezen?

- Zie posts op <https://wooverheid.nl>

Meedoen met Woogle?

- Zie [Woogle](#), Doe mee! tab.
- Liefst via openpub API. (Vraag Holger Peters).