Herkenning van weggelakte tekst in Wob-documenten: een onderzoek naar effectieve methodes

Roderick Majoor

Begeleider: M.J. Marx

Wob / Woo

- Wet openbaring van bestuur
- Regelt openbaring van informatie door overheid
- Vertrouwelijke en privacygevoelige informatie weggelakt

[redacted]

Van: [redacted]
Verzonden: maandag 30 juni 2014 16:40
Aan: [redacted] de
Onderwerp: RE: POSTZEGELPARK BUURT BIJEENKOMST

[redacted]

Beste [redacted]

In tegenstelling tot mijn eerder gedane toezegging om
aanwezig te zijn vanavond- laat ik je bij deze weten niet meer in de gelegenheid te zijn wegens een volle agenda; met optimale werkdruk.

Natuurlijk blijf ik belangstellend en ontvang met plezier de updates van het projekt.

Met vriendelijke groet, [redacted]

Van: [redacted]
Verzonden: Monday, June 23, 2014 12:18 PM
Aan: [redacted]
Onderwerp: FW: gebeidsagenda's voor netwerken

Beste buurtbewoner van Oud West,

Belangrijke informatie over uw buurt!

Wat gaat er het komende jaar in uw buurt gebeuren?

Lees het na in de gebiedsagenda

www.west.amsterdam.nl/gebiedsagenda.

Uw straat die opnieuw wordt ingericht, een nieuwe speelplek voor kinderen, of extra inzet op veiligheid?
Wilt u weten wat stadsdeel West het komende jaar in uw buurt gaat doen? U kunt het nu nalezen in de gebiedsagenda. Kijk hier voor meer informatie.

Met vriendelijke groet,

[redacted]
Buurtcoördinator Cremer-, Helmers- & Vondelparkbuurt
Afdeling Wijken, Stadsdeel West
M [redacted]
E [redacted]

Bezoekadres: Bos en Lommerplein 250, 1055 EK
Postadres: Postbus 57239, 1040 BC Amsterdam
Werkdagen: maandag, dinsdag, woensdag en vrijdag

1

Traag en te weinig

Ministeries overtreden op grote schaal eigen regels bij vrijgeven documenten

Door **Koen de Regt & Roland Strijker**

16 januari 2021 15:55 • Aangepast 16 januari 2021 15:55

nieuwsuur

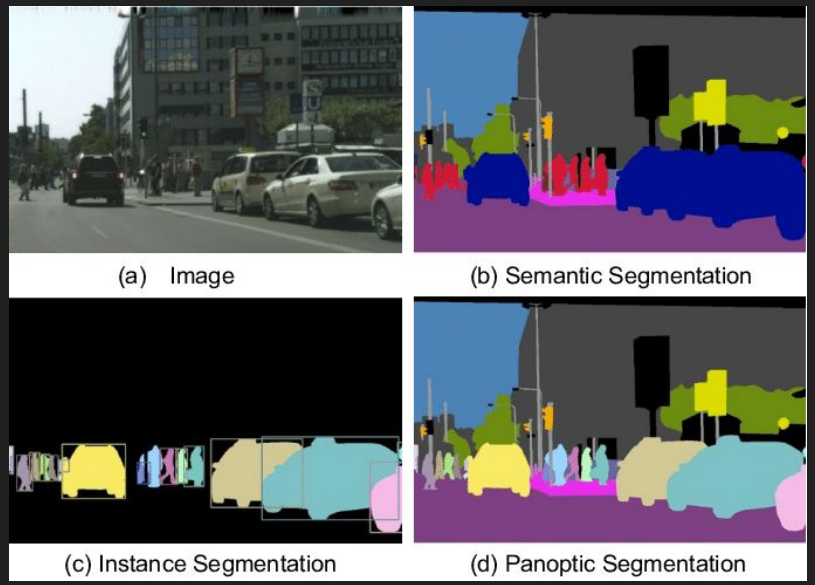
Donderdag 19 januari, 22:22

Adviescollege: 'Lakken is binnen de overheid topsport geworden, dat moet stoppen'

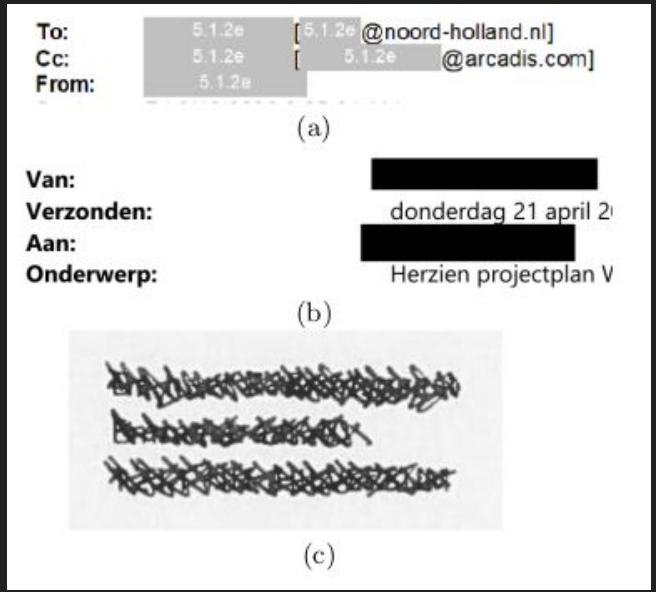
Doel

- Herkennen van weggelakte stukken in documenten
- Locatie van weggelakte stukken bepalen
- Evaluatie over hoeveelheid lak

Segmentatie

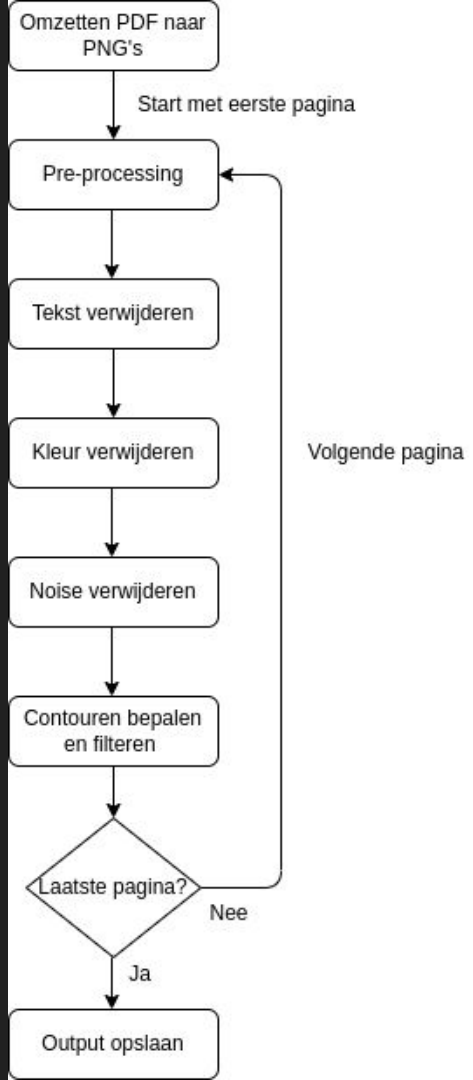


Datasets & Ground Truth



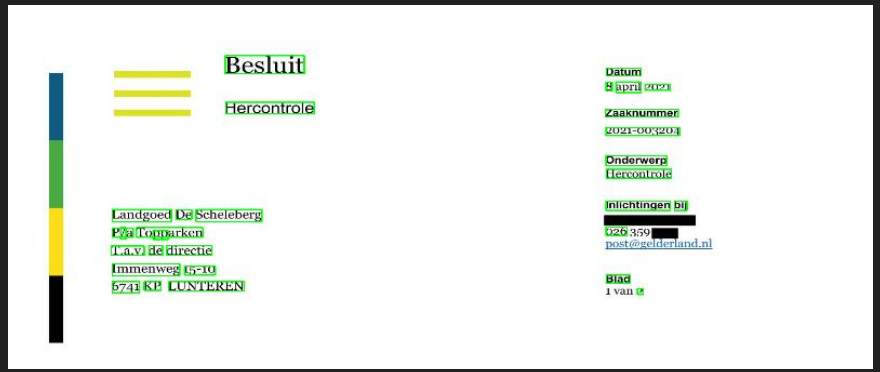
Methode

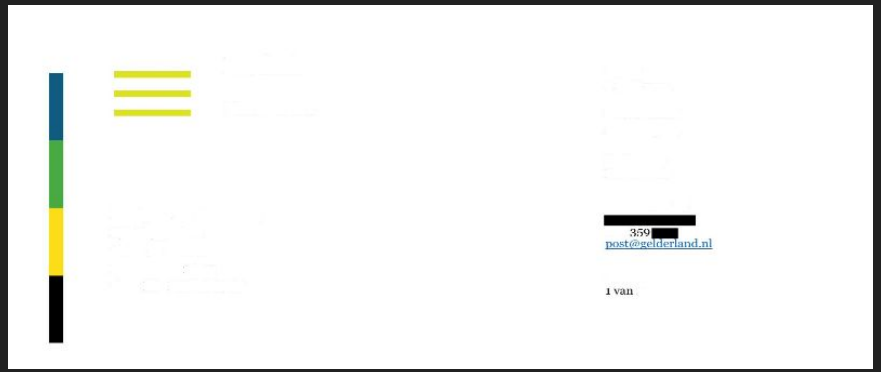
- Regel-gebaseerd vs. Machine Learning
- Tesseract & OpenCV
- Mask R-CNN



Regel-gebaseerd Implementatie (1/3)

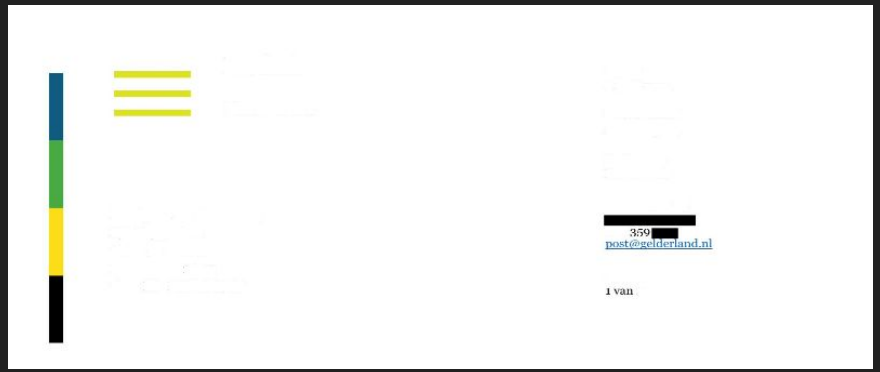
- Tekst bepalen

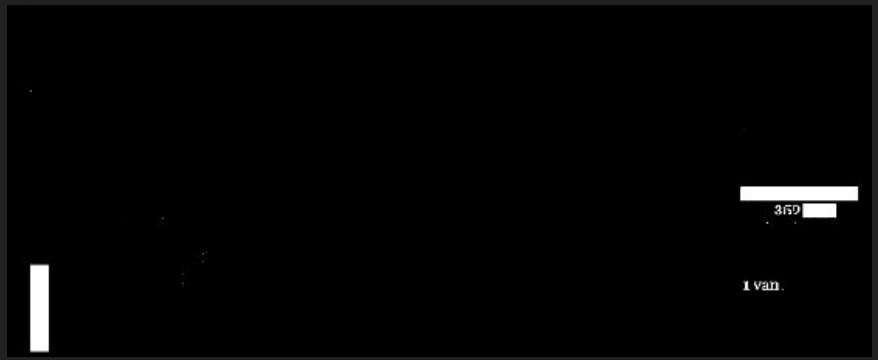




Regel-gebaseerd Implementatie (2/3)

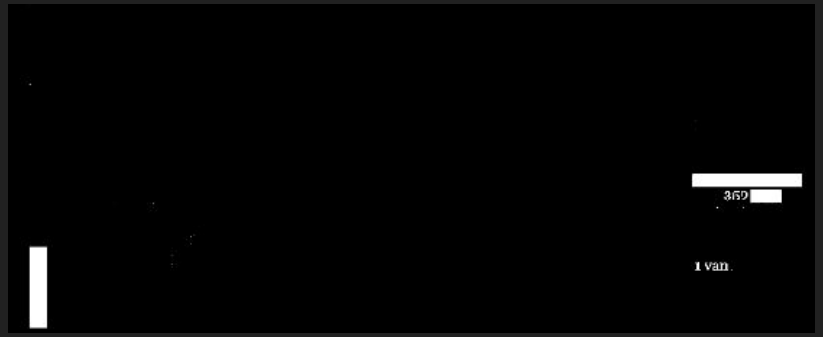
- Kleur bepalen

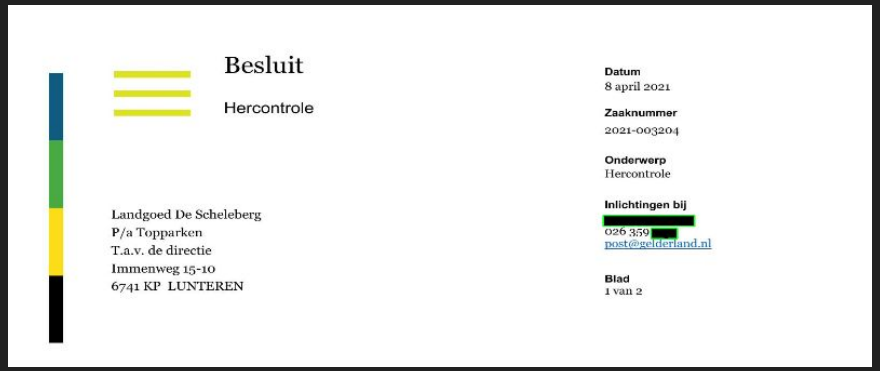




Regel-gebaseerd Implementatie (3/3)

- Verwijderen van noise en contouren selecteren





Besluit

Hercontrole

Landgoed De Scheleberg
P/a Topparken
T.a.v. de directie
Immenweg 15-10
6741 KP LUNTEREN

Datum
8 april 2021

Zaaknummer
2021-003204

Onderwerp
Hercontrole

Inlichtingen bij
026 359
post@gelderland.nl

Blad
1 van 2

Mask R-CNN

(art 67 AWR). Ze mogen daarbij niet meer gegevens verstrekken dan de begroting aangeeft en de toelichting bij de verordening waarin wordt aangegeven wat de beoogde opbrengst is. Een wat formeel juridisch antwoord maar meer kan ik er niet van maken. We kunn je wel definitieve bedragen mbt tot de opbrengst VRM halen uit de jaarrekening 2018. Ik heb nagevraagd wanneer deze gepubliceerd wordt. Dit laat ik je nog weten."

Uit de inmiddels gepubliceerde Jaarrekening 2018 en evenmin uit de Voorjaarsnota 2019 valt op geen enkele manier op te maken welk bedrag aan VMR-land in 2018 is afgedragen, noch worden hiermee de andere vragen beantwoord.

Met uitzondering van vraag twee zijn de vragen anoniem en generiek, waarmee Artikel 67 van de AWR niet toepassing is. Tevens worden in het antwoord dat wij op 26 maart 2018 van het College ontvingen, enkele bedrijven met naam genoemd. Vandaar dat wij hetzelfde ook in het kader van het onderhavige verzoek mogen verwachten.

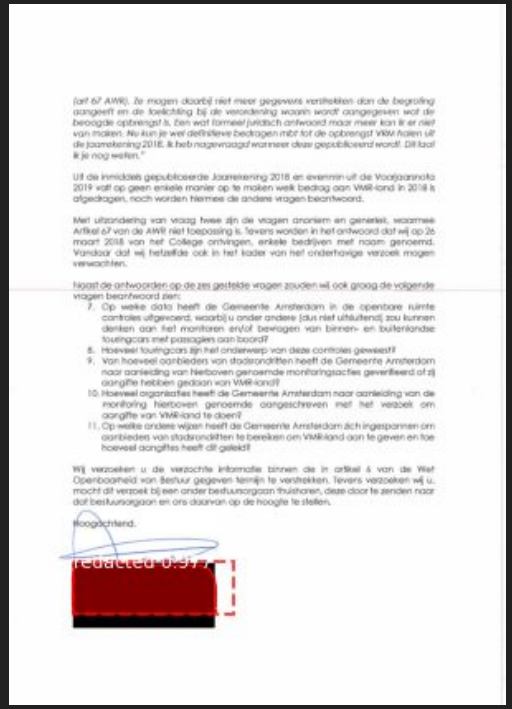
Naast de antwoorden op de zes gestelde vragen zouden wij ook graag de volgende vragen beantwoord zien:

7. Op welke data heeft de Gemeente Amsterdam in de openbare ruimte controles uitgevoerd, waarbij u onder andere (dus niet uitsluitend) zou kunnen denken aan het monitoren en/of bevragen van binnen- en buitenlandse touringcars met passagiers aan boord?
8. Hoeveel touringcars zijn het onderwerp van deze controles geweest?
9. Van hoeveel aanbieders van stadsrondritten heeft de Gemeente Amsterdam naar aanleiding van hierboven genoemde monitoringsacties geverifieerd of zij aangifte hebben gedaan van VMR-land?
10. Hoeveel organisaties heeft de Gemeente Amsterdam naar aanleiding van de monitoring hierboven genoemde aangeschreven met het verzoek om aangifte van VMR-land te doen?
11. Op welke andere wijzen heeft de Gemeente Amsterdam zich ingespannen om aanbieders van stadsrondritten te bereiken om VMR-land aan te geven en hoe hoeveel aangiftes heeft dit gelekt?

Wij verzoeken u de verzochte informatie binnen de in artikel 6 van de Wet Openbaarheid van Bestuur gegeven termijn te verstrekken. Tevens verzoeken wij u, mocht dit verzoek bij een ander bestuursorgaan thuishoren, deze door te zenden naar dat bestuursorgaan en ons daarvan op de hoogte te stellen.

Hoogachtend,

[redacted]



(art 67 AWR). Ze mogen daarbij niet meer gegevens verstrekken dan de begroting aangeeft en de toelichting bij de verordening waarin wordt aangegeven wat de beoogde opbrengst is. Een wat formeel juridisch antwoord maar meer kan ik er niet van maken. We kunn je wel definitieve bedragen mbt tot de opbrengst VRM halen uit de jaarrekening 2018. Ik heb nagevraagd wanneer deze gepubliceerd wordt. Dit laat ik je nog weten."

Uit de inmiddels gepubliceerde Jaarrekening 2018 en evenmin uit de Voorjaarsnota 2019 valt op geen enkele manier op te maken welk bedrag aan VMR-land in 2018 is afgedragen, noch worden hiermee de andere vragen beantwoord.

Met uitzondering van vraag twee zijn de vragen anoniem en generiek, waarmee Artikel 67 van de AWR niet toepassing is. Tevens worden in het antwoord dat wij op 26 maart 2018 van het College ontvingen, enkele bedrijven met naam genoemd. Vandaar dat wij hetzelfde ook in het kader van het onderhavige verzoek mogen verwachten.

Naast de antwoorden op de zes gestelde vragen zouden wij ook graag de volgende vragen beantwoord zien:

7. Op welke data heeft de Gemeente Amsterdam in de openbare ruimte controles uitgevoerd, waarbij u onder andere (dus niet uitsluitend) zou kunnen denken aan het monitoren en/of bevragen van binnen- en buitenlandse touringcars met passagiers aan boord?
8. Hoeveel touringcars zijn het onderwerp van deze controles geweest?
9. Van hoeveel aanbieders van stadsrondritten heeft de Gemeente Amsterdam naar aanleiding van hierboven genoemde monitoringsacties geverifieerd of zij aangifte hebben gedaan van VMR-land?
10. Hoeveel organisaties heeft de Gemeente Amsterdam naar aanleiding van de monitoring hierboven genoemde aangeschreven met het verzoek om aangifte van VMR-land te doen?
11. Op welke andere wijzen heeft de Gemeente Amsterdam zich ingespannen om aanbieders van stadsrondritten te bereiken om VMR-land aan te geven en hoe hoeveel aangiftes heeft dit gelekt?

Wij verzoeken u de verzochte informatie binnen de in artikel 6 van de Wet Openbaarheid van Bestuur gegeven termijn te verstrekken. Tevens verzoeken wij u, mocht dit verzoek bij een ander bestuursorgaan thuishoren, deze door te zenden naar dat bestuursorgaan en ons daarvan op de hoogte te stellen.

Hoogachtend.

redacted 0.91

PQ score

- Panoptic Quality

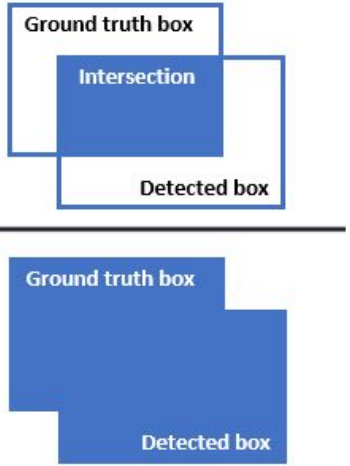
$$\mathrm{PQ} = \underbrace{\frac{\sum_{(p,g)\in TP} \mathrm{IoU}(p,g)}{|TP|}}_{\text{segmentation quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{recognition quality (RQ)}}$$

TP: true positive

FP: false positive

FN: false negative

$$\mathrm{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} =$$



RMSE

- Root Mean Square Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{N}}$$

- Testen kwaliteit van de evaluatiematen van ons programma

Evaluatie

1. Aantal weggelakte stukken
2. Aantal weggelakte karakters
3. Weggelakte tekst in tekstvlak (%)
4. Weggelakte tekst in karakteroppervlakte (%)

Evaluatiemaat	RMSE
1	2.24
2	22.94
3	0.71
4	2.20


ondertekende papieren brief behoef in te dienen.
De beslistermijn is verstreken, u bent te laat met beslissen, graag nu de beslissing tot verlening van de gevraagde inzage.

Inmiddels hebben we een paar keer gewisseld over de strekking van het verzoek, natuurlijk geef ik nadere toelichting nog steeds graag,
ook in de zin dat u concrete stukken noemt met de vraag of ik die wel of niet wil inzien, maar zonder uw tegenbericht ga ik er van uit dat we over & weer nu duidelijkheid hebben.

mtvrgrt
[redacted]


ondertekende papieren brief behoef in te dienen.
De beslistermijn is verstreken, u bent te laat met beslissen, graag nu de beslissing tot verlening van de gevraagde inzage.

Inmiddels hebben we een paar keer gewisseld over de strekking van het verzoek, natuurlijk geef ik nadere toelichting nog steeds graag,
ook in de zin dat u concrete stukken noemt met de vraag of ik die wel of niet wil inzien, maar zonder uw tegenbericht ga ik er van uit dat we over & weer nu duidelijkheid hebben.

mtvrgrt
[redacted]

Resultaten (1/2)

- Regel-gebaseerd vs. Machine Learning
- Getest op dezelfde pagina's en op dezelfde CPU

Methode	Regel-gebaseerd	Machine Learning
Gemiddelde SQ Score	0.93	0.91
Gemiddelde RQ Score	0.95	0.88
Gemiddelde PQ Score	0.88	0.80
Gemiddelde Uitvoertijd per Pagina	1.92s	6.68s

Resultaten (2/2)

- Regel-gebaseerde methode op verschillende datasets

Dataset	Zylab (N=46)	Corpus1 (N=55)	Corpus2 (N=45)	Totaal (N=146)
Gemiddelde SQ Score	0.92	0.88	0.95	0.92
Gemiddelde RQ Score	0.81	0.89	0.95	0.89
Gemiddelde PQ Score	0.74	0.82	0.90	0.83

Discussie

- Beide methodes goed op zwarte lak
- Aangrenzende stukken

- Regel-gebaseerd effectiever dan Machine Learning

Website

Vragen?