
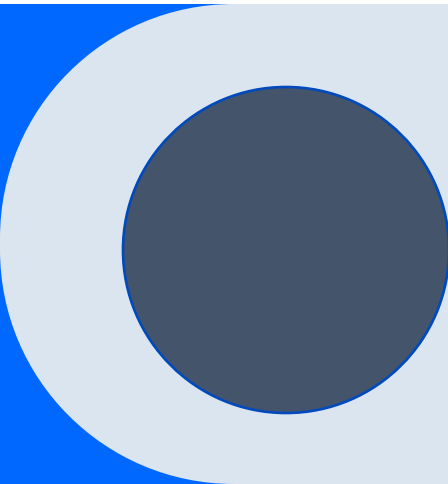




Maak slecht toegankelijke PDFs beter toegankelijk

Afstudeerproject bachelor Informatica



Lars Reddering, 12398675

Begeleider: Maarten Marx

Het probleem

- 300.000 mensen in Nederland met een visuele beperking
- Digitale toegankelijkheid voor iedereen, vastgelegd in wetgeving
- WCAG 2.1 richtlijnen
- Wet openbare overheid (woo)
- Veel documenten voldoen niet aan richtlijnen

Achtergrond

Portable Document Format (PDF)

- Verschillende PDF-standaarden
- PDF/Universal Accessibility (PDF/UA)

Web Content Accessibility Guidelines (WCAG)

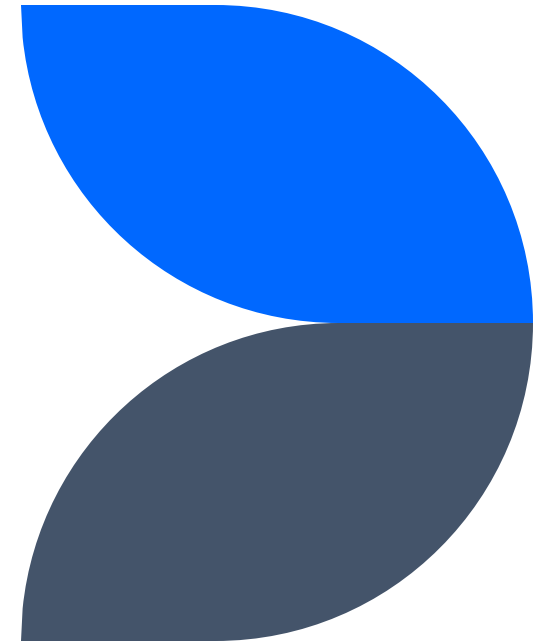
- Regels die bepalen of een document voldoet aan eisen
- De overheid heeft zich gebonden aan de WCAG 2.1 regels

Uitdagingen op dit moment

- Altijd menselijk controle
- Verschillende doelen
- Veel regels
- Onkunde (ctrl + p)

Onderzoeksvraag

Wat zijn de technische uitdagingen bij het automatisch oplossen van toegankelijkheidsproblemen in PDF-documenten en hoe kunnen deze problemen worden overwonnen?



Deelvragen

Onderzoeksvraag 1

Welke WCAG-toegankelijkheidsproblemen spelen er in de Nederlandse Woo dossiers, hoe vaak komen ze voor en welke daarvan lenen zich voor een geautomatiseerde oplossing?

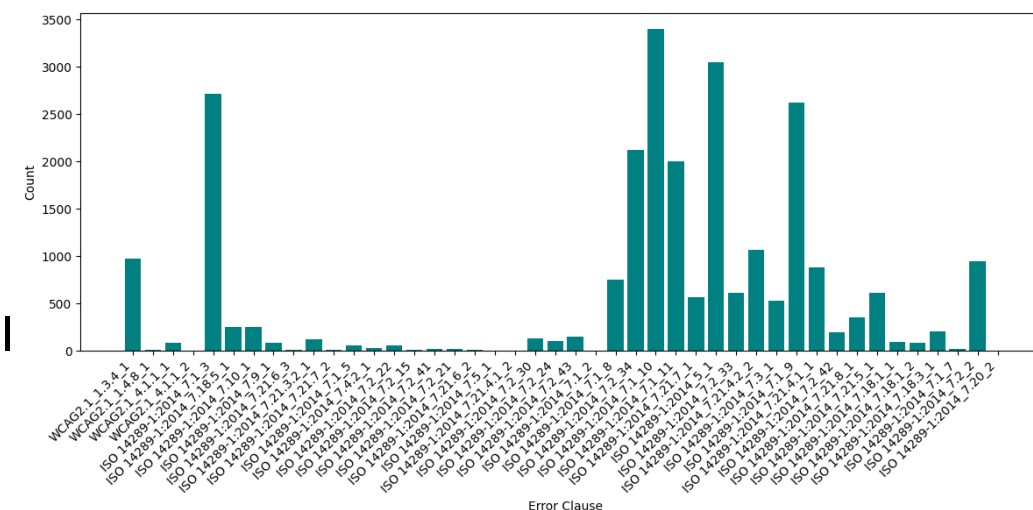
Onderzoeksvraag 2

Hoe kunnen deze fouten automatisch worden opgelost zodat het document beter toegankelijk is?



Toegankelijkheid meten

- Gebruik maken van VeraPDF
- Controle op alle WCAG richtlijnen
- Dataset van 6.768 PDF bestanden
- Ernst van het probleem via schaal PDFchecker.nl
- Critical, serious, error, warning, ignore



Figuur 1: Aantal gevonden fouten per WCAG regel

Een woo besluit

0%

 Nederlands

Het document bevat kritieke toegankelijkheidsproblemen die hieronder worden toegelicht. Los deze eerst op en check het document dan nogmaals.


Het document is niet gecodeerd

Als het document geen code heeft, kunnen screenreaders het niet in de juiste volgorde voorlezen. Dit is een kritieke fout. Je kan dit waarschijnlijk oplossen door je document nogmaals uit je kantoorapplicatie (Word of LibreOffice) te exporteren als 'tagged' (gelabelde) PDF. In Word doe je dit door in het menu 'Bestand - Opslaan als - (kies PDF) - Opties' de opties 'Labels voor documentstructuur voor toegankelijkheid' en 'Voldoet aan ISO 19005-1 (PDF/A)' aan te vinken. Gebruik NOOIT 'afdrukken - Adobe PDF' om je document naar PDF te exporteren.


Toon alle toegankelijkheidsproblemen


Waarschuwingen verbergen


Metadata (1) ^

 Het document mist essentiële metadata v

Code (3) ^

 Een lettertype heeft geen eenduidige letterbreedte. Dit betreft lettertype(n): Verdana. v

 Een gebruikt lettertype is niet in het document ingesloten. Dit betreft lettertype(n): ArialMT. v

 Het document bevat inhoud die niet gelabeld is in de code v

Figuur 2: WCAG 2.1 fouten in het besluit

Ministerie van Economische Zaken
en Klimaat

TER BESLUITVORMING

Aan de Staatssecretaris Mijnbouw
Cc aan de Minister voor Klimaat en Energie

**Programma DG Groningen en
Ondergrond**
Directie Transitie Diepe
Ondergrond
Auteur
[redacted]
[redacted]

Datum
6 december 2023
Kenmerk
PDGGO-DTDO / 41366814

Kopie aan
Bijlage(n)
1

nota

Lithiumwinning bij geothermie, haalbaarheidsonderzoek
n.a.v. motie Kröger

Parafenroute

[redacted] [redacted] [redacted]

[redacted] [redacted] [redacted]

[redacted] [redacted] [redacted]

Aanleiding

Op 17 februari 2022 is een motie aangenomen van Kröger (GroenLinks) c.s. (kamerstuk 35 531, nr. 28) waarin u verzocht wordt om de haalbaarheid van lithiumwinning bij geothermie te onderzoeken. EBN is gevraagd om onderzoek te doen naar dit onderwerp. Op 1 december 2023 zijn de resultaten van dat onderzoek (in de vorm van een rapport) gedeeld met het Ministerie van Economische Zaken en Klimaat.

Geadviseerd besluit

Met bijgevoegde Kamerbrief informeert u de Kamer over de uitkomsten van het onderzoek. U kunt de Kamerbrief ondertekenen.

Kernpunten

- Het rapport geeft aan dat lithiumwinning op dit moment nog niet rendabel is bij Nederlandse geothermieprojecten. Dat heeft te maken met de vastgestelde concentraties en de staat van de technologie. Te zeggen, tot nog toe aangetroffen concentraties zijn met de huidige staat van de technologie te laag voor rendabele extractie van lithium.
- Een bredere blik op het veld toont wel dat de ontwikkelingen snel gaan en dat het in de toekomst wel mogelijk kan zijn om ook bij de tot nog toe lage concentraties een rendabele extractie van lithium te realiseren. Uiteraard is het ook mogelijk dat hogere concentraties aangetroffen zullen worden.
- Verder bevat geothermiewater naast lithium een breder scala aan materialen die kritisch zijn voor de energietransitie. Dat is in het huidige rapport niet meegenomen, maar het is de moeite waard om daar in de toekomst dieper in te duiken. Dit zou een vervolgvraag aan EBN kunnen zijn.

Ontvangen EBR

Toelichting

Pagina 1 van 2

Figuur 3: Een woo besluitsnota

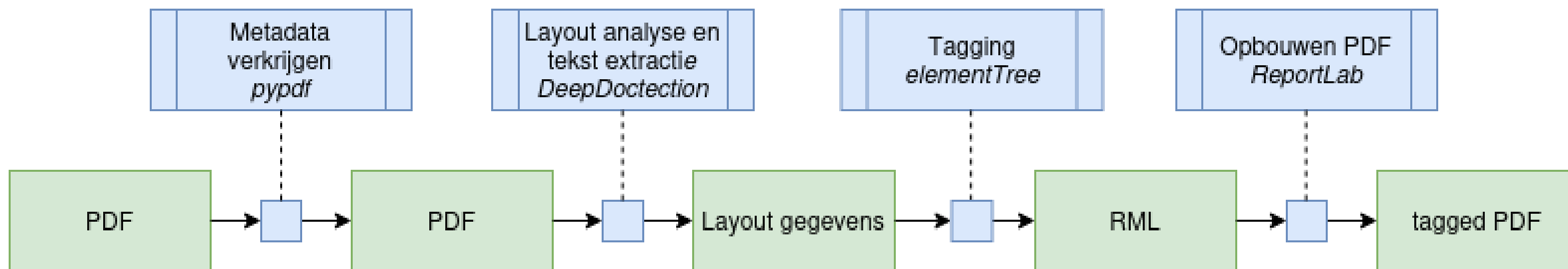
Resultaten

Specificatie	Regel	Check	Status	Aantal fouten	Beschrijving
ISO 14289-1	7.1	10	negeer	3396 (13.4%)	Require title in the title bar.
ISO 14289-1	5.1	1	negeer	3052 (12.1%)	No PDF/UA declaration.
ISO 14289-1	7.1	3	serieus	2714 (10.7%)	Content is not tagged.
ISO 14289-1	7.1	9	error	2622 (10.4%)	Title not in metadata.
ISO 14289-1	7.2	34	error	2122 (8.3%)	Language not in metadata.
ISO 14289-1	7.21.4.2	2	error	1069 (4.2%)	Not all glyphs are described.
WCAG 2.1	1.3.4	1	negeer	978 (3.9%)	Semantics expressed by formatting text.
ISO 14289-1	7.2	2	error	951 (3.8%)	Caption of a table is not tagged correctly
ISO 14289-1	7.21.4.1	1	error	883 (3.5%)	Font program not embedded.
ISO 14289-1	7.1	8	serieus	757 (3.0%)	No metadata stream

Tabel 1: Meest gevonden fouten in dataset 1

- Metadata ontbreekt vaak (regel 7.x)
- Tagging ontbreekt (regel 5.1)
- Metadata kan op zekere hoogte makkelijk
- Omzetten naar een getagde PDF lastig

Implementatie converter



Figuur 4: Schematisch overzicht van de converter

Implementatie Deepdoctection

Titel



Aanleiding

Indiening van een gezamenlijke verklaring bij het Internationaal Gerechtshof (IGH) die strekt tot interventie in de procedure tussen Gambia en Myanmar vanwege genocide op de Rohingya.

Paragraaf



Geadviseerd besluit

Uw akkoord met verzending van bijgevoegde Kamerbrief alsmede datering en ondertekening van de brief indien U daarmee akkoord bent.

Lijst



Kernpunten

In de motie Koopmans c.s. van 1 juli 2019 werd uw voorganger verzocht de mogelijkheden te onderzoeken Myanmar aan te spreken op schendingen van het Genocideverdrag bij het Internationaal Gerechtshof. Gambia is kort daarna een procedure gestart tegen Myanmar voor het IGH op grond van het Genocideverdrag. Een zelfstandige zaak van NL, naast een procedure van Gambia, zou geen meerwaarde hebben en de lopende procedure compliceren (met vertraging als gevolg). Daarom heeft Nederland de mogelijkheid tot interventie in de zaak van Gambia onderzocht. Vervolgens hebben NL en CAN besloten gezamenlijk te interveniëren in de procedure. Met het indienen van de interventie is echter gewacht totdat het IGH zich bevoegd had verklaard, en Gambia en Myanmar hun initiële schriftelijk verweerschrift ingediend hadden. CAN en NL zagen geen reden voor een interventie over het vraagstuk van de bevoegdheid omdat de jurisprudentie van het Hof op dat punt voldoende uitgekristalliseerd was (hetgeen werd bevestigd door het Hof in zijn uitspraak daarover). Een bijdrage aan de uitleg van de materiële bepalingen van het Genocideverdrag had volgens CAN en NL daarentegen wel toegevoegde waarde, in het bijzonder met betrekking tot de vaststelling van genocide opzet, seksueel en gender-gerelateerd geweld tegen vrouwen en meisjes, de speciale status van kinderen onder het Genocideverdrag, en de verplichting tot het waarborgen van een eerlijk proces in de tenuitvoerlegging van de verplichting tot vervolging en berechting van verdachten van genocide. De interventie is dan ook toegespitst op deze onderwerpen.

Pagina 1 van 2

00001

1424565

Implementatie ReportLab

```

1  <!DOCTYPE document SYSTEM "rml_1_0.dtd">
2  <document filename="41335180_fixed.pdf" compression="0" invariant="1" tagged="1">
3    <<<docinit>
4      <<<<registerTTFont faceName="Helvetica" fileName="font/Helvetica.ttf" />
5      <<<<registerTTFont faceName="Helvetica-Bold" fileName="font/HelveticaBold.ttf" />
6    <<</docinit>
7  > <<<<template pageSize="a4" title="41335180" subject="Lithiumwinning bij geothermie, haalbaarheidsonderzoek">
50 <<<</template>
51 <<<<stylesheet>
52 <<<<<paraStyle name="h1" fontName="Helvetica-Bold" fontSize="12" />
53 <<<<<paraStyle name="normal" fontName="Helvetica" fontSize="10" />
54 <<<<<listStyle name="tagged_list" bulletFontName="Helvetica" bulletFontSize="10" />
55 <<<</stylesheet>
56 <<<<story>
57 <<<<<keepInFrame id="p0ff7" frame="p0f7">
58 <<<<<<para style="h1" tagType="H1">TER BESLUITVORMING</para>
59 <<<<</keepInFrame>
60 <<<<<keepInFrame onOverflow="shrink" id="p0ff5" frame="p0f5">
61 <<<<<<para style="normal" tagType="P">Aan de Staatssecretaris Mijnbouw Co aan de Minister
62 <<<<<<<para style="normal" tagType="P">voor Klimaat en Energie</para>
63 <<<<</keepInFrame>
64 <<<<<keepInFrame id="p0ff10" frame="p0f10">
65 <<<<<<para style="h1" tagType="H1">nota</para>
66 <<<<</keepInFrame>
67 <<<<<keepInFrame onOverflow="shrink" id="p0ff6" frame="p0f6">
68 <<<<<<para style="normal" tagType="P">Lithiumwinning bij geothermie, haalbaarheidsonderzoek
69 <<<<<<<para style="normal" tagType="P">n.a.v. motie Kr&#246;ger</para>
70 <<<<</keepInFrame>

```

Titel

Paragraaf

Figuur 6: ReportLab Markup Language (RML)

Resultaten

	Origineel VeraPDF	Nieuw VeraPDF	Nieuw PAC
Totaal aantal documenten	15	15	15
Voldoet aan criteria	1	0	0
Bevat geen kritiek, serieus, of error fouten	2	1	14
Bevat geen kritiek of serieus fouten	3	3	1
Bevat geen kritiek fouten	10	12	0

Tabel 2: Overzicht gevonden fouten in dataset 2

	Zonder filter		Met filter	
	Jaccard	Cosine	Jaccard	Cosine
Gemiddelde	0,86	0,93	0.91	0.99
Mediaan	0,89	0,96	0.92	0.99
Standaardafwijking	0,10	0,10	0.07	0.02

Tabel 3: Overzicht van de similariteitsscores.

Resultaten

Originele PDF

- Geen tekst aanwezig in het document
- VeraPDF:
 - Error op 5 regels en 240 checks
- Onleesbaar voor schermlezers

Nieuwe PDF

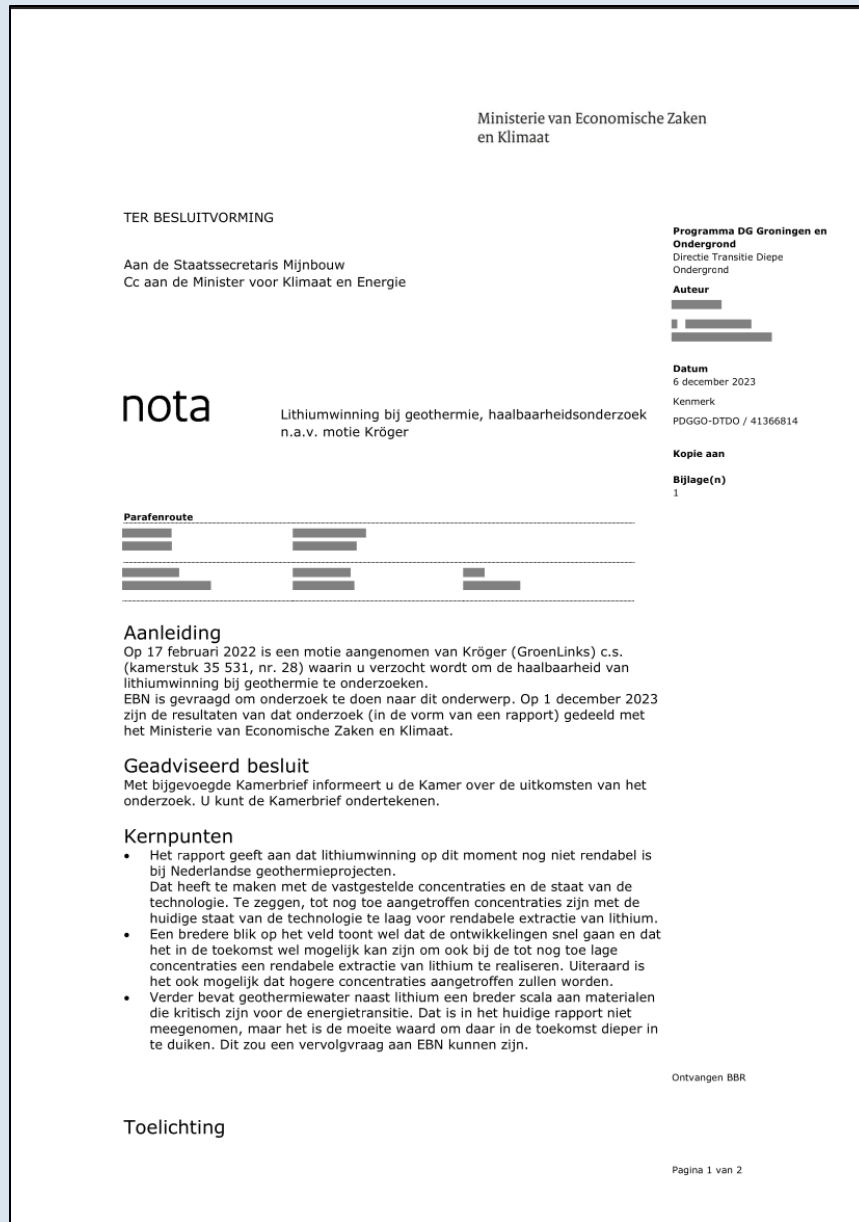
- Volledige “digital born” PDF
- Enkele OCR fouten
- Jaccard: 0,95
- Cosine: 1,0

Positief:

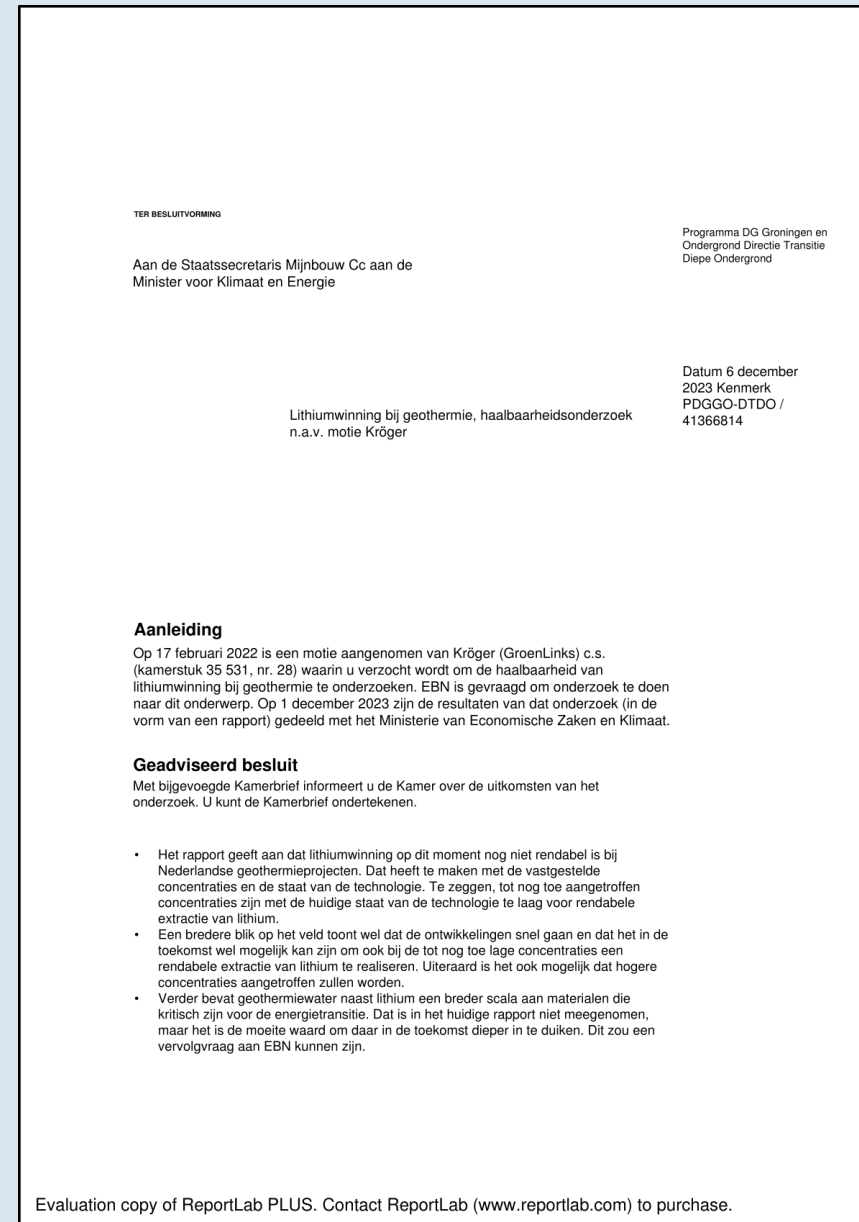
- Kritieke fouten zijn weg
- Indeling nagenoeg identiek

Negatief:

- OCR fouten in tekst
- Weglaten van tekst
- Geen afbeeldingen en tabellen



Figuur 7: Originale Woo besluit



Figuur 8: Gecreerde PDF door converter

Samenvatting

Er ontbreekt voornamelijk metadata en/of PDFs bevatten geen tagging. De converter kan dit tot zekere hoogte omzetten naar een nieuwe PDF die zowel in indeling als inhoud grotendeels overeenkomt met het origineel. Een stap dichterbij een toegankelijke toekomst voor iedereen.



Bedankt

Vragen?