

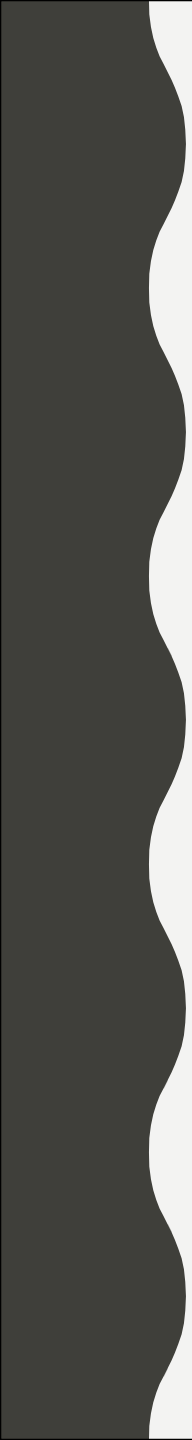
Nieuwe technieken voor toegankelijke pdf

Iacobien Riezebosch
Geïnviteerd expert PDF Accessibility LWG

Voorstellen: Iacobien Riezebosch

- ☐ Digitale toegankelijkheid sinds 2002
- ☐ Organisatieaanpak
- ☐ Complexe pdf-vraagstukken
- ☐ Publicatie 'rollenboekje'
- ☐ Geïnviteerd expert
PDF Accessibility LWG
- ☐ Freelancer
- ☐ Documenten-en-toegankelijkheid.nl
- ☐ Adviseur DigiToegankelijk





Inhoud

1. Waarom?
 2. PDF Association en Werkgroep LWG
 3. Standaarden
 4. Achtergrond nieuwe technieken
 5. Inhoud en opzet
- 

1. Waarom nieuwe technieken voor toegankelijke pdf?

Er zijn toch al pdf-technieken van het W3C?

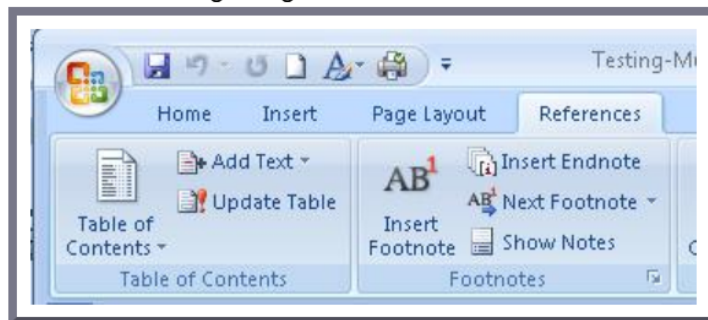
- Groot probleem is slechte software
- Behoefte aan uitleg, goede voorbeelden en fouten
- Onduidelijkheid over hoe toegankelijk pdf in elkaar zit.
- Goede informatie lastig te vinden

WCAG-technieken: Niet leverancier-neutraal

Example 1: Converting a table of contents created with Microsoft Word 2007 and creating bookmarks for Adobe Reader 9 and Acrobat 9 Pro

This example is shown with Microsoft Word and Adobe Acrobat Pro. There are other software tools that perform similar functions. See the list of other software tools in [PDF Authoring Tools that Provide Accessibility Support](#).

1. Create a table of contents at the beginning of the Word document.



2. Use Save as... > Adobe PDF to convert the Word document to PDF, specifying both of the following:
 - Enable Accessibility and Reflow with Tagged Adobe PDF
 - Convert Word Headings into Bookmarks

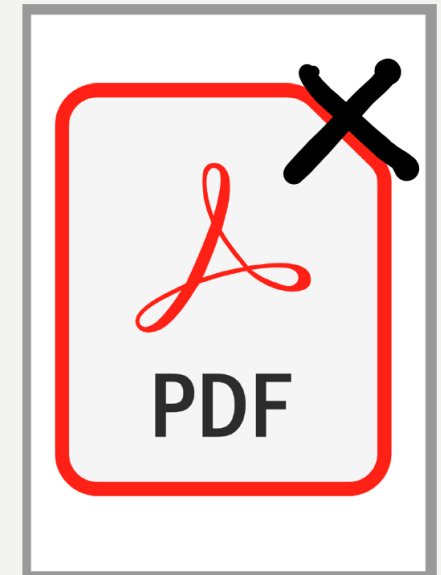
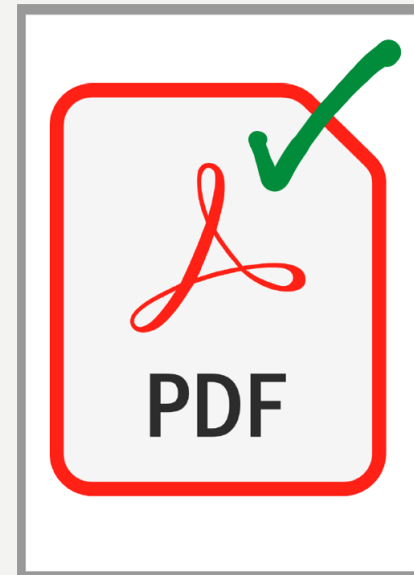
The table-of-contents entries in the converted document will be linked to the headings in the document.

In addition, the headings will appear as PDF Bookmarks in the left-hand Navigation pane.

Wat doe je zonder deze pakketten?

Gebrek aan goede voorbeelden en fouten van een pdf zelf

- Hoe ziet een goed voorbeeld of fout voorbeeld eruit?



WCAG-technieken gedateerd (1)

Example 2: Specifying the default document language in a PDF document using Microsoft Word 2007

This example is shown with Microsoft Word. There are other software tools that perform similar functions. See the list of other software tools in [PDF Authoring Tools that Provide Accessibility Support](#).

Documents authored in Microsoft Word: "In some instances, even if the document language has been specified in the source file, the information about document language is not conveyed to the PDFMaker. Setting the language for an entire document in the Document Properties dialog box [see Example 1] corrects all errors related to this option."([Adobe® Acrobat® 9 Pro Accessibility Guide: Creating Accessible PDF from Microsoft® Word](#))

Word 2007: 2007 – 2017

WCAG-technieken gedateerd (2)

Example 1: Adding an /E entry to an abbreviation using Adobe Acrobat 9 Pro's Tags panel

This example is shown with Adobe Acrobat Pro. There are other software tools that perform similar functions. See the list of other software tools in [PDF Authoring Tools that Provide Accessibility Support](#).

In a tagged PDF document:

1. Select the Tags panel, using Views > Navigation Panes > Tags
2. Select the first instance of the abbreviated text that needs to be expanded. If the selected text is part of a larger tag, access the Tags panel options menu, select Create Tag from Selection, and create a new Span tag. In this example, the text "WCAG2" (within the LBody tag) has been enclosed in a Span tag.

3. In the Tags panel, access the context menu for the created text and select Properties to open

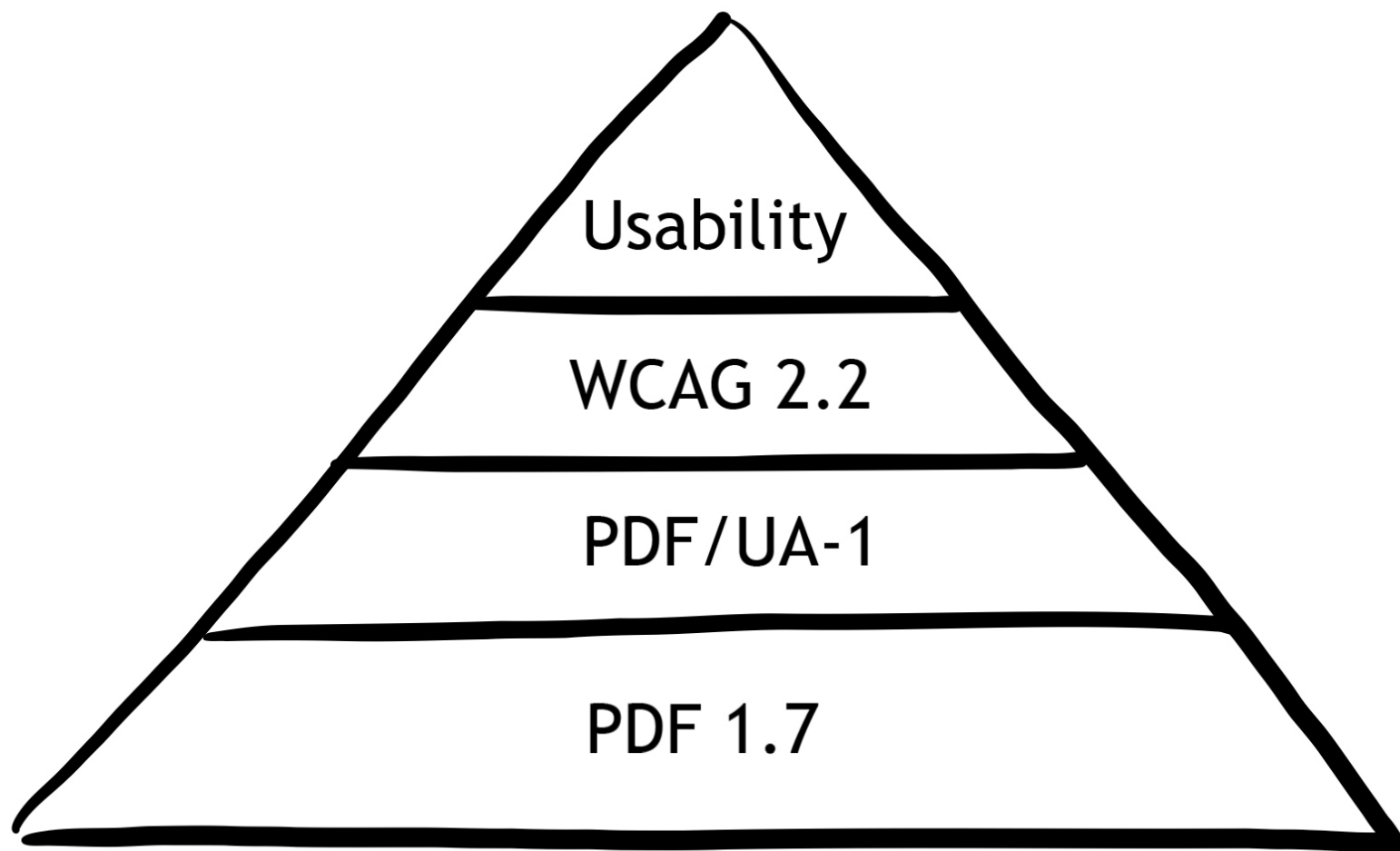
Acrobat 9: 2008 - 2013

WCAG-technieken niet onderhouden

- PDF-Technieken voor WCAG 2.0
- Verouderde informatie
- Dode links

Oops! That page can not be found.

PDF/UA voor technische toegankelijkheid



Grote oorzaak ontoegankelijke pdf

- Kwaliteit software voor creatie, export en bewerking
- Kwaliteit software voor lezen en gebruiken van pdf



2. PDF Association

Leverancier-neutraal platform voor het ontwikkelen van open specificaties en standaarden voor pdf-technologie.



PDF Accessibility LWG

- Werkgroep van de PDF Association
- Vanaf december 2018
- Doel: leverancier-neutraal alternatief maken voor (verouderde) pdf-technieken
- 'Technieken voor toegankelijke pdf' verschijnt in het Engels
- Openbaar

Speciale werkgroep



Solution Agent

About Us ▾

News

Events

Resources ▾

Community ▾

Member ▾



Technical resources



Get ISO 32000-2 at no cost



Become a Member!

PDF Accessibility LWG

The PDF Accessibility Liaison Working Group (LWG) was originally formed to continue the work started at the December 2018 [PDF Techniques Accessibility Summit](#); producing industry-supported example PDF files demonstrating techniques for achieving accessible PDF.

Today the PDF Accessibility LWG meets weekly to review example "pass" and "fail" PDF files ("techniques"), and develop appropriate metadata. Publication of the first set of Techniques is expected in early 2024.

The PDF Association is grateful to its Partner member [Dual Lab](#) for hosting the Jira instance used by this LWG.

COMMUNITY CHAIR



[Markus Erle](#)
axes4 GmbH



[Jochen Fehling](#)
axes4 GmbH



[Birgit Peböck](#)
Barrierefrei PDF OG

Standaarden en toegankelijke pdf

- PDF
 - PDF 1.7 met daarbij PDF/UA-1
 - PDF 2.0 met daarbij PDF/UA-2
- WCAG

Let op: er zijn meer standaarden voor pdf, zoals PDF/A

Matterhorn-protocol

- Testprotocol: Maakte PDF/UA-1 testbaar
- Beschreven **fouten** PDF/UA-1
- 31 meetpunten, 134 beschreven meetbare foutcondities
- Handmatig te meten: 47
- Automatisch: 87

Checkpoints

1. Real content tagged
2. Role Mapping
3. Flickering
4. Color and Contrast
5. Sound
6. Metadata
7. Dictionary
8. OCR Validation
9. Appropriate tags
10. Character Mappings
11. Declared Natural language
12. Stretchable Characters
13. Graphics
14. Headings
15. Tables
16. Lists
17. Mathematical Expressions
18. Page headers and footers
19. Notes and references
20. Optional content
21. Embedded files
22. Article Threads
23. Digital Signatures
24. Non-interactive forms
25. XFA
26. Security
27. Navigation
28. Annotations
29. Actions
30. Xforms
31. Fonts

Foutcondities Matterhorn

- Beschreven afwijking van standaard
- Toegankelijkheidsissue

Voorbeeld ijkpunt 1: Echte inhoud is getagd

- 01-001 Een artefact is getagd als echte inhoud
- 01-002 Echte inhoud is gemarkeerd als een artefact
- 01-003 Content die gemarkeerd is als een artefact staat binnen getagde inhoud
- 01-004 Getagde inhoud staat in een artefact
- 01-005 Content is niet gemarkeerd als artefact of als echte inhoud
- 01-006 De gebruikte tag en de attributen zijn semantisch niet geschikt voor het structuurelement. Bijvoorbeeld Document, Part, Art, Sect, Blockquote,

Vergelijk: WCAG voor pdf

- 1.3.1 Info en relaties

Informatie, structuur, en relaties overgebracht door presentatie kunnen door software bepaald worden of zijn beschikbaar in tekst.

Technieken W3C

Sufficient Techniques

Note: Other techniques may also be sufficient if they meet the success criterion. See [Understanding Techniques](#).

Situation A: The technology provides semantic structure to make information and relationships conveyed through presentation programmatically determinable:

- [G115](#): Using semantic elements to mark up structure **AND**
- [G117](#): Using text to convey information that is conveyed by variations in presentation of text
- [G140](#): Separating information and structure from presentation to enable different presentations
- Making information and relationships conveyed through presentation programmatically determinable using the following techniques:
 - [G138](#): Using semantic markup whenever color cues are used
 - [PDF6](#): Using table elements for table markup in PDF Documents
 - [PDF20](#): Using Adobe Acrobat Pro's Table Editor to repair mistagged tables
 - [PDF10](#): Providing labels for interactive form controls in PDF documents
 - [PDF12](#): Providing name, role, value information for form fields in PDF documents
 - [PDF9](#): Providing headings by marking content with heading tags in PDF documents
 - [SCR21](#): Using functions of the Document Object Model (DOM) to add content to a page
 - [PDF11](#): Providing links and link text using the Link annotation and the /Link structure element in PDF documents
 - [PDF17](#): Specifying consistent page numbering for PDF documents
 - [PDF21](#): Using List tags for lists in PDF documents

Beschreven fouten W3C

Failures

- F2: Failure of Success Criterion 1.3.1 due to using changes in text presentation to convey information without using the appropriate markup or text
- F33: Failure of Success Criterion 1.3.1 and 1.3.2 due to using white space characters to create multiple columns in plain text content
- F34: Failure of Success Criterion 1.3.1 and 1.3.2 due to using white space characters to format tables in plain text content
- F42: Failure of Success Criteria 1.3.1, 2.1.1, 2.1.3, or 4.1.2 when emulating links
- F43: Failure of Success Criterion 1.3.1 due to using structural markup in a way that does not represent relationships in the content
- F46: Failure of Success Criterion 1.3.1 due to using th elements, layout tables
- F48: Failure of Success Criterion 1.3.1 due to using the pre element to markup tabular information
- F87: Failure of Success Criterion 1.3.1 due to inserting non-decorative content by using ::before and ::after pseudo-elements and the 'content' property in CSS
- F90: Failure of Success Criterion 1.3.1 for incorrectly associating table headers and content via the headers and id attributes
- F91: Failure of Success Criterion 1.3.1 for not correctly marking up table headers
- F92: Failure of Success Criterion 1.3.1 due to the use of role presentation on content which conveys semantic information

Fouten voor PDF/UA:

1. Echte inhoud is getagd

- 01-001 Een artefact is getagd als echte inhoud
- 01-002 Echte inhoud is gemarkeerd als een artefact
- 01-003 Content die gemarkeerd is als een artefact staat binnen getagde inhoud
- 01-004 Getagde inhoud staat in een artefact
- 01-005 Content is niet gemarkeerd als artefact of als echte inhoud
- 01-006 De gebruikte tag en de attributen zijn semantisch niet geschikt voor het structuurelement. Bijvoorbeeld Document, Part, Art, Sect, Blockquote, Caption, TOC, TOCI, Index, NonStruct, Private, P, H, H1, H2 [...]

Beschreven fouten voor elk checkpoint

1. Real content tagged
2. Role Mapping
3. Flickering
4. Color and Contrast
5. Sound
6. Metadata
7. Dictionary
8. OCR Validation
9. Appropriate tags
10. Character Mappings
11. Declared Natural language
12. Stretchable Characters
13. Graphics
14. Headings
15. Tables
16. Lists
17. Mathematical Expressions
18. Page headers and footers
19. Notes and references
20. Optional content
21. Embedded files
22. Article Threads
23. Digital Signatures
24. Non-interactive forms
25. XFA
26. Security
27. Navigation
28. Annotations
29. Actions
30. Xforms
31. Fonts

4. Achtergrond nieuwe technieken voor toegankelijke pdf

- Voor wie?
- Wat bieden de technieken wel en niet?

Voor wie?

- Experts
- Softwaremakers (creatie)
- Softwaremakers (pdf-lezers)
- Fabrikanten en ontwikkelaars
hulptechnologie
- Developers
- Onderzoekers
- Pdf-verbeteraars

Wat bieden deze technieken?

- Praktische voorbeelden van eisen voor toegankelijke pdf
- Praktische voorbeelden van toegankelijke pdf's en niet-toegankelijke pdf's
- Duidelijke(re) taal
- Testprocedures
- Uitleg, naslag en verwijzingen

Wat bieden deze technieken niet?

- Geen voorstel om altijd voor pdf-formaat te kiezen. Denk aan de gebruiker.
- Geen instructies voor specifieke software.
- Geen uitleg voor het maken van toegankelijke pdf's of het repareren van ontoegankelijke pdf's.

PDF-Technieken

- Nu PDF 1.7 (PDF/UA-1), later ook PDF 2.0 (PDF/UA-2)

5. Inhoud en opzet nieuwe technieken voor toegankelijke pdf

Pass / Fail: **PASS**

Description

The objective of this technique is to show how to tag non-text content that is intended to be consumed as text. In this example, graphics were used to place the word "paragraph" on the page, instead of text objects. To make the characters of this graphic extractable, ActualText on a separate Span tag is used. The ActualText matches the visual appearance of the related graphics.

Note: The Span tag is always the recommended choice if ActualText is required because ActualText removes the semantics of tags. In spite of the previous statement, ActualText may also work when the tag is one of the following: Lbl (outside of List structures), P, BibEntry, BlockQuote, Caption, Code, Note, Quote. However, ActualText should never be assigned to other tags.

Voorbeeldbestand met alleen PASS-voorbeeld of beschreven fout

Example file

[UA_Tpdf-0160-F2-Image-representing-text](#)

Testprocedure

Test

Procedure

1. Check that the related characters are present as extractable characters for content intended to be consumed as text.
2. Check that Unicode can be derived for any text content.
3. Check that the extractable characters match their visual appearance.
4. Check that the Natural Language is set for any text content.

Expected Results

Checks #1 through #4 are all true.

Verwijzingen en vindbare technieken

Applicable WCAG 2.2 Success Criteria

[1.1.1](#), [1.3.1](#), [1.4.5](#), [3.1.1](#), [3.1.2](#)

Applicable **Matterhorn Protocol** checkpoints

01-006, 08-001, 08-002, 10-001, 11-001, 11-002, 11-006, 13-001, 13-004, 13-005, 13-008

Technique finder: UA_Tpdf-F2_07

Twee soorten technieken

- Fundamentals
- Use cases

Voorbeelden: Pass en Fail

- Alleen relevante functionaliteit en/of inhoud
- PASS: foutenvrije pdf
- FAIL: alleen de besproken fout
- Gevalideerd met een veelheid aan tools

Verklarende woordenlijst

- 'Glossary of accessibility terminology in PDF'
- In meer eenvoudige taal (zoals 'tags')
- Al online: pdfa.org/glossary-of-accessibility-terminology-in-pdf/

Fundamentals

- 5 basiseisen
- Voor elke pdf, ongeacht inhoud of functionaliteit
- Nu 36 Fundamentele technieken

Fundamentals in hoofdlijnen

1. Technische basisregels worden gevolgd.
2. Tekstinhoud is machineleesbaar.
3. Echte inhoud en artefacten worden van elkaar onderscheiden.
4. De volgorde van inhoud is logisch.
5. Juiste semantiek wordt gebruikt.

1. Technische basisregels worden gevolgd

- Pdf voldoet aan technische eisen voor een toegankelijke getagde pdf
- Dus conformiteit met PDF 1.7 en PDF/UA-1
- Testen met validator

Voorbeeld van een Techniek (Automatisch testen)

Controleer of de technische basisregels van ISO 32000 en PDF/UA worden gevolgd door een PDF/UA-validator uit te voeren.

Voorbeeld: Beschrijving

Het doel van deze techniek is om te laten zien hoe je technische basisregels correct volgt. In dit voorbeeld wordt een aangepaste tag "Paragraph" toegewezen aan de tag P.

Voorbeeld van de metadata

- Naam van de techniek (nummer)
- Matterhorn-protocol: 02-001
- WACG 2.2-succescriterium: 1.3.1

Voorbeeld pdf: PASS-Voorbeeld

Accessibility tags

Tags

<Document>

<H1>

This is an H1

<P>

This is the first paragraph.

<P>

This is the second paragraph.

<P>

This is the third paragraph.

This is an H1

This is the first paragraph.
Second line of first paragraph.
Third line of first paragraph.
Fourth line of first paragraph.
Fifth line of first paragraph.

This is the second paragraph.
Second line of second paragraph.
Third line of second paragraph.
Fourth line of second paragraph.
Fifth line of second paragraph.

This is the third paragraph.
Second line of third paragraph.
Third line of third paragraph.
Fourth line of third paragraph.
Fifth line of third paragraph.

Custom Style en Roltoewijzing

Logical Structure Tree

- ✓  Document
 - >  H1
 - >  Paragraph
 - >  Paragraph
 - >  Paragraph

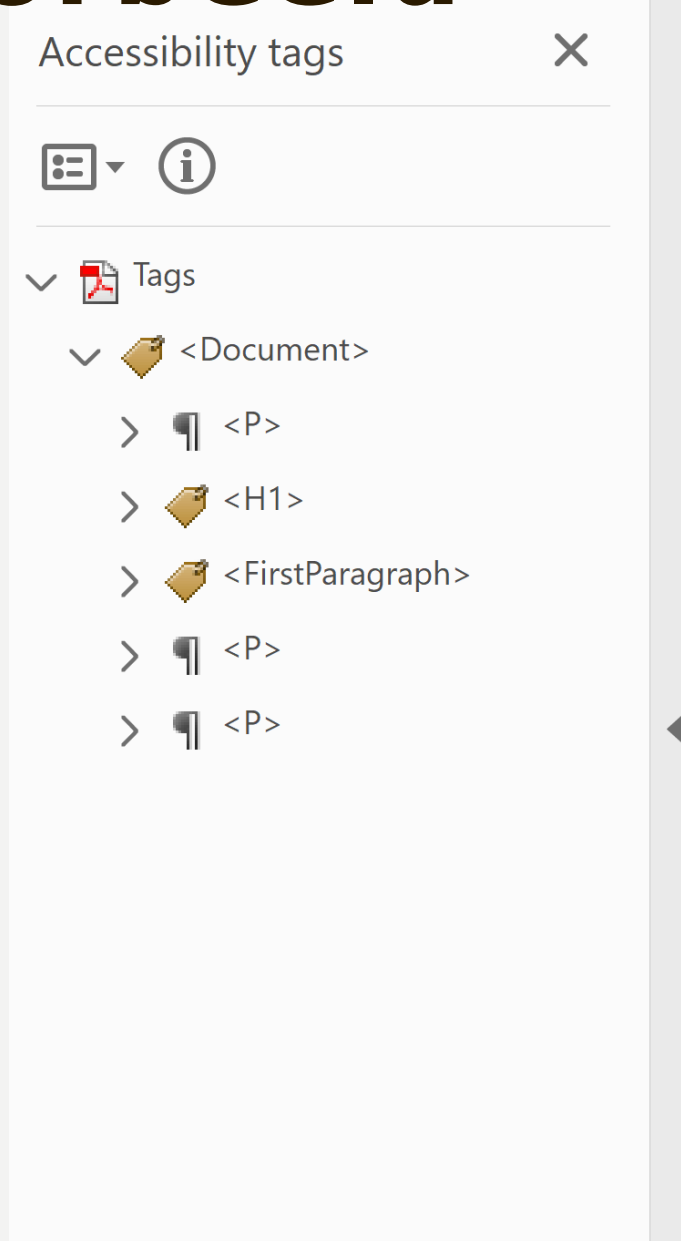
Rolemap

Structure Type	Role
Paragraph	P

Voorvertoning schermlezer (pass)

Document	H1	This is an H1
	P	This is the first paragraph. Second line of first paragraph. Third line of first paragraph. Fourth line of first paragraph. Fifth line of first paragraph.
	P	This is the second paragraph. Second line of second paragraph. Third line of second paragraph. Fourth line of second paragraph. Fifth line of second paragraph.
	P	This is the third paragraph. Second line of third paragraph. Third line of third paragraph. Fourth line of third paragraph. Fifth line of third paragraph.

Fail-voorbeeld



FAIL

This is an H1

This is the first paragraph.
Second line of first paragraph.
Third line of first paragraph.
Fourth line of first paragraph.
Fifth line of first paragraph.

This is the second paragraph.
Second line of second paragraph.
Third line of second paragraph.
Fourth line of second paragraph.
Fifth line of second paragraph.

This is the third paragraph.
Second line of third paragraph.
Third line of third paragraph.
Fourth line of third paragraph.
Fifth line of third paragraph.

Voorvertoning schermlezer (fail)

Document	P	FAIL
	H1	This is an H1
		This is the first paragraph. Second line of first paragraph. Third line of first paragraph. Fourth line of first paragraph. Fifth line of first paragraph.
	P	This is the second paragraph. Second line of second paragraph. Third line of second paragraph. Fourth line of second paragraph. Fifth line of second paragraph.
	P	This is the third paragraph. Second line of third paragraph. Third line of third paragraph. Fourth line of third paragraph. Fifth line of third paragraph.

2. Inhoud is machineleesbaar

- Alle inhoud moet in tekst (unicode) beschikbaar zijn.
- Deze tekst moet overeenstemmen met de inhoud.
- Er is een taal ingesteld voor alle tekst.
- Spaties zijn aanwezig in machine-leesbare tekst.

Voorbeeld

ActualText biedt de juiste tekens en vervangt de tekst met OCR-fouten.

Voorbeeld (OCR)

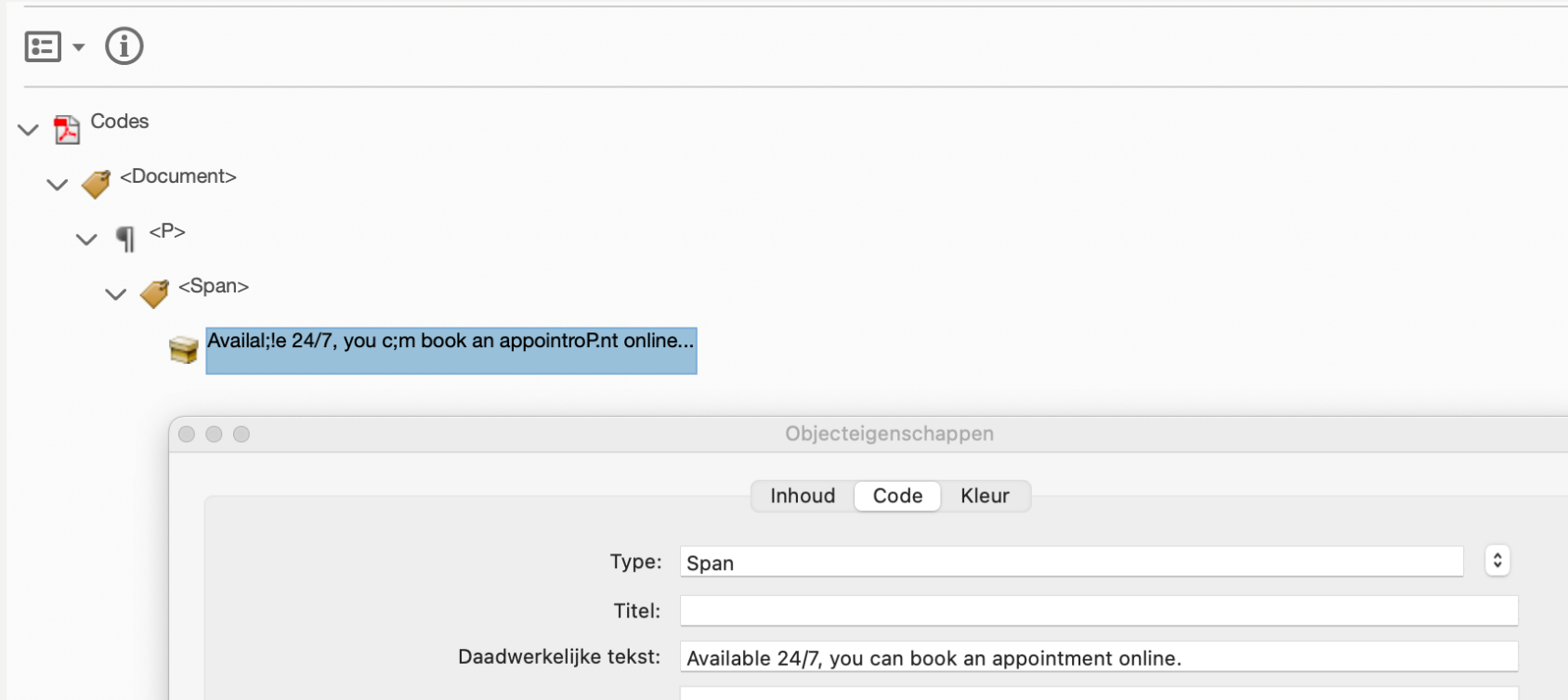
Available 24/7, you can book an appointment online.

- ✓  Codes
- ✓  <Document>
- ✓  <P>
- ✓ 

 Availal;le 24/7, you c;m book an appointroP.nt online...

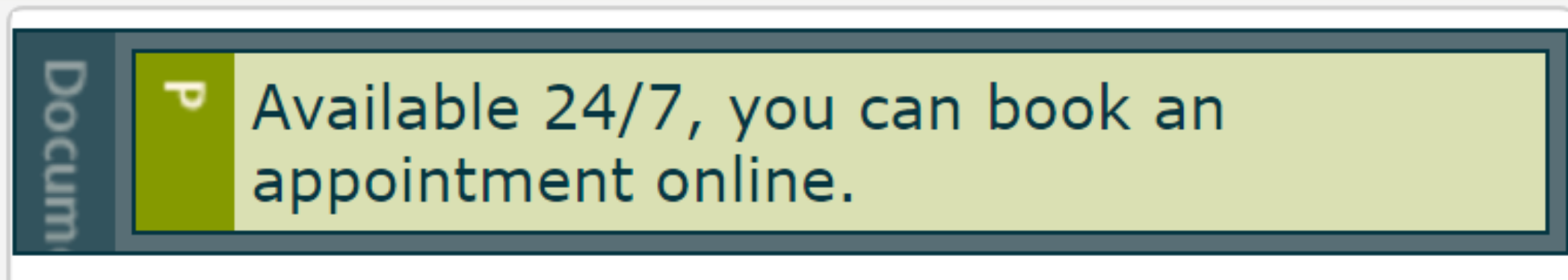
Availal;le 24/7, you c;m book an
appointroP.nt online....

Machineleesbaar



Available 24/7, you can book an appointment online

Voorvertoning schermlezer



3. Echte inhoud en artefacten worden van elkaar onderscheiden

- Niet alle inhoud van een pdf biedt informatie.
- 'Echte inhoud' geeft informatie en is getagd.
- Decoratieve inhoud is gemarkeerd als artefact.
- Inhoud kan niet tegelijk échte inhoud en een artefact zijn.

Fout voorbeeld

FAIL

This paragraph is real content, therefore it is tagged with a P structure element. The blue background is decorative and therefore it is artifacted.

Voorvertoning schermlezer (fout voorbeeld artefact)

Document	P	FAIL
	P	This paragraph is real content, therefore it is tagged with a P structure element. The blue background is decorative and therefore it is artifacted.
	Figure	
	Alt	Blue background

4. De volgorde van inhoud is logisch

- Tags staan in een logische volgorde in de structuurboom
- Software kan hiermee de inhoud op een logische wijze presenteren
- Weerspiegelt de echte inhoud, zoals bedoeld door auteur

Volgorde van inhoud correct

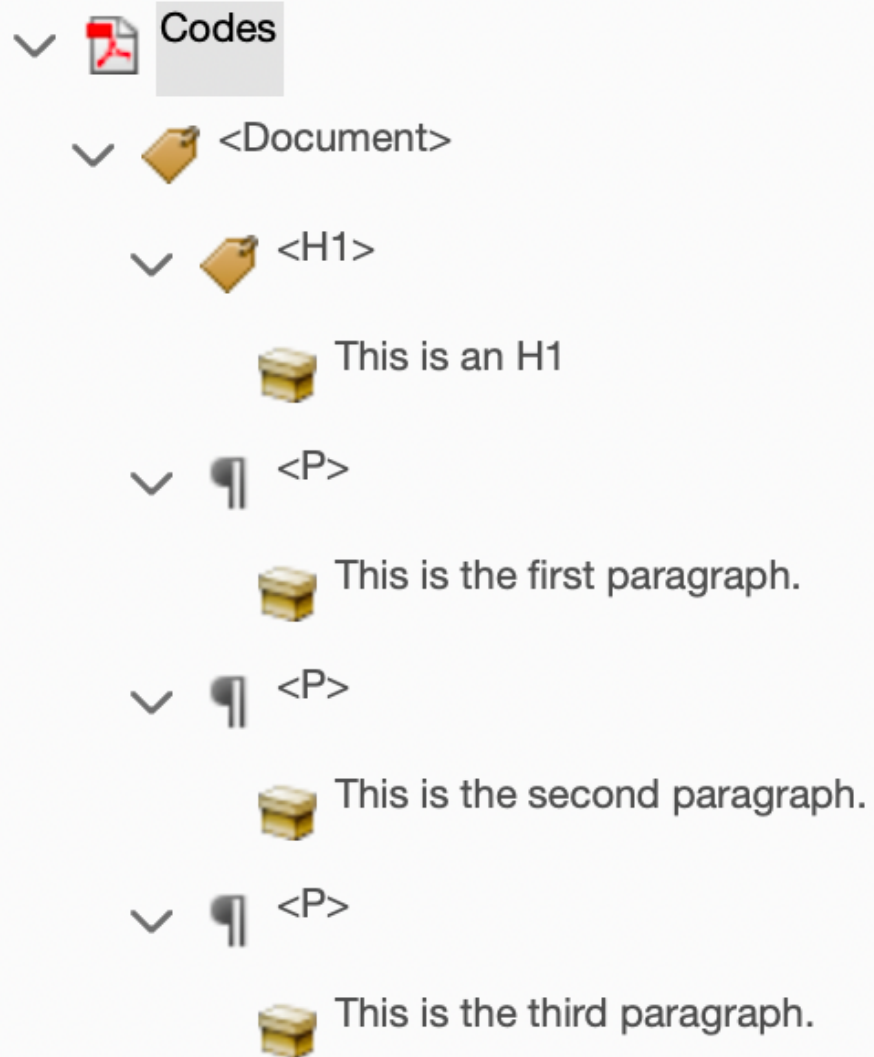
This is an H1

This is the first paragraph.

This is the second paragraph.

This is the third paragraph.

Goed voorbeeld



Volgorde van inhoud incorrect

FAIL

This is an H1

This is the first paragraph.

This is the second paragraph.

This is the third paragraph.

Geen goede leesvolgorde

  Codes

  <Document>

  <P>

 FAIL

  <P>

 This is the second paragraph.

  <H1>

 This is an H1

  <P>

 This is the third paragraph.

  <P>

 This is the first paragraph.

5. Juiste semantiek

- De semantisch meest geschikte tag moet worden gebruikt.
- Elke eenheid van echte inhoud moet met 1 enkele tag gemarkeerd worden, zelfs als delen visueel gescheiden zijn.

Goede voorbeelden

- Een enkele alinea die over 2 kolommen verspreid staat of onderbroken wordt door een afbeelding is getagd met 1 enkele P-tag.
- Een kop op niveau 1 is getagd met een H1
- Een lijst is getagd met L, LI, Lbl en LBody tags.

Goed voorbeeld: Link staat op 2 regels

Link: <https://www.pdfa.org/participating-in-the-pdf-techniques-accessibility-summit/>



Foute voorbeelden

- Voor 1 alinea zijn meerdere P-tags gebruikt.
- Een kop is getagd met een P-tag.
- Een tabel gaat door op de volgende pagina en is getagd met meerdere Table-tags.
- Een enkele lijst is semantisch opgesplitst in meerdere lijsten.

Bij elk voorbeeld dus:

- Beschrijving
- Indien nodig links naar definities
- Metadata
- Testprocedure
- Goed voorbeeld / fout voorbeeld in pdf

Technieken voor Use Cases

- Voor documenten met specifieke inhoud of technologie
- Zoals: koppen, afbeeldingen, lijsten, tabellen, voetnoten, kopteksten en voetteksten, links, inhoudsopgave, paginainde, kolomeinde et cetera.

Use cases

- Eerste eis: voldoet aan de 5 Fundamentals
- Voor elke use case: test criteria, goede voorbeelden en foute voorbeelden, mogelijke technieken en/of methodes
- 160 in concept klaar

Later: Technieken voor PDF 2.0 en PDF/UA-2

- PDF 2.0 nog niet breed ondersteund
- Technieken hiervoor volgen later
- Je kan niet aan PDF 1.7 en PDF 2.0 tegelijk voldoen

Lancering

- Verwacht in 2024

Wat heb ik hier aan?

- Software moet beter: info voor developers
- Meer en duidelijkere technieken met voorbeelden
- Informatie beter ontsloten
- Meer informatie voor onderzoekers en pdf-verbeteraars
- Verwijs je leveranciers, vormgevers en ontwikkelaars
- Weet waarvoor PDF/UA nodig is
- Duidelijkheid en duurzaamheid

PDF Association en toegankelijkheid

Onder andere:

- Ontwikkeling PDF/UA-2
- Matterhorn Protocol
- Tagged PDF Best Practice Guide
- PDF/UA Reference Suite
- Well-Tagged PDF (WTPDF) for Accessibility and Reuse in PDF 2.0

Meer informatie

- documenten-en-toegankelijkheid.nl
- [PDF Association](https://pdfa.org/)
- Webinar
DigiToegankelijk.nl op 18
juli 2024:
[PDF/UA en andere PDF-standaarden](#)

Vragen



Iacobien Riezebosch

[linkedin.com/in/iacobien/](https://www.linkedin.com/in/iacobien/)

documenten-en-toegankelijkheid.nl

Hand-out

Deze hand-out is een weergave van de presentatie die Iacobien Riezebosch van documenten-en-toegankelijkheid.nl verzorgde voor het Nationaal Congres Digitale Toegankelijkheid 2024.

De hand-out is ondersteunend bij de presentatie waar alle concepten en voorbeelden zijn uitgelegd.

Copyright afbeeldingen: documenten-en-toegankelijkheid.nl, 2024