

Status des Tagged PDF Projekts

Ulrike Fischer, Bonn
L^AT_EX Project Team

4–6.4.2024

Dante 2024

Tagging ist möglich

<https://pdfa.org/wtpdf>

WTPDF examples

The PDF Association invites its members to contribute example PDF files demonstrating conformance to WTPDF. To be included such files must be:

- Valid PDF 2.0 (a clean report from the Arlington PDF Model is required);
- Valid PDF/UA-2 (and [PDF/A-4](#) and/or [PDF/X-6](#), for bonus points :-);
- Valid WTPDF (that is, conform to at least one WTPDF conformance level).

Additional [PDF Declarations](#), such as [Validated as conforming to PDF/UA-2](#) or Validated as [WTPDF, conformance level accessibility](#), if applicable.

Note: The PDF Association does not endorse these developers' representations but presents the examples "as is".

PDF Association Member	Example(s)
BFO (Big Faceless Organization)	BFO-WTPDF-&-PDFua2.pdf
LaTeX Project	Examples of various types of content²

<https://github.com/latex3/tagging-project/discussions/72>



davidcarlisle last week Maintainer

WTPDF / PDF/UA-2 Examples by the LaTeX Project

The following files demonstrate various aspects of [Well Tagged PDF](#) documents conforming to [PDF/UA-2](#).

They were all generated with LuaLaTeX (`lualatex-dev` in [TeX Live 2024](#)).

Ziele des Tagged PDF Projekts

- *automatisch* und *einfach*
- für (fast) jedes Dokument
- für (fast) jeden Nutzer

- native Unterstützung im LaTeX Kernel
- direkt nutzbar Zwischenergebnisse
- Paketautoren einbinden
- Feedback ermöglichen



On September 4, 1882, the electrical age began. [...] By the end of the month, they had 59 customers. By the next year, they had 513.

PDF 2.0 ist wichtig!

- mehr Tags für Text (Title, Aside, FEnote, ...)
 - Namespaces für Tags
 - “structure destinations” für Links zu Strukturen
 - “associated files”
 - MathML Tags
 - neue Standards: PDF/UA-2, Well-Tagged PDF (WTPDF)
- } wichtig für Math Tagging

PDF 2.0 Support ist noch mau aber es bewegt sich ...

<https://pdfa.org/wtpdf/>

Codeorganisation: viele Github-Repos betroffen

<https://github.com/latex3>

[latex3](#) l3kernel, l3backend

[pdfresources](#) Code für PDF-Management

[tagpdf](#) Core Code für Tagging

[latex2e](#)

- base: lthooks, ltproperties, ltsockets, ltagging, ...
- required/tools: array, longtable ...

[required/latex-lab](#) Experimentelle Module für Taggingsupport

Generelle Issues: <https://github.com/latex3/tagging-project>

- Standardklassen
- “Leslie Lamport Befehle”
- hyperref, biblatex, natbib, footmisc
- einfache Tabellen, tabularx, longtable
- Mathematik xmit manuellen Eingriffen

Was geht noch nicht

- Andere Klassen: KOMA, memoir, beamer, ...
- `\savebox`, `\usebox`
- `tabbing`
- komplizierte Tabellen
- voll automatische Mathematik
- viele externe Pakete

Demos: Tagging aktivieren

```
\DocumentMetadata{
  pdfversion=2.0,
  pdfstandard=ua-2,
  testphase= %latex-lab Module laden
  {
    phase-III,
      %lists,footnotes,sectioning,toc,marginpar,bibliography,
      %floats,graphics ...

    math,
    table, %tabular, tabularx, longtable
    title, %maketitle
    firstaid %kleinere Anpassungen externer Pakete
  }
}
```

- Dies ist *neuer* Code: Man braucht ein aktuelles TeX-System.
- Dies ist Code in *Entwicklung*: Dinge können sich ändern

- Die Demodateien benutzen `lualatex-dev + unicode-math`

- Goethe: Sprache + Alttexte
- Max und Moritz: Sprachwechsel + Alttexte
- Bibel: an den Grenzen des Machbaren ...
- tagpdf: Listings, tikz und tcolorboxen
- Mathematik: arxiv-Dokument arxiv-3-tagged

Drei Tagging-Methoden

- Formula-Tag interne tagging Struktur mit mathml Tags
- Formula-Tag mit /Alt-Text
- mathml und/oder TeX-Source als “associated file” (AF)
 - Development Version von foxit
`https://github.com/foxitsoftware/PDF_UA-2/blob/main/installer/README.md`
 - Development Version von NVDA (selbe Url)
 - MathCAT plug-in für NVDA

Math

$$|-1| = 1$$

Output in NVDA Sprachviewer

the absolute value of negative 1 is equal to 1

Math

The equations

$$2x + y = 3$$

$$x - y = 0$$

have solution $x = y = 1$.

Output in NVDA Sprachviewer

The equations 2 lines

line 1 2 x plus y is equal to 3

line 2 x minus y is equal to 0

have solution x is equal to y is equal to 1

 `arxiv3-tagged-mathml.html`

Wie kommt man dran?

- 1 latexml (arXiv)
- 2 tex4ht
- 3 In einem Browser mit mathjax
- 4 manuell
- 5 ...

Nächste Schritte

- enumitem
- caption
- Sprachunterstützung
- Mathml-Erzeugung per Skript automatisieren
- ...

