

The newpax package, v0.1

Reinserting annotations from included pdf file

Ulrike Fischer*

2020-03-30

1 Introduction

When including PDF-files in a document – may it be with `\includegraphics` or with `\includepdf` – clickable links and other annotations of these documents are lost.

The newpax package offers some tools to reinsert these annotations. It is based in large parts on the pax package from Heiko Oberdiek.

2 Quick use instructions

2.1 First step: extract the annotations

The luacode offers two functions which take as argument the name of a PDF without the extension. The functions can be used in some lua scripts but also in some document which then must be compiled with lualatex:

```
\documentclass{article}
% load the lua code
\directlua{require("newpax.lua")}
% write .newpax files for newpax.sty
\directlua
{
  newpax.writenewpax("pax-input")
  newpax.writenewpax("figure/pax-input2")
  newpax.writenewpax("example-image-a")
}
% and/or write .pax files for pax.sty
\directlua
{
```

*fischer@troubleshooting-tex.de

```

    newpax.writepax("pax-input")
    newpax.writepax("figure/pax-input2")
    newpax.writepax("example-image-a")
}
\begin{document}
\end{document}

```

2.2 Step two: Using the .pax-file with pax.sty

Ensure that the .pax file created in the first step can be found by your main document. You can then insert your PDF files together with they annotations like in the following listing.

- This works with pdf_latex and lua_latex. lua_latex needs the extra code demonstrated in the document.
- It needs two or three compilations until every reference is correct.
- There is small typo in pax.sty which affects clipping, the patch shown in the listing correct this.
- Don't include PDFs with destinations twice as this will lead to duplicates.
- If annotations should not be reinserted remove the .pax-file.

```

\documentclass{article}
\usepackage{ifluatex,etoolbox}
\usepackage{pdfpages}
%pax needs this to run with lualatex
\ifluatex
\usepackage{pdfltexcmds}
\makeatletter
\let\pdflstrlcmp\pdf@strlcmp
\let\pdflescapename\pdf@escapename
\makeatother
\usepackage{lualtex85}
\fi
%load pax
\usepackage{pax}
%correct a bug in pax affecting clipping
\makeatletter
\patchcmd\PAX@pdf@annot{\PAX@pagellx}{\PAX@page@llx}{}{\fail}
\makeatother
\begin{document}
\includegraphics[scale=0.5,trim=5cm 15cm 8cm 3cm,clip,page=2]{pax-input}
\includegraphics[scale=0.5,trim=5cm 15cm 8cm 3cm,clip,page=1]{pax-input}

\includepdf[pages=-]{figure/pax-input2}
\end{document}

```

2.3 Experimental!!!

Alternative step two: Using the .newpax-file with newpax.sty

3 Background

Clickable links in a PDF are one example of an annotation. Annotations are areas on a page which are associated with an action. A typical annotation object could look like this in the PDF:

```
15 0 obj
<<
/Type /Annot
/Subtype/Link
/Rect [147.716 654.025 301.887 665.15]
/Border[0 0 1]/BS<</S/U/W 1>>/H/I/C[0 1 1]
/A<</Type/Action/S/URI/URI(https://www.latex-project.org)>>
>>
endobj
```

This is an object of type `Annot` and subtype `Link`. The `/Rect` value describes the rectangle of this annotation. The coordinates are absolute coordinates related to the current page. It is important to understand that an annotation is not connected to some page content! The `/Border` setting and the other values on this line describe the look and color of annotation. The `/A` value contains the action, in this case it is an url to an external website.

To “reactivate” the annotations of an included pdf one has to do a number of tasks.

- One must *retrieve and store* the annotations of the included pdf. For links to external links this requires to find only one object like the one shown above. But e.g. internal links point to a destination object and these must be found too.
- One must *recalculate* the rectangle coordinates to fit to the coordinate system of the target page: as the included pdf can be placed at various positions, scaled, rotated and even clipped this is not an easy task. Destinations have rectangles too that must be recalculated.
- One must *reinsert* the annotation and related objects. This has to take into account the a pdf is perhaps not included completely, a link shouldn't point to a missing page or a clipped annotation. It also has to take into account that a pdf is perhaps inserted more than once or in steps.

3.1 Retrieving and storing annotations

Theoretically one can do it manually: Uncompress the PDF (or when using \LaTeX , create directly an uncompressed one), open it in an editor and copy and paste all needed objects. Practically one naturally want some tool.

The pax package from Heiko Oberdiek consists of a perl script and a java-jar file `PDFAnnotExtractor` which can extract the necessary objects. It writes the information to a file with the extension `pax`. When it has been successfully installed it works quite fine. Problems with this approach are

- `PDFAnnotExtractor` requires an external, old version of the java library of `PDFbox` which must be installed manually.
- It requires a java installation and
- it is not extensible.

The `newpax` package comes with a lua-file. It uses the `pdfx` library embedded in `luatex` to extract the annotations and other needed information. `newpax` writes the information to a file with the extension `pax` or `newpax`. The content of the files is (nearly) identical to the content of the `pax`-file written by `PDFAnnotExtractor`. The lua code was written by looking at example outputs from `PDFAnnotExtractor` and reproducing it in lua. The ordering of elements is a bit different and some strings are output in a different way but for the examples I used the resulting `pax`-files can be used together with the original `pax.sty`. It is possible that the lua code is not yet handling all objects or options that `PDFAnnotExtractor` outputs, but the code can rather easily be extended later.

4 Importing annotations with the pax package

The pax package from Heiko Oberdiek consists of a perl script and a java-jar file `PDFAnnotExtractor` which can extract the necessary objects. When it has been successfully installed it works quite fine. Problems are

- `PDFAnnotExtractor` requires an external, old version of the java library of `PDFbox` which must be installed manually.
- It requires a java installation and
- it is not extensible.

which extract the data of all annotations from a pdf and stores them in a text file with the extension `pax`, a style `pax.sty` which patches the `\includegraphics` command to load the `pax`, process it to calculate the rectangles and reinsert destinations and annotations.

The package works actually very well. Problems are

- `PDFAnnotExtractor` requires an external, old version of the java library of `PDFbox`.
- The style works only with `pdflatex` and partly with `lualatex` (with the `luatex85`) and not at all with `xelatex`.

5 Importing annotations with the newpag package