PDF handling in LibreOffice

Jaume Pujantell

Software Engineer

jaume.pujantell@collabora.com

Collabora Online

LibreOffice Conference Bucharest 2023

Overview

- PDFs as graphical documents
- PDFium
- Inserted PDF images

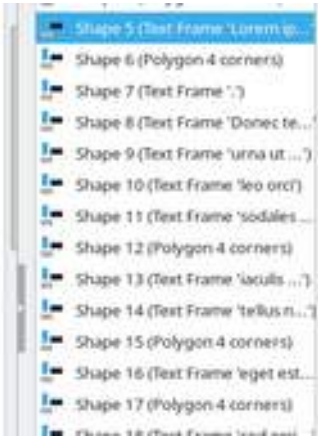
PDFs as graphical documents

Import PDF elements as Shapes

Text is imported into Text Frames

Lorem ipsum dolor sit amet, consectetur adipiscing elit

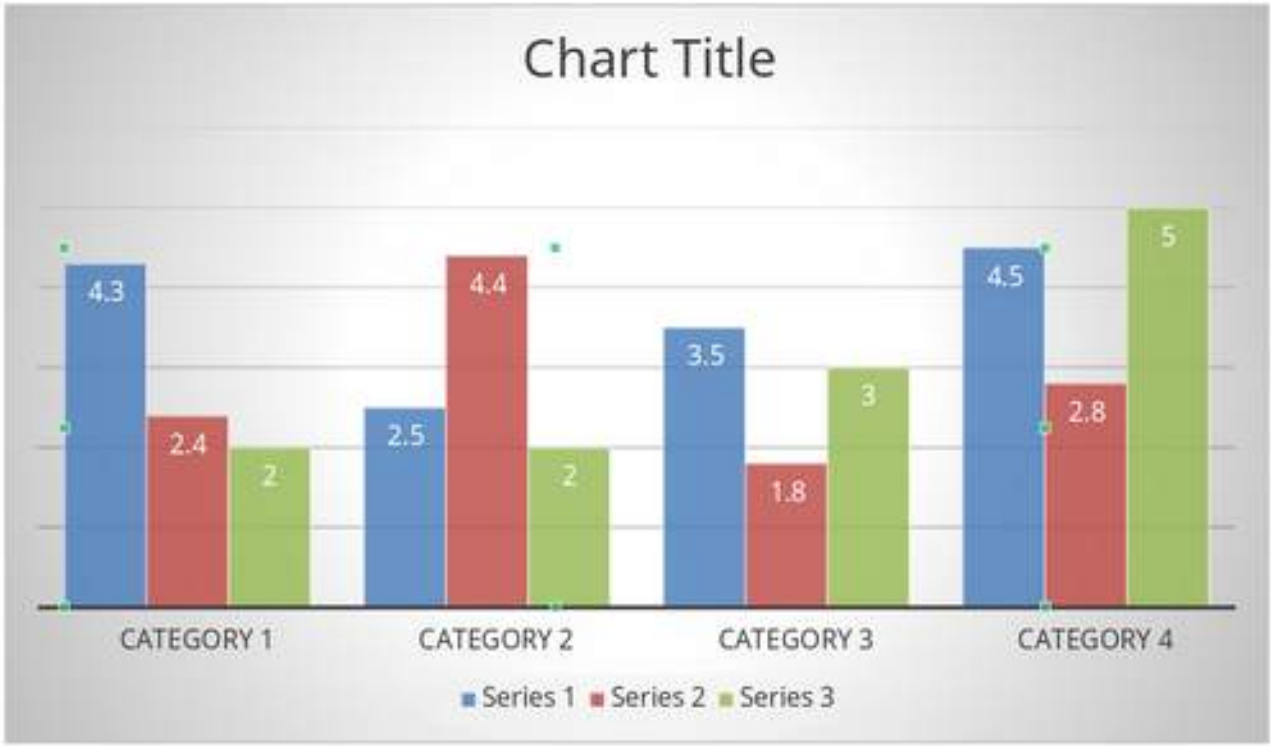
Donec tempor, urna ut pretium imperdiet, leo orci sodales ligula, iaculis pellentesque lorem tellus non felis. Ut ut nisl eget est tristique cursus. Proin nec vehicula neque. Integer eu quam sed orci tristique euismod in ac urna. Sed sed aliquam purus. Etiam venenatis dolor nisi. Morbi et felis quis sapien blandit maximus viverra vel ex. Curabitur egestas lobortis ex, et rutrum ligula rutrum ac. Donec in nibh sem. Nam aliquam massa a consequat sollicitudin. Etiam tincidunt nunc sit amet consectetur consectetur.



LibreOffice Technology

Import PDF elements as Shapes

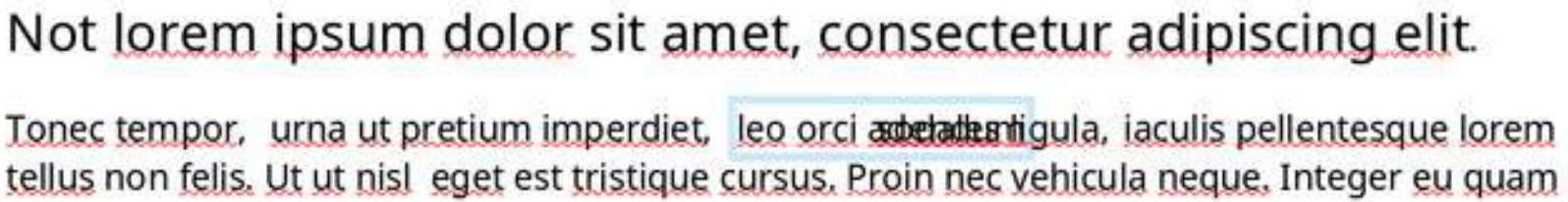
Polygons, curves and images are imported as their equivalent shape



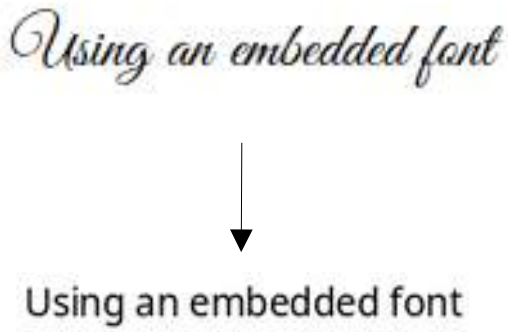
Collabora Online

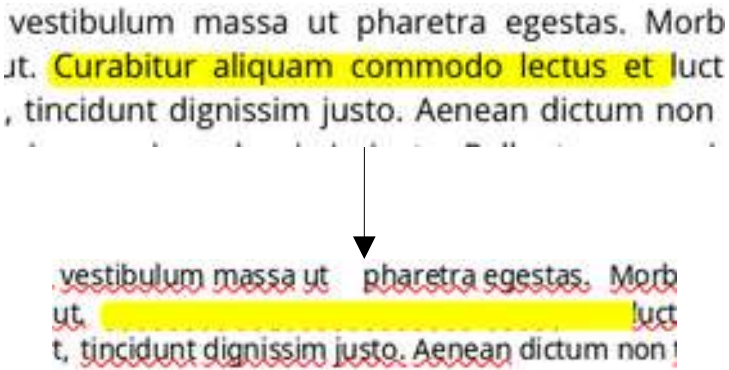
Granular but confusing editing

Some small text edits work fine, but others require remaking parts of the document.



Lost information





PDFium

What is PDFium

- Open Source PDF Rendering Engine
- Generates the bitmap representation of PDF pages
- Provides an API to consult the elements of a PDF

PDFium in LibreOffice

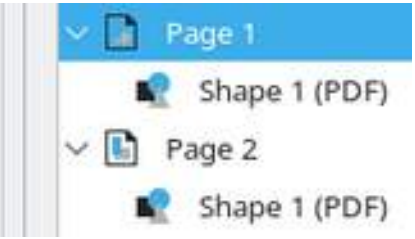
- In Draw is an experimental function, controlled by the environment variable LO_IMPORT_USE_PDFIUM
- In LibreOfficeKit it is always enabled
- When inserting a PDF as an image in another document

Import PDF content as bitmaps

- No possibility to edit directly the document
- Correctly uses embedded fonts and styles

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Donec tempor, urna ut pretium imperdiet, leo orci sodales ligula, iaculis pellentesque lorem tellus non felis. Ut ut nisl eget est tristique cursus. Proin nec vehicula neque. Integer eu quam sed orci tristique euismod in ac urna. Sed sed aliquam purus. Etiam venenatis dolor nisi. Morbi



PDF annotations compatibility

- Most common annotations are supported

Curabitur aliquam commodo lectus et luctus.

- Some still need better support

Inserted PDF images

Ideal use case of PDFium

- The endpoint is an image
- Users won’t expect the ability to edit
- Better quality and fidelity than importing as graphical document and converting to image

LibreOffice Technology

Re-exporting to PDF

- Can reuse the original PDF contents inside the new
  - Preserves embedded fonts and styles
  - Preserves vector graphics

Collabora Online

Collabora Online

Thank you!

By Jaume Pujantell

@CollaboraOffice
hello@collaboraoffice.com
www.collaboraoffice.com