



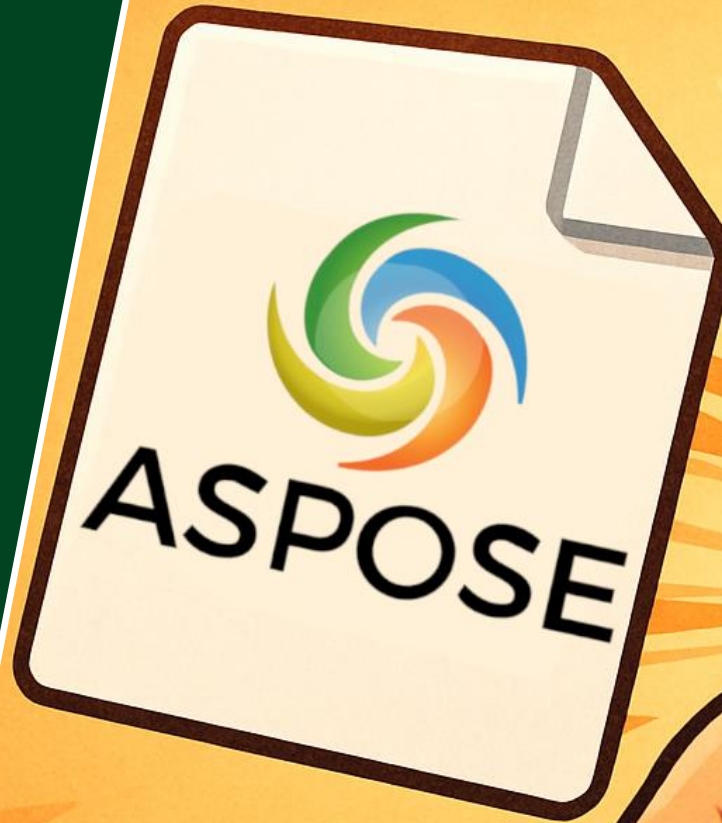
PhUSE EU Connect 2025

I Propose Aspose! Automating PDFs and Microsoft Office Files



Dominik Habel

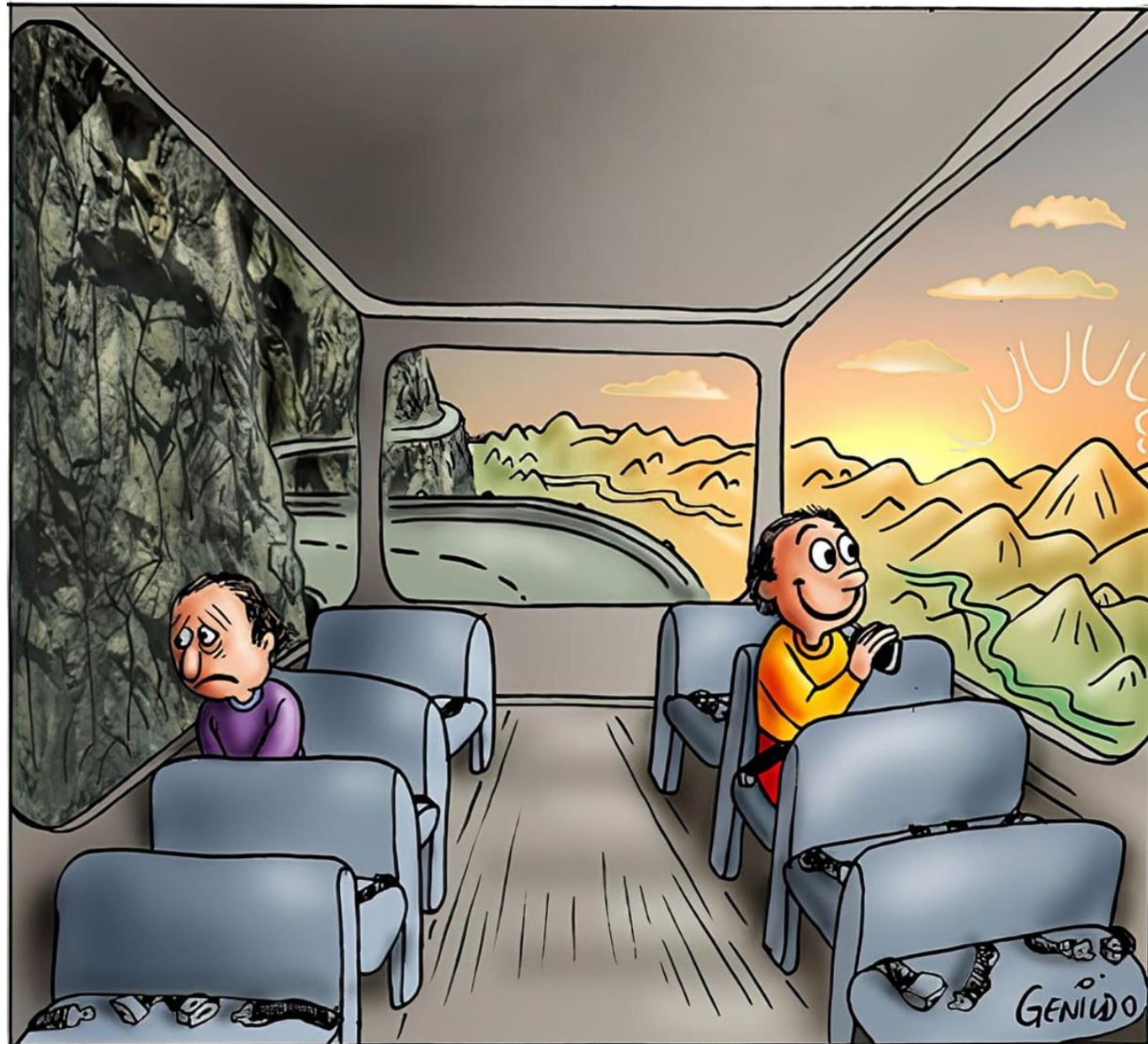
PhUSE Paper SM08 / November 2025

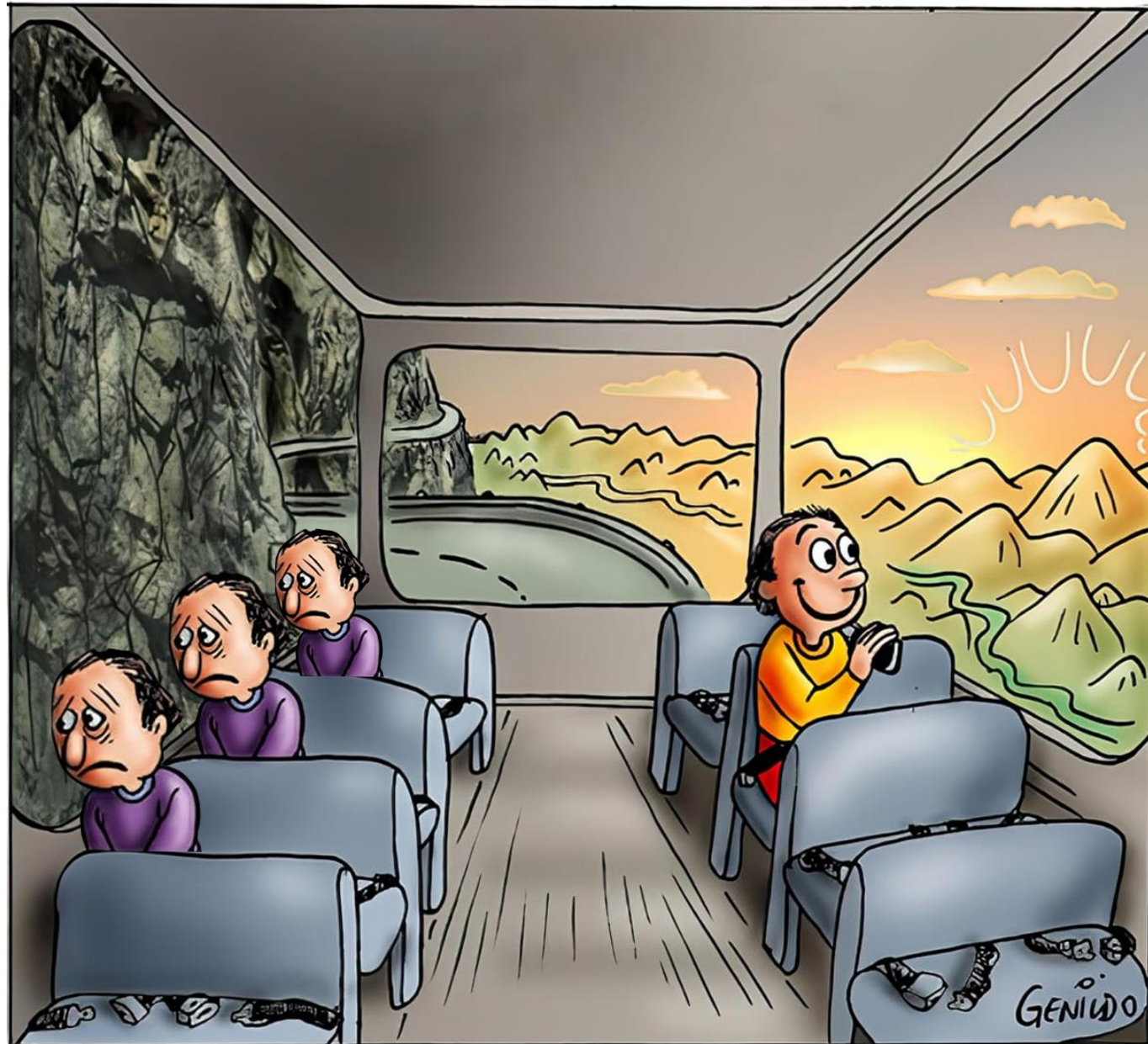




Co-Worker 16:29
Hey









BAD NEWS 1

It's pretty expensive!

Coming Soon to Open Source

Powerful Python libraries that will transform how you handle documents, images, and data



Aspose.Words for Python
Word Document Automation

Python document automation is about to get easier. This free library will handle mail merges, PDF conversions, watermarks, and more with proven tools.



Aspose.PDF for Python
PDF Management & Security

Unlike other PDF libraries, our upcoming open-source Python tools will use enterprise-grade technology. You'll create, edit, convert, secure, and sign PDFs.



Aspose.Cells for Python
Excel Spreadsheet Processing

Python's data power is about to meet open source Excel automation. You'll calculate formulas, build charts, transform formats, and manage workflows.



Aspose.Imaging for Python
Image Editing & Conversion

You'll easily change images using Python. One



Aspose.BarCode for Python
Barcode Generation and Recognition

This open-source Python library will read and



Aspose.Tasks for Python
Project and Task Management

Managing Microsoft Project files is about to get

BAD NEWS 2

I am not paid for this 😞



BAD NEWS 3

It is not a SAS or R tool...



GOOD NEWS

It is *amazing*!



Example 1

Word to PDF

```
%LET basepath = /var/swan/root/bhc/general/playground/gfeqr/dev/phuse;  
%LET docx = &basepath./word.docx;  
%LET output_path = &basepath./converted.pdf;  
  
PROC GROOVY;  
  ADD CLASSPATH = "&basepath./aspose-words-20.9-jdk17.jar";  
  SUBMIT "&docx." "&output_path.";  
  
  import com.aspose.words.Document;  
  import com.aspose.words.License;  
  
  License tmpLic = new License();  
  tmpLic.setLicense(Document.class.getClassLoader().getResourceAsStream("aspose.lic"));  
  
  String file = args[0];  
  String outputFileName = args[1];  
  
  Document docx = new Document(file);  
  
  docx.save(outputFileName);  
  
ENDSUBMIT;  
QUIT;
```

Proc Groovy
Boilerplate Code

Generic Java Code

Example 1

Word to PDF

```
List<FontSourceBase> tmpFontSrc = new ArrayList<>(Arrays.asList(FontSettings.getDefaultInstance().getFontsSources()));
tmpFontSrc.add(new FolderFontSource(fontPath, scanSubfolders: true));
FontSourceBase[] tmpNewFontSrc = tmpFontSrc.toArray(new FontSourceBase[0]);
FontSettings.getDefaultInstance().setFontsSources(tmpNewFontSrc);

Document docx = new Document(file);
docx.save(outputFileName);
```

Example 2

Extract Images from Word

```
Document docx = new Document(file);

NodeCollection<Shape> shapes = docx.getChildNodes(NodeType.SHAPE, isDeep: true);
int imageCount = 0;
for (Shape shape : shapes) {
    if (shape.hasImage()) {
        imageCount++;
        String imageFileName = outputPath + File.separator + "image_" + imageCount + ".png";
        shape.getImageData().save(imageFileName);
    }
}
```

Example 3

Update Header in Word

```
Document docx = new Document(file);

for (Section section : docx.getSections()) {
    for (HeaderFooter headerFooter : section.getHeadersFooters()) {
        Range headerRange = headerFooter.getRange();

        FindReplaceOptions options = new FindReplaceOptions();
        options.setMatchCase(true);
        options.setFindWholeWordsOnly(true);

        headerRange.replace(pattern: "12345", replacement: "ABCDE", options);
    }
}

docx.save(outputPath);
```

Example 4

Merge PDFs

```
Document pdf = new Document(file1);  
Document pdf2 = new Document(file2);  
  
pdf.getPages().add(pdf2.getPages());  
  
pdf.save(outputFileName);
```

Example 5

Extract Bookmarks from PDFs

```
void outputBookmarks() {  
    String file = "bookmarks.pdf";  
    Document pdf = new Document(file);  
  
    for (OutlineItemCollection outline : pdf.getOutlines()) {  
        extractBookmarkNames(outline);  
    }  
}  
  
void extractBookmarkNames(OutlineItemCollection outlines) {  
    for (OutlineItemCollection subOutline : outlines) {  
        System.out.println(subOutline.getTitle());  
  
        if (subOutline.size() > 0) {  
            extractBookmarkNames(subOutline);  
        }  
    }  
}
```

Summary

Aspose is extremely powerful!



Docx to PDF



Compare TFLs



Document
Style Check



Submission Ready
Document Generation



Question Time!

Thank you!

////////

Contact:

dhabel@mailbox.org

dominik.habel@bayer.com

