Change request log

1 Team

BugHunters
Members:
Prashant K Thakur
Sadaf Ghaffari
Bibek Raj Shrestha

2 Change Request

Change Request #1

The Split by size module of PDFsam allows to split a PDF file in files of a given size (see Figure 4). Ideally, the created files should be smaller than the given size, but this does not always happen (see Figure 5). You are requested to fix this functionality, so that the created files never exceed the specified size, unless the created file contains a single page that by itself exceeds the limit size.

3 Concept Location

The following is the detail on a concept location process for the change request "Enable/Disable scroll bars":

Step #	Description	Rationale
1	*We ran the system*	
2	*We looked for the module that handles split-by-size.*	
3	*From the class files in that module, we figured out that the module just handles the GUI and that in backend they are calling their open-sourced sejda engine.*	
4	*PDFsam project has added sedja engine as third party dependency for maven build.*	*We checked the maven dependency in parent pom.xml.*
5	*We downloaded the sedja project from Github.*	*Because it was open sourced by the same autho who wrote PDFsam.*
6	*We searched “split by size” in the whole project using Cltr+Shift+F.*	
7	*We found the sedja-sambox module that handles splits.* `(org.sejda.impl.sambox.component.split)`	
8	`Within that package (above), there are 6 class files of which AbstractPdfSplitter is the abstract class`	

	which is the skeletal implementation of the split execution.	
9	*It has a method split that does the actual split.*	*The comment gives us good idea about that.*
10	*The implementation is such that they opens a page, appends it and keep appending new pages until the size is above the user given value. The bug is they first append a page and then only checks if the total size of new pdf is larger than split size.*	
11	*We modified the code in line 93-127 such that now runtime append new page only if the split size is not overshooted.*	*Fixed the bug.*

Time spent (in minutes): 30

4 Impact Analysis

The following table summarizes the impact analysis we did for this change request.

Step #	Description	Rationale
1	*Since, we were working with abstract class, we checked if any subclasses overrides `AbstractPdfSplitter` class.*	*To track the classes that could be impacted by the change.*
2	*As no subclass was overriding the split method (as it was not asbtract itself), we could build the project and start testing.*	
3	*Since, PDFsam was using the sedja jar as dependency, we maven build sedja and exported only jar associated with sedja-sambox module to PDFsam project.*	*To check the version compatibility.*

Time spent (in minutes): 10

5 Actualization

The table below summarizes the actualization we performed to implement this change.

Step #	Description	Rationale
1	*As we had pin-pointed the logic in split method in `AbstractPdfSplitter class, we reversed the logic to fix the bug.`*	*We had to move when and how a page is appended to create a new pdf split.*
2	*We added some print statements to check the logical flow of program at different split-size.*	*To verify if the reshuffle that we did, didn't have any corner cases.*

Time spent (in minutes): 10

6 Validation

The table below summarizes the validation phase of the change request.

Step #	Description	Rationale
1	*We created sedja binary before we exported the sedja-sambox jar to PDFsam. Sedja provides sedja-console module that we used to test our changes in command-line.*	*The test passed.*
2	*We verified our change by downloading pre-build sedja-console binary and running same test.*	*The test passed. The default sedja-console returns some pdfs that overshoots the split-size.*
3	*Finally, in PDFsam project, we imported our custom sedja-sambox jar and removed the corresponding maven dependency from the pom.xml file. We build the project and verified its compatibility.*	*Same test verified. This time using PDFsam's GUI.*

Time spent (in minutes): 15

7 Timing

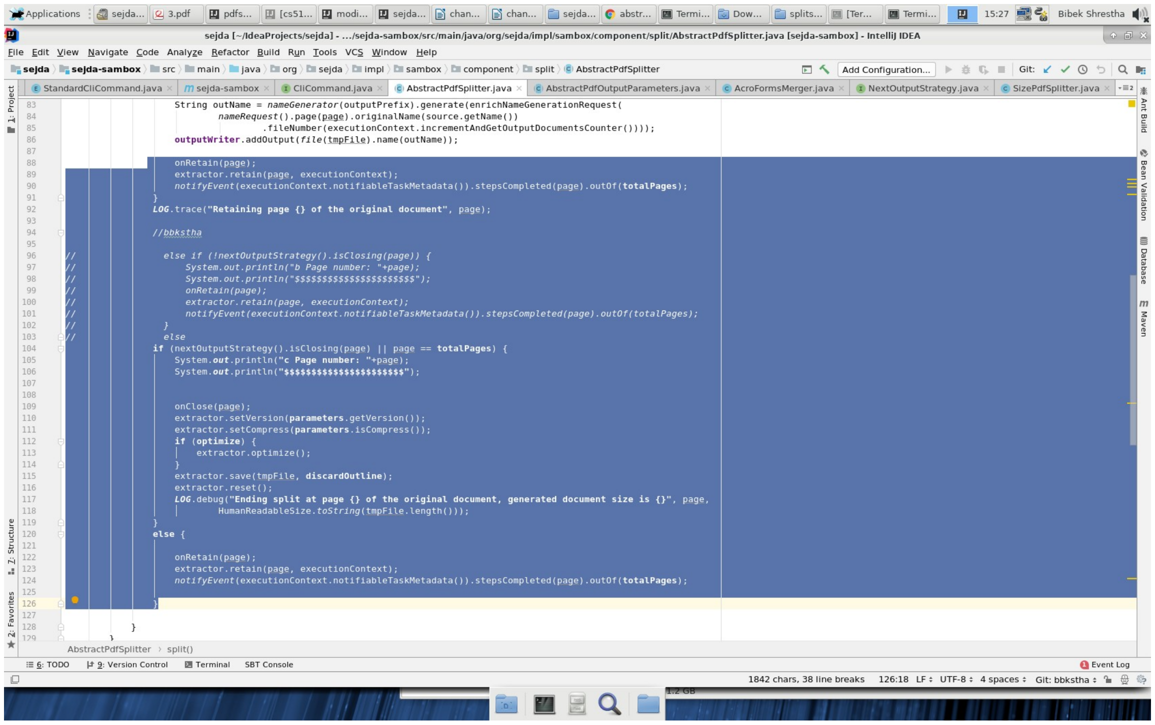
Summarize the time spent on each phase.

Phase Name	Time (in minutes)
Concept location	40
Impact Analysis	10
Prefactoring	-
Actualization	10
Postfactoring	-
Verification	15
Total	75

9 Conclusions

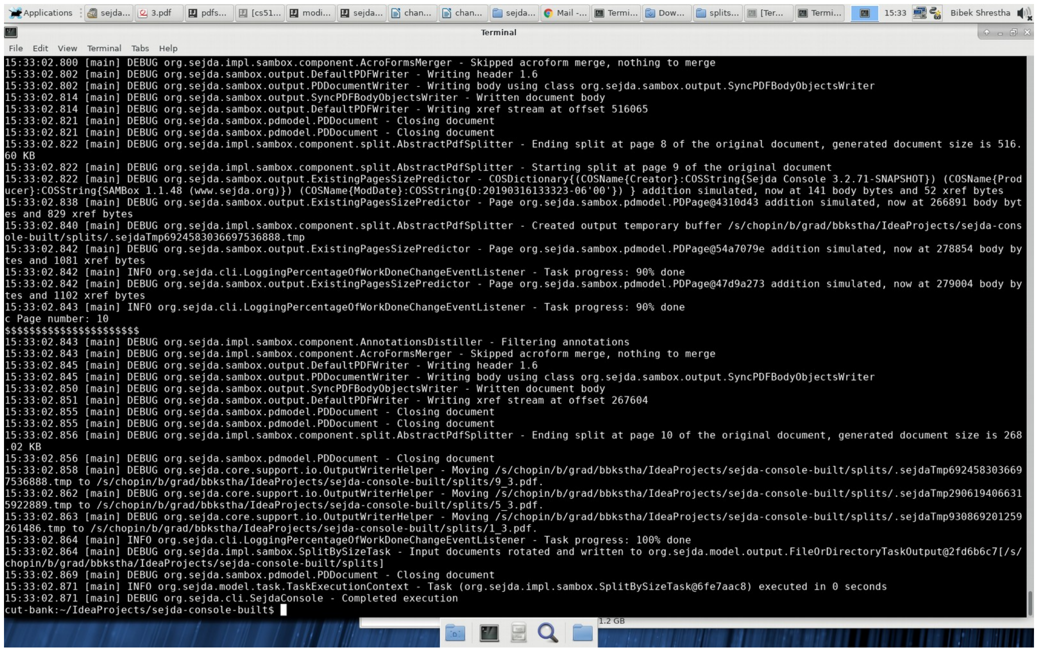
GitHub Repository: https://github.com/bbkstha/sejda-modified

Some screenshots of changes made.

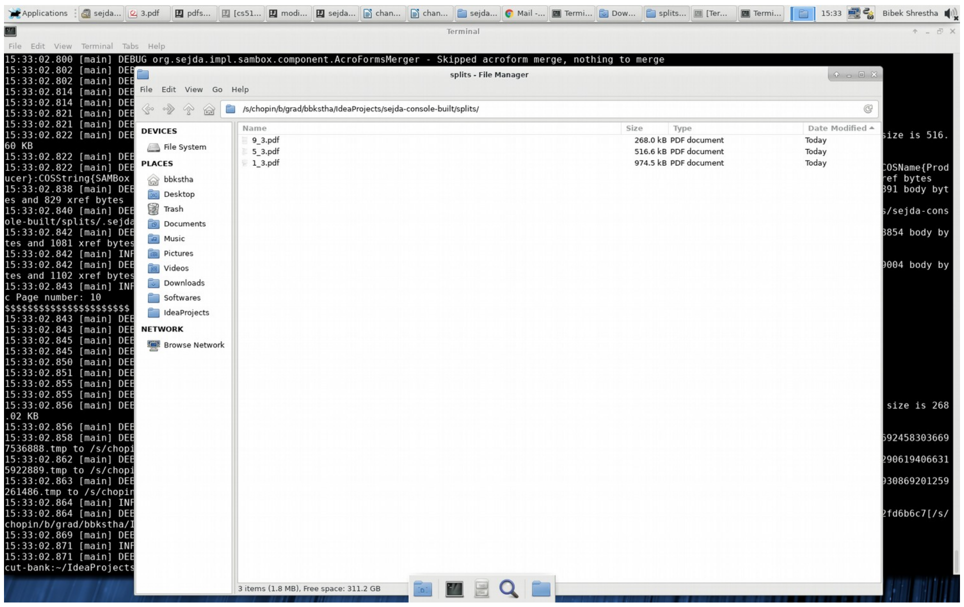


```java
                String outName = nameGenerator(outputPrefix).generate(enrichNameGenerationRequest(
                        nameRequest().page(page).originalName(source.getName())
                                .fileNumber(executionContext.incrementAndGetOutputDocumentsCounter())));
                outputWriter.addOutput(file(tmpFile).name(outName));

                onRetain(page);
                extractor.retain(page, executionContext);
                notifyEvent(executionContext.notifiableTaskMetadata()).stepsCompleted(page).outOf(totalPages);
            }
            LOG.trace("Retaining page {} of the original document", page);

            //bbkstha

//            else if (!nextOutputStrategy().isClosing(page)) {
//                System.out.println("b Page number: "+page);
//                System.out.println("$$$$$$$$$$$$$$$$$$$$$$");
//                onRetain(page);
//                extractor.retain(page, executionContext);
//                notifyEvent(executionContext.notifiableTaskMetadata()).stepsCompleted(page).outOf(totalPages);
//            }
//            else
            if (nextOutputStrategy().isClosing(page) || page == totalPages) {
                System.out.println("c Page number: "+page);
                System.out.println("$$$$$$$$$$$$$$$$$$$$$$");


                onClose(page);
                extractor.setVersion(parameters.getVersion());
                extractor.setCompress(parameters.isCompress());
                if (optimize) {
                    extractor.optimize();
                }
                extractor.save(tmpFile, discardOutline);
                extractor.reset();
                LOG.debug("Ending split at page {} of the original document, generated document size is {}", page,
                        HumanReadableSize.toString(tmpFile.length()));
            }
            else {

                onRetain(page);
                extractor.retain(page, executionContext);
                notifyEvent(executionContext.notifiableTaskMetadata()).stepsCompleted(page).outOf(totalPages);

            }

        }
    }
```

Figure: Changed code in AbstractPdfSplitter.



```
15:33:02.800 [main] DEBUG org.sejda.impl.sambox.component.AcroFormsMerger - Skipped acroform merge, nothing to merge
15:33:02.802 [main] DEBUG org.sejda.sambox.output.DefaultPDFWriter - Writing header 1.6
15:33:02.802 [main] DEBUG org.sejda.sambox.output.PDDocumentWriter - Writing body using class org.sejda.sambox.output.SyncPDFBodyObjectsWriter
15:33:02.814 [main] DEBUG org.sejda.sambox.output.SyncPDFBodyObjectsWriter - Written document body
15:33:02.814 [main] DEBUG org.sejda.sambox.output.DefaultPDFWriter - Writing xref stream at offset 516065
15:33:02.821 [main] DEBUG org.sejda.sambox.pdmodel.PDDocument - Closing document
15:33:02.821 [main] DEBUG org.sejda.sambox.pdmodel.PDDocument - Closing document
15:33:02.822 [main] DEBUG org.sejda.impl.sambox.component.split.AbstractPdfSplitter - Ending split at page 8 of the original document, generated document size is 516.
60 KB
15:33:02.822 [main] DEBUG org.sejda.impl.sambox.component.split.AbstractPdfSplitter - Starting split at page 9 of the original document
15:33:02.822 [main] DEBUG org.sejda.sambox.output.ExistingPagesSizePredictor - COSDictionary{(COSName{Creator}:COSString{Sejda Console 3.2.71-SNAPSHOT}) (COSName{Prod
ucer}:COSString{SAMBox 1.1.48 (www.sejda.org)}) (COSName{ModDate}:COSString{D:20190316133323-06'00'}) } addition simulated, now at 141 body bytes and 52 xref bytes
15:33:02.838 [main] DEBUG org.sejda.sambox.output.ExistingPagesSizePredictor - Page org.sejda.sambox.pdmodel.PDPage@4310d43 addition simulated, now at 266891 body byt
es and 829 xref bytes
15:33:02.840 [main] DEBUG org.sejda.impl.sambox.component.split.AbstractPdfSplitter - Created output temporary buffer /s/chopin/b/grad/bbkstha/IdeaProjects/sejda-cons
ole-built/splits/.sejdaTmp6924583036697536888.tmp
15:33:02.842 [main] DEBUG org.sejda.sambox.output.ExistingPagesSizePredictor - Page org.sejda.sambox.pdmodel.PDPage@54a7079e addition simulated, now at 278854 body by
tes and 1081 xref bytes
15:33:02.842 [main] INFO org.sejda.cli.LoggingPercentageOfWorkDoneChangeEventListener - Task progress: 90% done
15:33:02.842 [main] DEBUG org.sejda.sambox.output.ExistingPagesSizePredictor - Page org.sejda.sambox.pdmodel.PDPage@47d9a273 addition simulated, now at 279004 body by
tes and 1102 xref bytes
15:33:02.843 [main] INFO org.sejda.cli.LoggingPercentageOfWorkDoneChangeEventListener - Task progress: 90% done
c Page number: 10
$$$$$$$$$$$$$$$$$$$$$$
15:33:02.843 [main] DEBUG org.sejda.impl.sambox.component.AnnotationsDistiller - Filtering annotations
15:33:02.843 [main] DEBUG org.sejda.impl.sambox.component.AcroFormsMerger - Skipped acroform merge, nothing to merge
15:33:02.845 [main] DEBUG org.sejda.sambox.output.DefaultPDFWriter - Writing header 1.6
15:33:02.845 [main] DEBUG org.sejda.sambox.output.PDDocumentWriter - Writing body using class org.sejda.sambox.output.SyncPDFBodyObjectsWriter
15:33:02.850 [main] DEBUG org.sejda.sambox.output.SyncPDFBodyObjectsWriter - Written document body
15:33:02.851 [main] DEBUG org.sejda.sambox.output.DefaultPDFWriter - Writing xref stream at offset 267604
15:33:02.855 [main] DEBUG org.sejda.sambox.pdmodel.PDDocument - Closing document
15:33:02.855 [main] DEBUG org.sejda.sambox.pdmodel.PDDocument - Closing document
15:33:02.856 [main] DEBUG org.sejda.impl.sambox.component.split.AbstractPdfSplitter - Ending split at page 10 of the original document, generated document size is 268
.02 KB
15:33:02.856 [main] DEBUG org.sejda.sambox.pdmodel.PDDocument - Closing document
15:33:02.858 [main] DEBUG org.sejda.core.support.io.OutputWriterHelper - Moving /s/chopin/b/grad/bbkstha/IdeaProjects/sejda-console-built/splits/.sejdaTmp692458303669
7536888.tmp to /s/chopin/b/grad/bbkstha/IdeaProjects/sejda-console-built/splits/9_3.pdf.
15:33:02.862 [main] DEBUG org.sejda.core.support.io.OutputWriterHelper - Moving /s/chopin/b/grad/bbkstha/IdeaProjects/sejda-console-built/splits/.sejdaTmp290619406631
5922889.tmp to /s/chopin/b/grad/bbkstha/IdeaProjects/sejda-console-built/splits/5_3.pdf.
15:33:02.863 [main] DEBUG org.sejda.core.support.io.OutputWriterHelper - Moving /s/chopin/b/grad/bbkstha/IdeaProjects/sejda-console-built/splits/.sejdaTmp930869201259
261486.tmp to /s/chopin/b/grad/bbkstha/IdeaProjects/sejda-console-built/splits/1_3.pdf.
15:33:02.864 [main] INFO org.sejda.cli.LoggingPercentageOfWorkDoneChangeEventListener - Task progress: 100% done
15:33:02.864 [main] DEBUG org.sejda.impl.sambox.SplitBySizeTask - Input documents rotated and written to org.sejda.model.output.FileOrDirectoryTaskOutput@2fd6b6c7[/s/
chopin/b/grad/bbkstha/IdeaProjects/sejda-console-built/splits]
15:33:02.869 [main] DEBUG org.sejda.sambox.pdmodel.PDDocument - Closing document
15:33:02.871 [main] INFO org.sejda.model.task.TaskExecutionContext - Task (org.sejda.impl.sambox.SplitBySizeTask@6fe7aac8) executed in 0 seconds
15:33:02.871 [main] DEBUG org.sejda.cli.SejdaConsole - Completed execution
cut-bank:~/IdeaProjects/sejda-console-built$
```



```
15:33:02.800 [main] DEBUG org.sejda.impl.sambox.component.AcroFormsMerger - Skipped acroform merge, nothing to merge
15:33:02.802 [main] DEB
15:33:02.802 [main] DEB
15:33:02.814 [main] DEB
15:33:02.814 [main] DEB
15:33:02.821 [main] DEB
15:33:02.821 [main] DEB
15:33:02.822 [main] DEB                                                                                   size is 516.
60 KB
15:33:02.822 [main] DEB
15:33:02.822 [main] DEB                                                                                   COSName{Prod
ucer}:COSString{SAMBox                                                                                    ref bytes
15:33:02.838 [main] DEB                                                                                   391 body byt
es and 829 xref bytes
15:33:02.840 [main] DEB                                                                                   s/sejda-cons
ole-built/splits/.sejda
15:33:02.842 [main] DEB                                                                                   3854 body by
tes and 1081 xref bytes
15:33:02.842 [main] INF
15:33:02.842 [main] DEB                                                                                   9004 body by
tes and 1102 xref bytes
15:33:02.843 [main] INF
c Page number: 10
$$$$$$$$$$$$$$$$$$$$$$
15:33:02.843 [main] DEB
15:33:02.843 [main] DEB
15:33:02.845 [main] DEB
15:33:02.845 [main] DEB
15:33:02.850 [main] DEB
15:33:02.851 [main] DEB
15:33:02.855 [main] DEB
15:33:02.855 [main] DEB
15:33:02.856 [main] DEB                                                                                    size is 268
.02 KB
15:33:02.856 [main] DEB
15:33:02.858 [main] DEB                                                                                   692458303669
7536888.tmp to /s/chopi
15:33:02.862 [main] DEB                                                                                   290619406631
5922889.tmp to /s/chopi
15:33:02.863 [main] DEB                                                                                   930869201259
261486.tmp to /s/chopir
15:33:02.864 [main] INF
15:33:02.864 [main] DEB                                                                                   2fd6b6c7[/s/
chopin/b/grad/bbkstha/I
15:33:02.869 [main] DEB
15:33:02.871 [main] INF
15:33:02.871 [main] DEB
cut-bank:~/IdeaProjects
```

/s/chopin/b/grad/bbkstha/IdeaProjects/sejda-console-built/splits/

Name	Size	Type	Date Modified
9_3.pdf	268.0 kB	PDF document	Today
5_3.pdf	516.6 kB	PDF document	Today
1_3.pdf	974.5 kB	PDF document	Today

3 items (1.8 MB), Free space: 311.2 GB