

## 13.7 Extract text

A special application is the extraction of pure (unformatted) text data from PDF documents. Therefore, PDFBox provides the so-called stripper class.

First the PDF document has to be loaded again:

```
PDDocument document = PDDocument.load("test.pdf");
```

Then an object of class PDFTextStripper has to be generated.

```
PDFTextStripper stripper = new PDFTextStripper();
```

With this stripper object and its method `getText()` the PDF document's text can be parsed into a String:

```
String text= stripper.getText(doc);
```