

PDF

Extracting text from a PDF document

In the event that you are going to index the content of a PDF, a good place to look first is a Java library called PDFBox <http://pdfbox.apache.org/cookbook/textextraction.html>