




Java Extraction [HOME](#)  [GUIDES](#)  [JAVA](#)  [EXTRACTION](#)

# Extract text from PDFs using Java



COPY PAGE



Extracting text from a PDF can be a complex task. For our libraries products, we distilled the text extraction process to a single function that provides all the information required to achieve whatever a user may require to accomplish any text processing task.

Text can be extracted on a per-page basis using the library page API. Each page's text is split into individual text lines as `TextBlock` objects that contain the string and page coordinates for each line. The following code shows how you might concatenate all the text on the first page of a document and print it to the console:

```
PdfDocument document = PdfDocument.open(new FileDataProvider(new File("documentWithText.pdf")));
List<TextBlock> textLines = document.getPage(0).getTextLines();
StringBuilder pageText = new StringBuilder();
for (TextBlock textLine : textLines) {
    pageText.append(textLine.getText());
    pageText.append("\r\n");
}
System.out.println("The text on the first page reads:\n" + pageText);
```

The location and size of each line in the form of a `Rect` can also be read from each page text lines:



ASK AI

```
PdfDocument document = PdfDocument.open(new FileDataProvider(new File("documentWithText.pdf")));
List<TextBlock> textLines = document.getPage(0).getTextLines();
Rect firstLineRect = textLines.get(0).getRect();
```

See our API reference for more specifics on `getTextLines`.

---

## Was this helpful?

✓ YES

✗ NO



Copyright 2025 Nutrient. All rights reserved.