

<https://github.com/huilam/hl-doc-extractor-pdf>

**PDF-TEST-03:** 1-Page-Portrait-Text-Image

# hl-doc-extractor-pdf

A simple lightweight (250 KB on disk) Java PDF extractor based on PDFBox.

Sample Code:

```
/** Initialize PDF Extract with a file
PDFExtractor extractor = new PDFExtractor(new File("file.pdf"));
extractor.setStartPageNo(0);
extractor.setEndPageNo(0);

/** Extract all selected pages
ExtractedContent data = extractor.extractAll();

/** Render page layout as BufferedImage
MetaData meta = data.getMetaData();
List page1Data = aExtractData.getContentItemListByPageNo(1);
BufferedImage imgLayout = ContentUtil.renderPageLayout(
    meta.getPageWidth(), meta.getPageHeight(),
    Color.WHITE, // background color
    false, // render text
    page1Data); // page's extracted item
```

