

PDFBox - PDF Metadata

Java PDF Library, XMP, PDF metadata, pdfbox

Table of contents

1 Accessing Metadata.....	2
---------------------------	---

1. Accessing Metadata

See class:[org.apache.pdfbox.pdmodel.common.PDMetadata](#)

See example:[AddMetadataFromDocInfo](#)

See Adobe Documentation:[XMP Specification](#)

PDF documents can have XML metadata associated with certain objects within a PDF document. For example, the following PD Model objects have the ability to contain metadata:

- PDDocumentCatalog
- PDPage
- PDXObject
- PDICCBased
- PDStream

The metadata that is stored in PDF objects conforms to the XMP specification, it is recommended that you review that specification. Currently there is no high level API for managing the XML metadata, PDFBox uses standard java InputStream/OutputStream to retrieve or set the XML metadata. For example:

```
PDDocument doc = PDDocument.load( ... );
PDDocumentCatalog catalog = doc.getDocumentCatalog();
PDMetadata metadata = catalog.getMetadata();

//to read the XML metadata
InputStream xmlInputStream = metadata.createInputStream();

//or to write new XML metadata
InputStream newXMPData = ...;
PDMetadata newMetadata = new PDMetadata(doc, newXMPData, false );
catalog.setMetadata( newMetadata );
```