J. Plasma Fusion Res. Vol.93, No.8 (2017) 387-390

サロン

多数のPDF文書を自動処理によりまとめ上げて一つのPDF版プロシーディングス論文集を作る方法

A Programming Method to Make a PDF Version of Proceedings Book by Combining Multiple PDF Documents

佐々木 明[1,3)]，村上 泉[2,3)]
SASAKI Akira[1,3)] and MURAKAMI Izumi[2,3)]
[1)]量子科学技術研究開発機構関西光科学研究所，[2)]核融合科学研究所，
[3)]特定非営利活動法人原子分子データ応用フォーラム

（原稿受付：2017年5月15日）

PDF文書の連結，編集をプログラムから簡単にできるJava iTextライブラリについて紹介します．研究会などを開くときに，著者から収集したPDF文書から，迅速にアブストラクト集やプロシーディングス論文集の冊子を編集および発行し，成果の普及やコミュニティの形成に役立てる方法を紹介します．

Keywords:
PDF, Java, iText library, DTP, IT

1．はじめに

研究会などの会合は日々の研究の中で重要な活動ですが，良い議論をするには，時間の制約がある中で，お互いに発表される研究内容，研究手法を良く理解することが必要です．そのためには研究会の主催者の仕事として，プログラムやアブストラクト集を正確でタイムリーに発行することが重要です．研究会の開催の前に他の参加者の発表内容を理解できると，自分の発表を研究会の趣旨に合わせてより良いものにするなど，モチベーションが高める効果もあります．

研究会のアブストラクトなどの資料は，PDF（Portable Document Format）ファイルでやりとりすることが普通になりました．資料を一冊の冊子にまとめたものがあると良いですが，できるだけ最新の結果を発表しようとする気持ちのために，投稿が研究会の直前になったり，原稿の頻繁な差し替えが必要になる傾向があります．AdobeAcobatなどを用いた手動での冊子の作成作業は決して難しくはありませんが，原稿の追加や変更のたびごとに作業の繰り返しが必要になると，もっと効率的な方法の必要性が感じられると思います．本サロンでは，PDF文書から冊子を作成するための便利な道具として，Java言語のiTextライブラリについて紹介します．

ここで紹介する内容は，特定非営利法人原子分子データ応用フォーラムにおいて原子分子データのニーズとシーズのマッチングの実現のために行ってきた活動の一部で，著者（佐々木）が核融合研共同研究「原子分子データ，原子分子モデルの開発，検証，利用のための技術基盤とネットワークの構築」の支援を受けて検討してきたIT技術に関するものです．そして，2016年12月に行われた核融合科学研究所素過程研究会との合同研究会[1]および，第15回X線レーザー国際会議（The 15th International Conference on X-Ray Lasers）[2]において，講演プログラムの編成，アブストラクト集，プロシーディングス論文集の編集，作成に活用した技術をまとめたものです．

2．iTextライブラリによる冊子の作成

iTextライブラリ[3]を使うと，Java言語を使って，プログラムからPDFを新たに作成したり，既存のPDFを操作することができます．AdobeAcrobatのようなGUI（Graphical User Interface）ツールを使わず，端末からコマンドを入力する処理，あるいは以前の言葉で言うバッチ処理で，PDFを生成するソフトウエアを開発することができます．

iTextの現在ダウンロード可能なバージョンには4，5，7がありますが，バージョン4[5]がLGPL（Lesser General Public License），MPL（Mozilla Public License）ライセンスに従うフリーのライブラリであるのに対し，バージョン5[6]はAGPL（Affero General Public License）ライセンスに従い，バージョン7は原則有償とされていて，新しいバージョンほど使用条件が厳しくなっています．ライセンスの取り扱い方法についてはインターネット上の情報[7]等も参照してください．本技術ノートでは参考文献[3]で用いられているバージョン5を想定して説明します．

iTextライブラリを利用するプログラムを作成するには，まず適当なバージョンの.jarファイルをダウンロード

Kansai Photon Science Institute, National Institute for Quantum and Radiological Science and Technology, Kizugawa, KYOTO 619-0215, Japan
author's e-mail: sasaki.akira@qst.go.jp

```
import java.io.*;
import java.io.nio.file.*;
import com.itextpdf.text.*;
import com.itextpdf.text.Font.*;
import com.itextpdf.text.pdf.*;
...
Document document=new Document();
PdfCopy copy=new PdfCopy(document,new FileOutputStream(outputfile));
document.open();
for (String inputfile: files){
        if (Files.exists(Paths.get(inputfile))){
                reader=new PdfReader(inputfile);
                int numberOfPages=reader.getNumberOfPages();
                for (int i=1; i<=numberOfPages; i++){
                        page=copy.getImportedPage(reader,i);
                        copy.addPage(page);
                }
                reader.close();
        }
}
copy.close();
document.close();
```

図1　iText ライブラリの PdfCopy クラスを用いて，pdf ファイルを連結するプログラム例.

したのちに，ソースプログラム内でimport 文によってAPI（Application Programing Interface）の使用を宣言すること，コンパイル時にライブラリの．jar ファイルをリンクする必要があります．Eclipse統合開発環境を利用しているときは，projectのpropertiesの設定で．jarファイルをlibraryに追加します．C，C++，Python などの言語にも PDF ライブラリがありますが[4]，オブジェクト指向言語を使うと，プログラム本体とライブラリの処理の干渉がないことや，動作のために必要なパラメータがデフォルトで適切な値に設定されるなど，ライブラリの機能を簡単に利用することができます．なお，以降の説明では，プログラムのことをクラスと呼び，関数，サブルーチンのことをメソッドと呼ぶなど，オブジェクト指向言語の用語を使うことにします．

2.1　PDF ファイルの連結

ここでは，iText ライブラリを利用して，複数の PDF ファイルを連結して，ひとつの PDF ファイル，冊子を作成する方法を紹介します．PDF は表示，印刷のためのフォーマットであり，編集を行うには適していません．一方，ワードプロセッサで複数のファイルを連結しようとすると，図のわずかな移動や，たった 1 文字の挿入，削除によって，リフローが発生し，ページ区切りが移動し，表示や印刷の結果が大きく変わってことがあります．iText ライブラリには，既存の PDF の内容，レイアウトには影響を与えることなく処理を行うという，プロシーディングス論文集などの作成に適した機能が備わっています．

PdfCopy は，プログラムで読み込んだPDFをコピーして新しい PDF を作成するために使うクラスです．**図 1** は，files というリストに格納されているファイル名を持つファイルを一つずつ読み込み，outputfile で指定される新しい PDF にコピーして，ひとつにまとめて出力するプログラム例です．inputfile で指定されている PDF ファイルの内容を，PdfReader クラスを使って読み込んだあと，getImportPage メソッドで 1 ページずつ取り出し，addPage メソッドで新しい PDF ファイルへと書き込む操作を示しています．getNumberOfPages は，文書のページ数を返すメソッドで，目次の作成にも役立ちます．

Files.exists（Paths.get（inputfile））メソッドは，inputfile で指定されたファイルがファイルシステム上に存在するときに，真を返します．このあとのファイルを連結する処理は，ファイルが存在したときだけ実行されます．Files クラスはJava言語の機能として提供されている，ファイル操作，ディレクトリの作成，ファイルのコピー，移動などの機能を持つクラスです[8]．これらの機能は，メールなどとして受理したあと，所定のディレクトリに保存されているファイルを調べて，正しいPDF文書を冊子の中に取り込んだり，原稿がまだ投稿されていないことを検出して著者に知らせるときにも役立ちます．それまで人が行っていたチェックをひとつずつプログラムに肩代わりさせることができるようになり，結果的に人の負担を減らしながら作業を迅速に進めるために役立つと考えられます．

2.2　ページ番号の付与

新しく資料を作成したとき，しばしば通しのページ番号をふることが必要になります．PdfStamper クラスは，既存

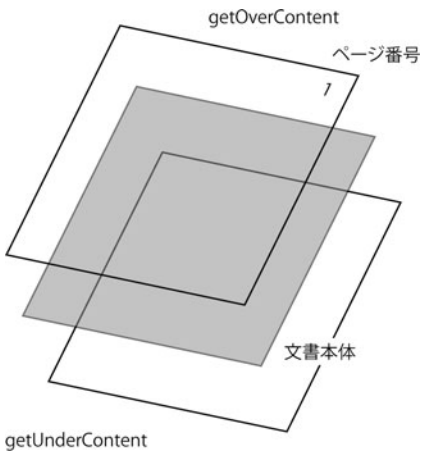

図2　iText ライブラリの PdfStamper クラスで考えられている pdf 文書の構造.

の PDF に，スタンプを押すようにヘッダ，フッタをつけ，ページ番号などを記載するため用いられます．**図 2** に示すように，iText では PDF はあたかも文書や図が描かれた透明なレイヤーが重なって構成されていると考えます．ページ番号は，文書のレイヤーの上に新しいレイヤーを作り，その上に記載すれば良いと考えられます．

図 3 はその処理を PdfStamper クラスを使って行うプログラム例を示します．まず PdfReader クラスを使って，PdfStamper クラスに PDF を読み込みます．次に getOverContent メソッドを使い，それぞれのページの上に，PdfContentByte canvas によって指定される新しい透明なレイヤーを作ります．そして，ColumnText.showTextAligned メソッドを使って，そのレイヤーの上に，絶対座標，フォントを指定して，ページ番号を書き込みます．この例では，ページ番号が見開きページの左右の端に記載されるよう，奇数ページではページの右端，偶数ページでは左端に記載することを示しています．この方法を使うと，既存の PDF に上書きされる形でページ番号が記載されることになるので，著者に原稿の執筆をお願いする際に，ページ番号と本文が重ならないように，マージンを適切に設定したテンプレートの利用をお願いすることが必要です．

3．まとめ

PDF ファイルを連結して冊子を作る作業は，mediabb.sty パッケージを読み込めば，**図 4** のように includegraphics コマンドで LaTeX でもできます．ただし，この方法はページの中に図を貼り込むという考え方なので，PDF が複数のページを持つ場合には正常に動作しませんし，LaTeX のレイアウトは内部のテキストや図の配置によってリフローされるので，正確なページ番号を把握することが困難です．

iText ライブラリを利用すると，PDF 文書の編集や，ホームページ作成における人手による作業を一つずつ減らすことができて，本来の研究開発の生産性の向上に役立つと考えています．処理の性能という点では，ここで示した PDF 冊子の作成では，合計約70件，400ページの論文集（容量約 100 MB）を標準的な PC で 1 分以内で作成することができました．

IT 技術の進歩によって，データベースと連携する web アプリケーションを用いて学術的な会合のロジスティクスを一元管理することも可能になりましたが，情報セキュリティに関することを含めて利用するために必要な技術は高

```
import java.io.*;
import java.text.DecimalFormat;
import com.itextpdf.text.*;
import com.itextpdf.text.Font.*;
import com.itextpdf.text.pdf.*;
...
PdfReader reader=new PdfReader( inputfile);
PdfStamper stamper=new PdfStamper(reader, new FileOutputStream(outputfile));
for (int i=1; i<=reader.getNumberOfPages(); i++){
        PdfContentByte canvas=stamper.getOverContent(i);
        int page=paper.getStartpage()+i-1;
        String str=new DecimalFormat("0").format(page);
        if (i%2==1) {
                ColumnText.showTextAligned(canvas, Element.ALIGN_RIGHT,
                        new Phrase(str,new Font(FontFamily.TIMES_ROMAN, 11)),470,730,0);
        }
        else {
                ColumnText.showTextAligned(canvas, Element.ALIGN_LEFT,
                        new Phrase(str,new Font(FontFamily.TIMES_ROMAN, 11)),125,730,0);
        }
}
reader.close();
stamper.close();
```

図 3　iText ライブラリの PdfStamper クラスを用いて，文書のヘッダにページ番号を記入するプログラム例．

```
\usepackage{mediabb}
...

\begin{tabularx}{18cm}{l}
        \fbox{\large label}\\
        \begin{minipage}{18cm}
                \includegraphics[width=17cm, trim=60 50 40 40]{filename.pdf}
        \end{minipage}
\end{tabularx}
```

図 4　LaTeX で PDF を図として取り込む例．

度になり，研究者からはしきいが高くなってしまったように思います．しかし，主催者の手元でのデータ処理や，静的なwebページの作成など，比較的簡単で生産性を高めるために役立つIT技術には，他にもさまざまな可能性があるのではないかと思います．

ここで紹介したプログラムのサンプルをwebサイトhttp://www.am-data-forum.com/pdf_example.html に公開しています．

謝　辞

本サロンに紹介した内容は核融合科学研究所一般共同研究「原子分子データ，原子分子モデルの開発，検証，利用のための技術基盤とネットワークの構築」（NIFS16KBAF024）の支援を受けて行われました．また，量子科学技術研究開発機構，関西光科学研究所，河内哲哉所長，光量子科学研究部，近藤公伯部長，X線レーザー研究グループ，錦野将元グループリーダーの支援に感謝します．

参 考 文 献

［1］http://www.am-data-forum.com/seminar28/seminar28program.html
［2］http://www.kansai.qst.go.jp/icxrl2016/
［3］B. Lwagie, "iText IN ACTION" SECOND EDITION, Manning Publications Co. 2011.
［4］https://texwiki.texjp.org/?PDF
［5］http://mvnrepository.com/artifact/com.lowagie/itext/4.2.1
［6］http://developers.itextpdf.com/apis
［7］http://qiita.com/toshi71/items/bc05d6e15edd645c8f46
［8］http://docs.oracle.com/javase/jp/7/api/java/nio/file/Files.html