

KDrasterコンバート RSTX3PST_OCR

ラスター+テキスト(OCR)/PostScript(PDF)変換プログラム



Kernel Computer System
カーネルコンピュータシステム株式会社

本社：パッケージ販売部
〒221-0056

横浜市神奈川区金港町 6-3 横浜金港町ビル
TEL：045-442-0500 FAX：045-442-0501
URL：<http://www.kernelcomputer.co.jp>

製品概要

- ・ラスターデータを PostScript, PDF データに変換する。
- ・ラスターデータ+テキストを合成し、PostScript, PDF データに変換する。
- ・ラスターデータを OCR し、PostScript, PDF データに変換する。
 - ・出力方法は下記の 3 種類から選択することが可能です。
 - ・ OCR したテキストと原図を重ねて
 - ・ OCR したテキストのみ
 - ・ OCR したテキストと原図が 2 ページ分けて
- ・ Windows 版では、Windows フォントを指定し、テキストと合成することが可能です。
- ・ UNIX/Linux 版では TTF 形式のフォントファイルを指定し、テキストと合成することが可能です。
- ・ 読み込み可能なラスターデータは次のとおりです。

・ BMP	・ CALSG4	・ CCRF	・ EDMICS	・ FAX
・ FORMTEK	・ FX_RASTER	・ GIF	・ GTX_G4	・ IOCA
・ JPEG	・ MH	・ MIEL	・ MMR	・ MR
・ NS-G4	・ PCX	・ PNG	・ PNM	・ SUN_RASTER
・ TIFF	・ TOSFILE	・ WBMP	・ XBITMAP	・ XWD

- ・ 出力可能なフォーマット

・ PostScript	レベル 1, レベル 2, レベル 2 (G4), EPSF, Acrobat 用 PostScript
・ PDF	1.4, 1.5, 1.6, 1.7, 1.7EX3

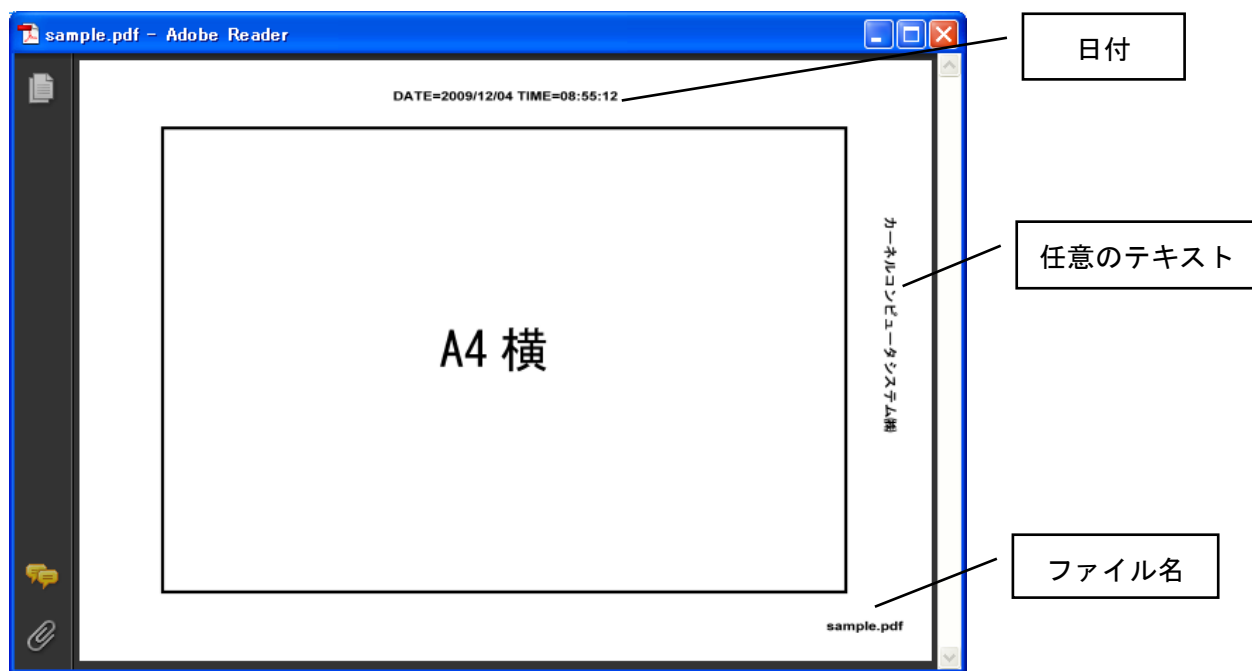
動作環境：

Windows XP (SP3 以降) (32bit)
Windows Vista (32bit)
Windows 7 (32bit/64bit)
Windows 8 (32bit/64bit)
Windows 8.1 (32bit/64bit)
Windows 10 (32bit/64bit)
Windows Server 2003 (32bit)
Windows Server 2008 (32bit)
Windows Server 2008 R2
Windows Server 2012
Windows Server 2008 R2
Windows Server 2016

価格：

40 万円 (税抜き)

合成サンプル :



1. 基本操作

- ・ラスターデータを PS、PDF データに変換するには、次のような形式でコマンドラインに入力します。

```
実行モジュール名  入力ファイル名  -0 出力ファイル名  -各オプション
```

- ・変換実行例：

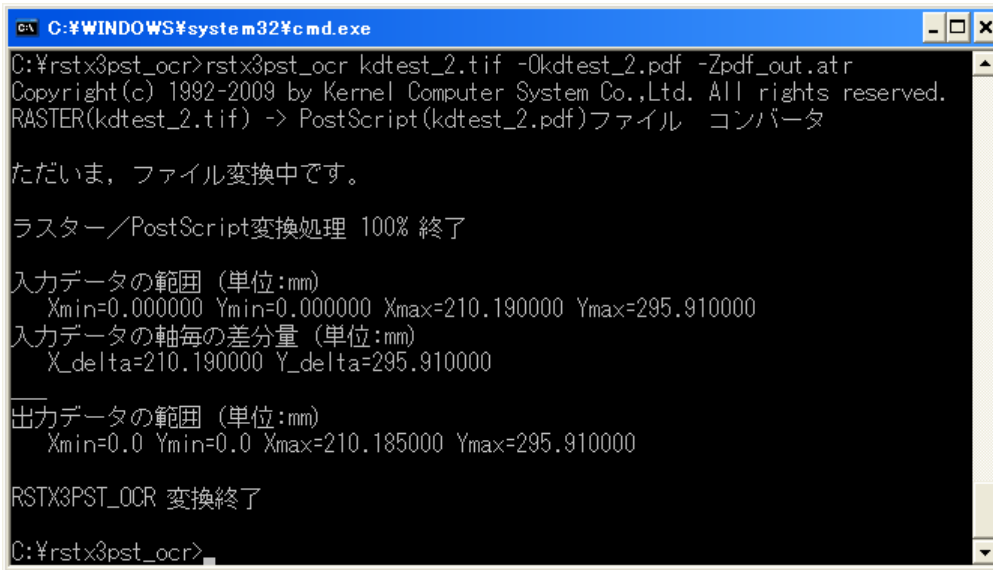
```
C:¥EXEC> RSTX3PST_OCR inputdata.tif -0outputdata.pdf -Zpdf_out.atr
```

RSTX3PST_OCR : 実行モジュール名
inputdata.tif : 変換元のラスターファイル名
outputdata.pdf : 変換後の PDF ファイル名

- 0 : 変換結果を出力するファイル名を指定するオプション。
マイナス記号とアルファベットの「オー」です。
-0 と変換後のファイル名の間には空白を入れないようにします。
変換前のファイル名と変換後のファイル名に同じものを指定することはできません。
- Z : 出力属性パラメータファイル名を指定するオプション。

※ オプション指定は大文字でも小文字でもかまいませんが、全角文字は使えません。
※ オプションの詳細についてはオプションの項目を参照ください。

- ・コマンドプロンプトでの実行画面：



```
C:¥WINDOWS¥system32¥cmd.exe
C:¥rstx3pst_ocr>rstx3pst_ocr kdtest_2.tif -0kdtest_2.pdf -Zpdf_out.atr
Copyright(c) 1992-2009 by Kernel Computer System Co.,Ltd. All rights reserved.
RASTER(kdtest_2.tif) -> PostScript(kdtest_2.pdf)ファイル コンバータ

ただいま、ファイル変換中です。

ラスター／PostScript変換処理 100% 終了

入力データの範囲 (単位:mm)
  Xmin=0.000000 Ymin=0.000000 Xmax=210.190000 Ymax=295.910000
入力データの軸毎の差分量 (単位:mm)
  X_delta=210.190000 Y_delta=295.910000

出力データの範囲 (単位:mm)
  Xmin=0.0 Ymin=0.0 Xmax=210.185000 Ymax=295.910000

RSTX3PST_OCR 変換終了

C:¥rstx3pst_ocr>
```

2. テキスト合成機能について

RSTX3PST_OCR のテキスト合成機能を利用すると、イメージデータ上に簡単なテキストを書き込むことや、日付、ファイル名、ページ番号の貼り付けなどができます。

- ・処理の手順は、次のようになります。
 - ① 書き込むテキストの情報を記述した「テキストデータファイル」と、テキスト合成用の各種パラメータを記述した「テキスト入力属性データ変換パラメータファイル」を作成しておく。
 - ② -TX[text_file]オプションでそのファイルを指定して、変換処理を実行する。
- ・変換実行例：

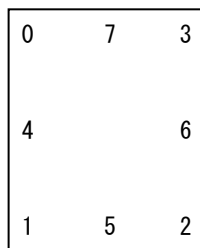
```
C:¥EXEC> RSTX3PST_OCR inputdata.tif -Ooutputdata.pdf -Zpdf_out.atr -TXtext_file.txt
```

・テキストデータファイルサンプル

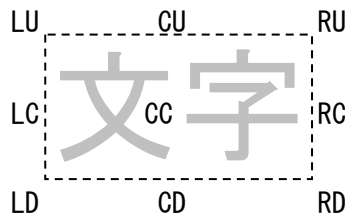
```
PAGE_NUMBER = 0
MOJI_FONT_NAME = MS ゴシック
KANJI_FONT_NAME = MS ゴシック
*GROUP
SET_PAGE_NO = Y
MOJI_HEIGHT = 5.08
MOJI_WIDTH = 5.08
POSITION_NO = 5
MOJI_REF = CD
PRINT_MODE = 1
PAGE_START_NO = 1
POSITION_OFFSET_Y = 5.08
*END
*GROUP
SET_FILENAME = Y
MOJI_HEIGHT = 3.81
MOJI_WIDTH = 3.81
POSITION_NO = 2
MOJI_REF = RD
POSITION_OFFSET_X = 0.0
POSITION_OFFSET_Y = 5.08
*END
*GROUP
DATE_TIME = Y
MOJI_HEIGHT = 3.81
MOJI_WIDTH = 3.81
POSITION_NO = 3
MOJI_REF = RU
POSITION_OFFSET_X = 0.0
POSITION_OFFSET_Y = 5.08
TEXT = DATE=%Y/%M/%D TIME=%h:%m:%s
*END
*GROUP
MOJI_HEIGHT = 5.0
MOJI_WIDTH = 5.0
ORG_X = 50.0
ORG_Y = 20.0
TEXT = RSTX3RS BY カーネル COMPUTER SYSTEM
*END
```

・テキストデータファイルのパラメータの説明

- *GROUP : テキストデータ定義の開始
- *END : テキストデータ定義の終了
- PAGE_NUMBER : マルチページファイルに文字列を合成する場合の、
合成先のページ番号を指定します(0:すべてのページに合成)
- MOJI_FONT_NAME : 使用するフォント名を指定します(※1)
- KANJI_FONT_NAME : 使用するフォント名を指定します(※1)
- MOJI_HEIGHT : 文字幅を指定します
- MOJI_WIDTH : 文字高さを指定します
- PRINT_MODE = 1 : ページ番号を割り付ける書式を指定します
- SET_PAGE_NO = Y : ページ番号の割り付けを指定します
- DATE_TIME = Y : 日付、時間の割り付けを指定します
- SET_FILENAME = Y : 出力データにファイル名を自動的に割り付けを指定します
- PAGE_START_NO : ページ番号を割り付ける開始ページを指定します
- POSITION_OFFSET_X : ページ番号とファイル名を割り付ける横余白を指定します
- POSITION_OFFSET_Y : ページ番号とファイル名を割り付ける縦余白を指定します
- TEXT : 表示文字列を指定します
- ORG_X : 文字の表示位置を指定します
- ORG_Y : 文字の表示位置を指定します
- POSITION_NO : ページ番号とファイル名と日付、時間の割り付け位置を指定します



MOJI_REF : 文字の基準位置を指定します



(※1)Windows 版では、フォント名称を指定します。
Unix/Linux 版ではフォントファイルを指定します。

・テキストデータファイル作成時の制限事項

- ・ファイル内に空白行は作成できません。
- ・コメント行を作成する場合は、先頭(1カラム目)に#を指定して下さい。
- ・テキスト合成、ページ番号の割付、日付時間の割付は独立の“*GROUP ~ *END”の中に指定しなければなりません。
- ・パラメータの値を変更する場合は、必ず指定して下さい。
複数指定した場合には、最後に指定したパラメータの値が有効になります。

3. OCR 処理機能について

イメージの文字は、そのままでは選択したり、コピーしたりできません。
イメージを OCR (文字認識) 処理してテキストデータ (PS, PDF 形式) に変換すると、その結果を検索したり、修正したり、コピーしたりできます。
本製品の OCR 機能の実現には、メディアドライブ株式会社の「活字文書 OCR ライブラリ」を使用しています。

- ・ 処理の手順は、次のようになります。
 - ① OCR 処理用属性ファイルに OCR 処理の各種パラメータを指定します。
 - ② “-RAS_OCR”, “-SAVE_OCR”, “-IMAGE_OCR”等 OCR 処理用オプションを指定して、変換処理を実行する。

・ 変換実行例 :

```
RSTX3PST_OCR in.tif -Oout.pdf -Zpdf_out.atr -RAS_OCR -SAVE_OCRO -IMAGE_OCR1
```

- RAS_OCR[atr_name] : OCR 処理を行います。
atr_name : OCR 処理用属性パラメータファイル (初期値 : ocr_in.atr)
- SAVE_OCR[n] : 出力ファイルの保存方法を選択します。
N = 0 : ラスターデータ + OCR 文字列 → 同一ページ (初期値)
表示しているラスタデータと OCR 変換した文字列を同一ページに出力します。
1 : ラスターデータ (1 ページ目) + OCR 文字列 (2 ページ目)
表示しているラスタデータを出力ファイルの 1 ページ目、OCR 変換した文字列を 2 ページ目に出力します。
2 : OCR 文字列
OCR 変換した文字列のみを出力ファイルに出力します。
- IMAGE_OCR[n] : ラスターデータと OCR 文字列を同一ページにするときにイメージを前面にする、テキストを前面にするのどちらかを選択します。
0 : イメージを前面 (初期値)
1 : テキストを前面

・ OCR 処理属性ファイル設定サンプル :

```
#symbol height_scale & width_scale & space(unit:mm)
OCR_SYMHT = 0.9
#OCR_SYMWD = 1.0
#OCR_SYMSP = 0.0
# Text output mode (0:word(default),1:stream)
TEXT_MODE = 1
# Text color pen number (1(default))
OCR_COLOR = 5
DICTIONARY_DIR = C:¥rstx3pst_ocr¥Dic¥
#LANGUAGE = 1
JAPANESE_MODE = Y
LEFT_X = 10
LEFT_Y = 10
RIGHT_X = 210
RIGHT_Y = 100
```