

Tworzenie plików w formacie DjVu z wykorzystaniem oprogramowania DocumentExpress Enterprise Edition



Jakub Bajer

Biblioteka Główna Politechniki Poznańskiej

Krzysztof Ober

Poznańska Fundacja Bibliotek Naukowych



Plan prezentacji

- I. Format DjVu
- II. Rodzina aplikacji DocumentExpress
- III. Komponenty DocumentExpress EE
- IV. Wymagania systemowe
- V. Obsługiwane typy plików
- VI. Configuration Manager
- VII. Workflow Manager
- VIII. Komendy uruchamiane z linii poleceń

I. Format DjVu

- Wysoce efektywna metoda kompresji obrazu.
- Obraz rozdzielany jest na odrębne warstwy. Każda warstwa jest oddzielnie poddawana optymalizacji i kompresji.
- Zastosowanie zarówno do kompresji kolorowych, skanowanych dokumentów, jak i do dokumentów generowanych elektronicznie.
- <http://www.djvu.pl/>

I. Format DjVu

Podstawowe cechy:

- bardzo małe rozmiary plików,
- doskonała jakość obrazu,
- przenośność,
- efektywne przeglądanie,
- optymalizacja pod kątem zastosowań WWW (WWW ready),
- możliwość przeszukiwania,
- wiarygodność dokumentów (niemodyfikowalność),
- otwarty format,
- znak wodny w dokumencie, ochrona hasłami, panele graficznych miniatur stron lub rozwijalnych spisów treści, zintegrowane w dokumencie metadane o publikacji.

II. Rodzina aplikacji DocumentExpress

Główne składniki i cechy rodziny DocumentExpress:

- DjVu Editor,
- wirtualna drukarka,
- DjVu Browser Plugin,
- system OCR,
- hyperlinki i adnotacje.

II. Rodzina aplikacji DocumentExpress

Produkty rodziny DocumentExpress:

- DocumentExpress Professional Edition
- **DocumentExpress Enterprise Edition**
- DocumentExpress SDK

III. Komponenty DocumentExpress EE

1) Graficzne (tylko Windows):

- Configuration Manager – interfejs graficzny do zarządzania profilami (zestawami parametrów przetwarzania) - umożliwia modyfikację istniejących, tworzenie nowych oraz testowanie działania profili;
- Workflow Manager – interfejs graficzny do wsadowego przetwarzania plików DjVu.

III. Komponenty DocumentExpress EE

2) Programy uruchamiane z linii poleceń (wszystkie platformy):

- ***documenttodjvu*** – konwersja obrazów rastrowych do formatu djvu z obsługą warstw;
- ***photododjvu*** – konwersja obrazów rastrowych do formatu djvu bez obsługi warstw;
- ***djvutotext*** – ekstrahowanie warstwy tekstowej do pliku tekstowego;
- ***djvudecode*** – konwersja plików djvu do obrazów rastrowych;
- ***djvutoxml*** – ekstrahowanie adnotacji, metadanych oraz warstwy tekstowej do pliku XML;

III. Komponenty DocumentExpress EE

- ***djvuparsexml*** – przetwarzanie informacji tekstowych zawartych w pliku XML, import do pliku djvu;
- ***djvubundle*** – konwersja pliku DjVu do formatu *bundled* (opcjonalnie tworzenie warstwy OCR oraz osadzanie miniatur);
- ***djvujoin*** - konwersja pliku DjVu do formatu *indirect* (opcjonalnie tworzenie warstwy OCR oraz osadzanie miniatur);
- ***watermarkdjvu*** – osadzanie znaku wodnego w dokumencie DjVu.

IV. Wymagania systemowe

Win32

- Microsoft Windows 98, 2000, NT4.0, XP,
- procesor Intel x86 400MHz,
- 256 MB (512 MB dla plików o rozdzielczości > 400 dpi).

Solaris (wersja 9)

- procesor Intel x86 lub Sparc 400MHz ,
- 256 MB (512 MB dla plików o rozdzielczości > 400 dpi).

Linux (kernel 2.2)

- procesor Intel x86 400MHz ,
- 256 MB (512 MB dla plików o rozdzielczości > 400 dpi).

IV. Wymagania systemowe

Wymagana przestrzeń dyskowa

- Na dysku muszą pomieścić się wszystkie pliki 😊
- Wielostronicowe dokumenty DjVu wymagają dodatkowego miejsca na dysku na pliki tymczasowe, które są tworzone w czasie konwersji. Można przyjąć, że wielkość plików tymczasowych jest porównywalna z wielkością wynikowego pliku DjVu.

V. Obsługiwane typy plików

Pliki wejściowe:

- TIFF 6.0 (.tiff, .tif)
- JPEG (.jpeg, .jpg)
- BMP (.bmp)
- GIF (.gif)
- PNM (.pnm, .ppm, .pgm)
- PBM (.pbm)
- DjVu® (.djvu, .djv)
- PDF (.pdf)

V. Obsługiwane typy plików

Pliki wyjściowe:

- TIFF 6.0 (.tiff, .tif)
- JPEG (.jpeg, .jpg)
- BMP (.bmp)
- PS (.ps)
- PNM (.pnm, .ppm, pgm)
- PBM (.pbm)
- DjVu (.djvu, .djv)

VI. Configuration Manager

- *documenttodjvu.conf* (stock profiles),
- profile ogólne (common profiles):
 - bitonal,
 - manuscript,
 - drawn,
 - scanned,
 - clean,
 - photo,
- profile specjalne (special profiles).

VI. Configuration Manager

Podstawowe parametry konwersji:

- Background Subsample
- Foreground Subsample
- Threshold Balance

VI. Configuration Manager

Configuration Manager umożliwia testowanie różnych profili konwersji przed rozpoczęciem właściwego procesu przetwarzania. Dzięki temu użytkownik może dobrać optymalną konfigurację dla konkretnego typu dokumentów źródłowych.

VII. Workflow Manager

Przykłady zastosowań:

- konwersja pojedynczych plików,
- przetwarzanie wsadowe,
- *watched folders* („gorące foldery”) + serwer FTP.

VIII. Komendy uruchamiane z linii poleceń

Przykłady zastosowań:

- ***printtodjvu***: konwersja dokumentu MS Word do formatu DjVu,
- ***djvujoin***: konwersja dokumentu DjVu typu *bundled* do *indirect* (dla potrzeb WWW).