



# Tworzenie i edycja plików DjVu za pomocą djevulibre



Źródło Słów

Poznań

27-29 października 2017

Autor: Ankry



# djvulibre



- Zestaw programów do tworzenia i edycji plików w formacie DjVu
- Licencja GNU GPL v2
- Wersje dla Linux, MacOS (32bit), Windows (32bit), OS/2
- Programy wiersza poleceń (command line)
- więcej informacji:

<http://djvu.sourceforge.net/index.html>



# djvulibre



- Podstawowe narzędzia pakietu
- djvm – tworzenie plików DjVu, ich modyfikacja i listowanie zawartości
- c44 – tworzenie jednostronicowego DjVu na podstawie obrazka wielobarwnego (np. jpg)
- cjb2 – tworzenie jednostronicowego DjVu na podstawie obrazka czarno-białego (bichromatycznego; nie: w skali szarości)
- djvutxt – ekstrakcja tekstu zawartego w pliku DjVu



# djvm



- Listowanie zawartości:
  - `djvm -l plik.djvu`
- pokazuje:
  - rozmiar strony
  - wewnętrzną nazwę strony
  - lokalizację plików iff

```
$ djvm -l Gabryel_D'Annunzio_-_Dzwony.djvu
Size      Type      Name
-----
   901    SHARED_ANNNO    shared_anno.iff
  9034    PAGE #1      p0001.djvu
  6746    PAGE #2      p0002.djvu
 31703    PAGE #3      p0003.djvu
 17365    PAGE #4      p0004.djvu
 45593    PAGE #5      p0005.djvu
 58103    PAGE #6      p0006.djvu
 60079    PAGE #7      p0007.djvu
 59040    PAGE #8      p0008.djvu
 57852    PAGE #9      p0009.djvu
 54020    PAGE #10     p0010.djvu
 54797    PAGE #11     p0011.djvu
 57660    PAGE #12     p0012.djvu
 58287    PAGE #13     p0013.djvu
 32578    PAGE #14     p0014.djvu
 42308    PAGE #15     p0015.djvu
 55538    PAGE #16     p0016.djvu
 57392    PAGE #17     p0017.djvu
 55160    PAGE #18     p0018.djvu
 55684    PAGE #19     p0019.djvu
 54914    PAGE #20     p0020.djvu
 58117    PAGE #21     p0021.djvu
 58454    PAGE #22     p0022.djvu
 57106    PAGE #23     p0023.djvu
```



# djvm



- Usuwanie stron
  - `djvm -d plik.djvu nr_strony`
- Usuwa pojedynczą stronę

```

$ djvm -d Gabryel_D'Annunzio_-_Dzwony.djvu 1
$ djvm -l Gabryel_D'Annunzio_-_Dzwony.djvu
Size      Type      Name
-----
   901     SHARED_ANN shared_anno.iff
31703     PAGE #1    p0003.djvu
17365     PAGE #2    p0004.djvu
45593     PAGE #3    p0005.djvu
58103     PAGE #4    p0006.djvu
60079     PAGE #5    p0007.djvu
59040     PAGE #6    p0008.djvu
57852     PAGE #7    p0009.djvu

```

```

Microsoft Windows [wersja 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. Wszelkie prawa zastrzeżone.

C:\Windows\system32>cd \Users\Ankry\Desktop

C:\Users\Ankry\Desktop>"\Program Files (x86)\DjVuLibre\djvm.exe" -d Gabryel_D'Annunzio_-_Dzwony.djvu 1

C:\Users\Ankry\Desktop>

```



# djvm



- Wstawianie stron
  - na końcu pliku:
    - `djvm -i plik.djvu strona.djvu`
  - jako podaną stronę:
    - `djvm -i plik.djvu strona.djvu nr_str`
- Można w ten sposób połączyć dwa wielostronicowe pliki DjVu

```
$ djvm -i Gabryel_D'Annunzio_-_Dzwony.djvu dumb.djvu 2
$ djvm -l Gabryel_D'Annunzio_-_Dzwony.djvu
Size      Type      Name
-----
  901     SHARED_ANN    shared_anno.iff
31703     PAGE #1      p0003.djvu
 2946     PAGE #2      dumb.djvu
17365     PAGE #3      p0004.djvu
45593     PAGE #4      p0005.djvu
58103     PAGE #5      p0006.djvu
60079     PAGE #6      p0007.djvu
59040     PAGE #7      p0008.djvu
57852     PAGE #8      p0009.djvu
54020     PAGE #9      p0010.djvu
```



# djvm



- Wstawiane strony mogą pochodzić z „rozdzielonego” pliku DjVu
  - Należy wtedy pamiętać, że z taką stroną jest skojarzony plik iff, bez którego strona jest bezużyteczna
- Djvulibre nie ma wygodnego mechanizmu ekstrakcji pojedynczej strony z pliku DjVu
  - Chyba jedyna metoda to zrobić kopię pliku i usunąć z niej wszystkie strony „przed” i „po” wybranej – bardzo nieefektywne
  - Lepiej korzystać z innych narzędzi do ekstrakcji stron



## c44



- Tworzy strony DjVu w oparciu o obrazki więcej niż dwukolorowe (także w skali szarości)
  - c44 -dpi 300 -decibel 40 plik.jpg plik.djvu
- Ograniczona liczba obsługiwanych formatów (pnm, jpg)
- Parametr *-decibel* w zakresie 16-50; przy wartościach poniżej 40 zdarzają się „przekłamanie literek” (z mojej praktyki)





# cjb2



- Tworzy strony DjVu w oparciu o bichromatyczne obrazki
  - `cjb2 -dpi 300 plik.pbm plik.djvu`
  - `cjb2 -dpi 300 -losslevel 30 plik.pbm plik.djvu`
  - `cjb2 -dpi 300 -lossless plik.pbm plik.djvu`
  - Parametr `losslevel` w zakresie 0-200
- Rzadko wykorzystywane – obrazki muszą być rzeczywiście bichromatyczne, a nie w skali szarości
- Formaty np. TIFF, PBM
  - niektóre PDF-y z Google Books zawierają „w środku” PBM-y



# djvm



- Pliki stron DjVu utworzone za pomocą cjb2/c44 można łączyć w jeden zagregowany plik DjVu za pomocą djvm:
  - `djvm -c całość.djvu strona1.djvu strona2.djvu ...`
  - `djvm -c całość.djvu *.djvu`
- Albo wstawiać do istniejącego DjVu (`djvm -i`)



# djvutxt



- Ekstrakcja warstwy tekstowej z całego pliku DjVu lub wybranych stron:
  - `djvutxt plik.djvu >plik.txt`
  - `djvutxt -page=10,12,30-115 plik.djvu >plik.txt`



# djvulibre



- Inne narzędzia, jak np.
  - djvused
  - djvuextract
  - djvumake
- wymagają dogłębnej wiedzy odnośnie formatu djvu i są przydatne głównie dla programistów tworzących narzędzia korzystające z djvulibre (vide narzędzia do ekstrakcji sformatowanego OCR na itwikisource czy rozmaite kreatory plików DjVu)



# DjVu browser plugin



- Zawiera DjVu viewer
- Pozwala na wygodne przeglądanie i **zapisywanie** plików DjVu z bibliotek cyfrowych
- Działa z Firefoxem (do wersji 51) i Internet Explorerem
  - [http://bin.us.lizardtech.com/ExpressView\\_6.6.1.zip](http://bin.us.lizardtech.com/ExpressView_6.6.1.zip)
  - <https://www.cuminas.jp/en/downloads/download/?pid=1>
- Lizardtech lub Cuminas (ta druga zawiera też viewer)
- Nie wspierany przez MS Edge, Chrome, Opera, Firefox v52+ (zarzucona technologia wtyczek NSAPI)