

# – DjVu –

Jan Ščotka (7. semestr)  
xscotka at fi.muni.cz  
8. prosince 2005

DjVu je velice slibný projekt určený pro přenos souborů a publikování na internetu. Jeho velikou výhodou je malá velikost souborů a progresivní dekompresní algoritmy, takže i při zobrazování velkých souborů není velká paměťová náročnost.

Základem komprese jsou tyto tři části.

**DjVuPhoto** – Známá jako IW44 používá wavelet continuous-tone kompresi. Používá progresivní vykreslování obrazu a dekompresi. Je určena pro kompresi obrázků s plynulými přechody barev. Je to něco jako JPEG.

**DjVuBitonal** – Známá jako JB2 je dvoubarevný kompresní algoritmus využívající opakování blízce podobných tvarů jako jsou třeba písmena. Nejlépe je použitelná pro černobílé obrázky reprezentující text nebo jednoduchou grafiku.

**DjVuDocument** – Určena pro kompresi barevných dokumentů, které obsahují text i obrázky. DjVuDocument rozděluje obrázky do oddělených vrstev. Přední vrstva, která obsahuje převážně text, je obvykle komprimována programem DjVuBitonal. Pozadí obsahuje grafiku nebo texturu a je komprimována algoritmem DjVuPhoto s menším rozlišením a ztrátovou kompresí.

Existují komerční produkty implementující tyto základní prvky, hlavním komerčním produktem je Document Express firmy Lizzardtech.

## DjVuLibre

<http://djvulibre.djvuzone.org/>

Základ projektu vznikl u firmy AT&T, která později prodala komerční licenci firmě Lizzardtech.

**Cíl projektu:** vytvořit referenční implementaci pro DjVu formát. Byla vytvořena firmou Lizzardtech a uvolněna pod *GNU GPL* licencí. Je kolekcí nástrojů, kódovacích a dekodovacích, pracujících s DjVu formátem dle výše uvedených principů. Ačkoli nástroje z referenční implementace nedosahují kompresních výsledků jako produkt firmy Lizzardtech, je jen otázkou času, kdy se tato situace zlepší. Pokud dojde k rozšíření tohoto formátu mezi veřejnost, bude se situace rapidně zlepšovat.

**Knihovna obsahuje následující nástroje:**

**bzz** – Obecně použitelný algoritmus na kompresi, používaný k zakódování interních struktur.

- c44** – Viz. DjVuPhoto, nástroj na kompresi fotek, grafiky.
- cjb2** – Viz. DjVuBitonal, nástroj pro zakódování dvoubarevných obrázků určen pro textovou vrstvu.
- cpaldjvu** – Viz. DjVuDocument, nástroj na kompresi dokumentů paletou barev (defaultně nastavené na 256) Dominantní barva je kódovaná jako pozadí, ostatní barvy jsou zakódovány do popředí.
- csepdjvu** – Nástroj jenž z předpřipravených obrázků popředí a pozadí vytvoří DjVuDocument soubor.
- djvm** – Nástroj na manipulaci s multi-page DjVu dokumenty. Umí skládat individuální stránky do dokumentu.
- djvmcvt** – Převéde dokumenty z formy jednoho dokumentu do nepřímé adresace stránek, tzn. indexový soubor v kterém jsou odkazy na jednotlivé stránky.
- djvused** – Nástroj na manipulaci s textovými informacemi v DjVu dokumentu. Muže upravovat následující informace: textovou vrstvu, metadata, anotace, odkazy do internetu.
- djvumake** – Příkaz na skládání DjVu obrázkových souborů.
- djvutoxml, djvuxmlparser** – nástroje pro editaci metadat v DjVu souboru pomocí XML.
- djvudigital** – skript, který spouští ghostscript s DjVu ovladačem a generuje z postscriptu nebo PDF DjVu dokument.  
Možné parametry lze nalézt v dokumentaci, mezi důležité patří:  
  --words --lines – Extrahují text z postscriptového souboru a začlení jej do DjVu dokumentu a nahraje jeho polohu. Jak napovídají názvy, buď pro slova, nebo řádky.  
  --dpi – Určující rozlišení.  
  --treshhold – Určuje rozsah pro separaci pozadí a popředí.  
  --bg-slice – Určuje kvalitu pozadí.

### DjVudigital

<http://www.research.att.com/~haffner/gsdjvu/>

[http://sourceforge.net/project/showfiles.php?group\\_id=32953](http://sourceforge.net/project/showfiles.php?group_id=32953)

Je převaděč, který je velice kvalitní a dosahuje dobrých kompresních poměrů i oproti PDF. Ačkoli narazil jsem na případy kdy PDF soubor byl menší než DjVu, byli tyto případy vyjimečné a rozdíl velikosti byl malý.

Porovnání bylo provedeno na mé bakalářské práci kterou byla částečně vypracovaná. Nejedná se o exaktní výsledky, protože nebyly použity žádné optimalizace formátů ani komprimačních programů. Ale poskytuje určitý přehled jak to může s velikostí souborů vypadat. Bakalářská práce v této fázi neobsahovala žádný

obrázek.

soubor	bez komprese	gzip komprese	bzip2 komprese
bp.tex	15K	6,4K	6,3K
bp.dvi	30K	12K	11K
bp.ps	172K	33K	28K
bp.pdf	91K	74K	76K
bp.djvu	21K	20K	21K

**Licenční problémy** – Základním problémem *djvudigital* je ovladač pro ghostscript. Tento ovladač pro ghostscript vytvořil León Bottou pro firmu AT&T. Firma AT&T se rozhodla uvolnit projekt pod otevřenou licencí, avšak použila licenci CPL (Common Public Licence), kterou používala již v minulosti. Problém nastává při použití v projektech pod GPL licencí, protože pokud je použita, musí být vše v projektu pod GPL. Proto není možno začlenit ovladač do distribuce GPL ghostscriptu, ani do projektu DjVuLibre.

Můžeme jen doufat že se projekt bude vyvíjet rychlým tempem alepší se kvalita použitých komprimačních algoritmů v knihovně DjVuLibre. Knihovna také obsahuje exportní funkce do grafického formátu PNM, TIFF a postscriptu. Rovněž je zahrnut pruhlížeč DjVu.

## JavaDjVu

<http://javadjvu.foxtrottechnologies.com/>

**Cíl projektu:** přidání podrory pro prohlížení DjVu formátu v jazyku Java. S následnou možností použití ve webovém prohlížeči. Byl vyvinut firmou *Foxtrot Technologies Inc.* a uvolněn pod GNU GPL licencí.

Tato knihovna samozřejmě umožňuje prohlížet DjVu soubory lokálně bez nutnosti použít webový prohlížeč. Použití: `java -jar javadjvu.jar jmeno_djvu.djvu`. Avšak pro tento typ prohlížení není primárně určen.

Právě z velikostí DjVu souborů pramení vyhoda prohlížení přes internet. Protože Java *jar* archiv se může lokálně uchovávat, tak velikost této knihovny není podstatná (velikost *javadjvu.jar* je 200Kb a *DjVuApplet.class* 137b). Výhoda formátu se projeví nejvíce v dokumentech rozdělených na jednotlivé stránky, kdy existuje malý hlavní soubor, obsahující odlazy na jednotlivé stránky (Obsah publikace), které se načítají individuálně v případě potřeby.

### Příklad použití v HTML kódu:

```
<applet codebase = "../.."
  code = "DjVuApplet.class"
  archive = "javadjvu.jar"
  width = "98%"
  height = "98%"
  style = "border:0px none;
  margin:0px 0px 0px 0px;
  width:100%;
  height:100%">
```

```
<param name = "data"
      value = "jmeno_djvu_souboru.djvu"/>
<param name = "cache_archive"
      value = "javadjvu.jar"/>
</applet>
```

Stav projektu umožňuje spolehlivě prohlížet dokumenty, ať již lokálně nebo přes webový prohlížeč a to i přeto že je projekt ve verzi *0.8.08*. Počítač na kterém se spouští musí mít nainstalováno J2SE Runtime Environment buď 1.1.4, nebo J2ME 2.0, či novější verze.

Projekty kolem formátu DjVu jsou slibné a čas ukáže jestli se formát ujme v širokém povědomí, či zkončí na smetišti digitálních dějin.

## Reference

- [1] *DjVu Project Pages*. <http://www.djvuzone.org>.
- [2] *DjVu Tutorial*. <http://www.djvuzone.org/support/tutorial/>.
- [3] *DjVuLibre Documentation*. <http://djvulibre.djvuzone.org/doc/index.html>.
- [4] David Evans. Postscript vs. pdf. Technical report, Adobe Systems Incorporated, 2005. Why do we offer two printing technologies? How do they differ?
- [5] Patrick Haffner, Yann LeCunn, Leon Bottou, Pascal Vincent Paul Howard, and Bill Riemers. Color documents on the web with djvu. Technical report, AT&T Labs – Research, 1999.
- [6] Foxtrot Technologies Inc. *Java DjVu Viewer Project*. <http://javadjvu.foxtrottechnologies.com/>.
- [7] Yann LeCun, Léon Bottou, and Patrick Haffner. Djvu: Results and comparisons. Technical report, AT&T Labs – Research, 2000.
- [8] Judith Wusteman. Formats for the electronic library. Department of Library and Information Studies of University College Dublin, 1998. <http://www.ariadne.ac.uk/issue8/intro.html>.