(CBZ) و(DjVu): صيغتان لحفظ الوثائق الرقمية خارج (PDF)

سيث كينلون



الملخص:

يستعرض المقال صيغتي (CBZ) و(DjVu) خيارين لحفظ الوثائق المصورة، انطلاقًا من رغبة الكاتب في رقمنة كتب عائلية. ويشرح بساطة أرشيف القصص المصورة بوصفه حزمة صور مضغوطة مرتبة، مع سهولة إنشائه وفكه ومحدودية توحيد بياناته الوصفية. ثم يعرض خصائص (DjVu) في الضغط والصفحات والعلامات المرجعية والنص المضمن، وأدوات إنشائه وتعديله وفهرسته. ويتناول ربط النص بإحداثيات الصورة وتعقيد هذه العملية، وإمكان استخدام (PDF) وسيطًا لأتمتة التحويل عند توفر سلسلة الأدوات اللازمة. ويوازن بين سهولة العمل والتحكم في البنية وقابلية العرض، مؤكدًا ملاءمة كل صيغة لاحتياجات معينة. وتبقى أمثلة الأوامر منقولة في سياقها التاريخي، مع الحفاظ على التنبيه إلى مراجعة معاملات الأدوات قبل استخدامها عمليًا.

الكلمات المفتاحية: حفظ الوثائق، (CBZ)، (DjVu)، الأرشفة الرقمية، البيانات الوصفية

اكتشفت مؤخرًا أن جد جدي ألّف قرب مطلع القرن العشرين كتابين: أحدهما عن الملاحة، والآخر عن عمله رئيسًا لإطفاء نيويورك. جمهورهما محدود، لكنهما من تاريخ أسرتي، وأردت حفظ نسخة رقمية منهما. فما صيغة الوثيقة المحمولة الأنسب؟

استبعدت (PDF) مبكرًا. فرغم فائدتها في تجهيز الطباعة، تبدو محكومة بتضخم الوظائف، وتنتج وثائق يصعب فحص بنيتها وتحريرها. أردت صيغة أذكى بخصائص شبيهة، فحضرت صيغة أرشيف القصص المصورة و(DjVu).

أرشيف القصص المصورة:

صيغة بسيطة تستخدم غالبًا للقصص المصورة، وتوجد أمثلتها في (Comic Book Plus) (The Digital Comicg (Museum. أكبر مزاياها هو نقطة ضعفها: إنها أقرب إلى عرف من صيغة؛ فهي ملف (ZIP) أو (TAR) أو (7Z) أو (RAR) بامتداد (CBZ) أو (CBT) أو (CB7) أو (CBR) على الترتيب، ولا معيار فيها لتخزين البيانات الوصفية. لكنها سهلة الإنشاء.

إنشاء الأرشيف:

أنشئ مجلد صور، وأعد تسميتها بتسلسل رقمي:

```
$ n=0 && for i in *.png ; do mv $i `printf %04d $n`.png ; ((n+=1)); done
```

اضغط الملفات بأداتك المفضلة. وجدت أن (CBZ) أوسع دعمًا:

```
$ zip comicbook.zip -r *.png
```

ثم غيّر الامتداد:

```
$ mv comicbook.zip comicbook.cbz
```

يفتح الناتج في أغلب أجهزتك. يدعمه (Evince) و(Okular) في لينكس، و(Document Viewer) و(Bubble) في أندرويد.

فك الأرشيف:

استعادة البيانات سهلة: فك ضغط (CBZ). وإذا لم تتعرف أداة الأرشفة إلى الامتداد، فأعده إلى امتداده الأصلي:

```
$ mv comicbook.cbz comicbook.zip
$ unzip comicbook.zip
```

صيغة (DjVu):

طورتها (AT&T) قبل أكثر من عشرين سنة، وتُنطق «ديجا فو». وهي صيغة رقمية ذات ضغط متقدم، تقرؤها تطبيقات أكثر مما قد تتصور: (Evince) و(Okular) و(DjVu.js) على الويب وإضافته لفايرفوكس و(GNU Emacs) و(Document Viewer) لأندرويد، وعارض (DjView) المفتوح متعدد المنصات في (Sourceforge). وتوجد معلومات وأمثلة في (djvu.org).

تدعم ضغط الصور وبنية الفهرس، أو العلامات المرجعية، والنص المضمن، ويسهل فحصها وتحريرها بأدوات حرة مفتوحة.

التثبيت:

مجموعة الأدوات المفتوحة هي (DjVuLibre)، وتوجد في مستودع توزيعتك. مثلًا في فيدورا:

```
$ sudo dnf install djvulibre
```

إنشاء ملف:

ملف (DjVu) صورة مرمزة بهذه الصيغة، ويمكن أن يضم صورة واحدة أو صورًا متعددة في صفحات. وللإنشاء اليدوي استخدم (c44) للصور العالية الجودة أو (cjb2) للصور الثنائية اللون البسيطة. يقبل الأول (PNM) و(JPEG)، والثاني (PBM) و(TIFF).

إن احتجت معالجة سابقة، فاستخدم (ImageMagick) مع (density) لتحديد الدقة:

```
$ convert -density 200 foo.png foo.pnm
```

ثم حوّل الصورة:

```
$ c44 -dpi 200 foo.pnm foo.djvu
```

ولصورة بسيطة، كنص أسود على بياض، جرب المرمز الأبسط. حوّلها أولًا إلى صيغة وسيطة ملائمة عند الحاجة:

```
$ convert -density 200 foo.png foo.pbm
```

ثم إلى (DjVu):

```
$ cjb2 -dpi 200 foo.pbm foo.djvu
```

صار لديك مستند بسيط من صفحة واحدة.

إنشاء مستند متعدد الصفحات:

مع فائدة الصفحة المفردة بضغط (DjVu) الجيد أحيانًا، يشيع استخدامه متعدد الصفحات. إذا كان المجلد يحوي ملفات (DjVu) متعددة، تجمع بالأمر (djvm):

```
$ djvm -c pg_1.djvu two.djvu 003.djvu mybook.djvu
```

خلافًا لـ(CBZ)، لا تؤثر أسماء الصور في الترتيب، بل يحفظ ترتيب تقديمها في الأمر. وإذا سميتها بتسلسل طبيعي: 001 و002 و003 و004، أمكن استخدام محرف شامل:

```
$ djvm -c *.djvu mybook.djvu
```

التعديل:

يمكن إدراج صفحة في مستند قائم:

```
$ djvm -i mybook.djvu newpage.djvu 2
```

تصبح (newpage.djvu) الصفحة الثانية في (mybook.djvu). ويمكن حذف الصفحة الرابعة هكذا:

```
$ djvm -d mybook.djvu 4
```

إضافة فهرس للتنقل:

يمكن إضافة بيانات وصفية مثل العلامات المرجعية. يدويًّا، أنشئ ملفًا نصيًّا بالفهرس، ببنية تشبه (Lisp) تبدأ بعنصر (bookmarks) ثم أسماء العلامات وأرقام الصفحات:

```
(bookmarks
("Front cover" "#1")
("Chapter 1" "#3")
("Chapter 2" "#18")
("Chapter 3" "#26")
)
```

تحدد الأقواس المستويات. في المثال علامات عليا فقط؛ ويمكن جعل قسم فرعي بتأخير قوس الإغلاق، مثل إضافة قسم إلى الفصل الأول:

```
(bookmarks
("Front cover" "#1")
("Chapter 1" "#3"
("Section 1" "#6"))
("Chapter 2" "#18")
("Chapter 3" "#26")
)
```

احفظ الملف وطبقه بأداة (djvused):

```
$ djvused -e 'set-outline outline.txt' -s mybook.djvu
```

افتح الملف لترى الفهرس.

تضمين النص:

يمكن تضمين عناصر نصية، يسميها (djvused) «النص المخفي»، لتتيح برامج مثل (Okular) و(DjView) التحديد والنسخ إلى الحافظة.

لكنها عملية معقدة: يجب أولًا امتلاك النص، بالتعرف الضوئي الجيد أو نسخ الصفحات يدويًّا، ثم ربطه بالصورة النقطية. وعند امتلاك النص وإحداثيات كل سطر أو كلمة، تستطيع كتابة سكربت (djvused) بكتل لكل صفحة:

```
select; remove-ant; remove-txt
# ------------------------
select "p0004.djvu" # page 4
set-txt
(page 0 0 2550 3300
(line 1661 2337 2235 2369 "Fires and Fire-fighters")
(line 1761 2337 2235 2369 "by John Kenlon"))

.
# ------------------------
select "p0005.djvu" # page 5
set-txt
(page 0 0 2550 3300
(line 294 2602 1206 2642 "Some more text here, blah blah blah."))
```

تمثل الأعداد الحدود الدنيا والعليا لإحداثيي (X) و(Y) لكل سطر: (xmin) و(ymin) و(xmax) و(ymax). وكل سطر مستطيل بوحدة البكسل، نقطة أصله أسفل يسار الصفحة.

يمكن تعريف الكلمات والأسطر والروابط، ورسم مناطق بأشكال غير المستطيلات. ويمكن تضمين بيانات محددة، مثل مفاتيح (BibTeX) الصغيرة (year) و(booktitle) و(editor) و(author)، ومفاتيح (DocInfo) المستعارة من (PDF) التي تبدأ بحرف كبير، مثل (Title) و(Author) و(Subject) و(Creator) و(Produced) و(CreationDate) و(ModDate).

أتمتة الإنشاء:

رغم متعة إعداد مستند دقيق يدويًّا، فإن اعتماده يوميًّا يكشف نقص وسائل الراحة المتاحة لـ(PDF)؛ فقليل من التطبيقات، إن وجد، يقدم «اطبع إلى (DjVu)» أو «صدّر إلى (DjVu)». ويمكن تجاوز ذلك باستخدام (PDF) وسيطًا.

للأسف، المكتبة اللازمة للتحويل الآلي السهل مرخصة بـ(CPL)، الذي له متطلبات لا تتوافق مع شفرة (GPL) في سلسلة الأدوات؛ لذا لا يمكن توزيعها مكتبةً مبنية مسبقًا، لكن تستطيع بناءها بنفسك.

والعملية بسيطة نسبيًّا بفضل سكربت بناء جيد من فريق (DjVuLibre). جهز النظام بأدوات التطوير. في فيدورا استخدم مجموعة (DNF):

```
$ sudo dnf group install @c-development
```

وفي أوبونتو:

```
$ sudo apt-get install build-essential
```

نزّل مصدر (GSDjVu) من (Sourceforge)، لا (DjVuLibre)؛ أي اختر الملف الأحدث بدل الزر الأخضر الكبير أعلى القائمة. فك الضغط وادخل المجلد:

```
$ cd ~/Downloads
$ tar xvf gsdjvu-X.YY.tar.gz
$ cd gsdjvu-X.YY
```

أنشئ مجلدًا اسمه (BUILD) بالضبط، فلا مجال هنا للإبداع:

```
$ mkdir BUILD
$ cd BUILD
```

نزّل حزم المصدر الإضافية، وبخاصة (Ghostscript) ؛ قد يكون مثبتًا لكنك تحتاج مصدره للبناء عليه. وتحتاج مصادر (jpeg) و(libpng) و(openjpeg) و(zlib). إن ظننت أنها موجودة فشغّل السكربت؛ وإذا لم يجدها فسيتوقف ويتيح تصحيح النقص ثم المحاولة.

شغّل سكربت (build-gsdjvu) التفاعلي المرفق؛ فهو يفك المصادر، ويضيف مشغل (gdevdjvu) إلى (Ghostscript)، ويبنيه وينقح ملفات الناتج غير اللازمة.

يمكن التثبيت في أي موضع ضمن (PATH)، الذي يمكن عرضه بـ(echo $PATH). مثلًا للتثبيت تحت (/usr/local/):

```
$ sudo cp -r BUILD/INST/gsdjvu  /usr/local/lib64
$ cd /usr/local/bin
$ sudo ln -s ../lib64/gsdjvu/gsdjvu gsdjvu
```

تحويل (PDF) إلى (DjVu):

بعد بناء المشغل يكفي أمر واحد:

```
$ djvudigital --words mydocument.pdf mydocument.djvu
```

ينقل الصفحات والعلامات المرجعية والنص المضمن. ويربط خيار (words) الكلمات المضمنة بنقاطها في (DjVu). وإذا لم يكن في (PDF) نص مضمن فلا ينتقل نص. هكذا تستخدم مزايا (PDF) المريحة في التطبيقات وتحصل في النهاية على (DjVu).

لماذا الصيغتان:

إنهما خياران مفيدان لحفظ الوثائق وأرشفتها. يبدو غريبًا حشر سلسلة صور داخل صيغة (PostScript) مثل (PDF)، بحسب وصف الكاتب، أو صيغة موجهة أساسًا للنص مثل (EPUB)؛ لذا يحسن امتلاك (CBZ) و(DjVu) خيارين آخرين. قد لا يصلحان لكل الوثائق، لكن الاعتياد عليهما يتيح اختيارهما حين يكون ذلك الأنسب.

المراجع:

- https://www.fireengineering.com/articles/print/volume-56/issue-27/features/chief-john-kenlon-of-new-york-city.html
- https://en.wikipedia.org/wiki/Comic_book_archive
- https://comicbookplus.com/
- https://digitalcomicmuseum.com/
- https://wiki.gnome.org/Apps/Evince

- https://okular.kde.org
- https://f-droid.org/en/packages/org.sufficientlysecure.viewer/
- https://f-droid.org/en/packages/com.nkanaev.comics/
- https://djvu.org/
- https://djvu.js.org/
- https://github.com/RussCoder/djvujs
- https://elpa.gnu.org/packages/djvu.html
- https://djvu.sourceforge.net/djview4.html
- https://djvu.org
- https://djvu.sourceforge.net
- https://www.imagemagick.org/
- https://en.wikipedia.org/wiki/Lisp_(programming_language)
- https://sourceforge.net/projects/djvu/files/GSDjVu/1.10/