

Kompresory

V nadtitulku uvedené skratky majú spoločné dve veci - prvou je autor Markus F. X. J. Oberhumer a druhou je, že všetky tieto programy sú tzv. komprimačné programy, resp. knižnice, t. j. môžete ich podobne ako RAR, ZIP použiť na komprimovanie binárnych, textových či iných ľubovoľných dát.

LZO

LZO je dátová kompresná knižnica, ktorá je určená pre kompresiu, resp. dekompresiu dát v reálnom čase. Pomenovanie LZO pochádza zo slov Lempel-Ziv-Oberhumer. Knižnica LZO bola napísaná v jazyku ANSI C. Zdrojový kód a kompresný dátový formát sú navrhnuté tak, aby mohli byť portované aj na iné platformy. K dnešnému dňu sú na stránke www.oberhumer.com/opensource/lzo k dispozícii zdrojové kódy v jazykoch Perl (verzia 1.08), Python (verzia 1.08), ANSI C (verzia 1.08) a Java (verzia 1.0). V LZO je implementovaných niekoľko algoritmov s nasledovnými funkciami:

- jednoduchá a veľmi rýchla dekompresia;
- pre dekompresiu nie je potrebná pamäť;
- pomerne rýchla kompresia;
- pre kompresiu je potrebná pamäť 64 Kb;
- bezpečný algoritmus;
- bezstratový algoritmus.

Knižnica LZO a LZO algoritmus sú distribuované pod licenciou GNU General Public License (GPL). Ak chcete získať špeciálnu licenciu pre komerčné použitie, treba kontaktovať autora. Knižnica LZO bola kompilovaná a testovaná na rozmanitých platformách, napr. DOS (16 + 32 bit), Windows 3.x (16-bit), Windows 95/98/ME/NT/2000/XP, Linux, HP-UX a DEC Alpha (64-bit). LZO tiež pracuje pod AIX, AmigaOS, ConvexOS, FreeBSD, IRIX, Mac OS, Mac OS X, NetBSD, Nintendo64, Palm OS (Pilot), PSX (Sony PlayStation), OpenBSD, Solaris, SunOS, TOS (Atari ST) a VxWorks.

Ako príklad uvádzam časy namerané na mojom starom Pentiu 133:

- memcpy(): ~60 MB/s;
- LZO1X dekompresia v jazyku C: ~16 MB/s;
- LZO1X dekompresia v optimalizovanom assembleri: ~20 MB/s;
- LZO1X-1 kompresia: ~5 MB/s.

Okrem toho môžete na citovanej stránke nájsť tzv. miniLZO, predstavujúce podskupinu, resp. časť funkcií knižnice LZO. Po rozbalení zdrojových kódov môžete tieto skompilovať napríklad takto:

```
Administrator@P1600 /tmp
$ make
Please choose one of the following targets:
gcc gcc-i386 gcc-rs6000
visualc watcomc watcomc16
hpux hpux9

Administrator@P1600 /tmp
$ make gcc
gcc -Wall -O2 -s -o testmini testmini.c minilzo.c
```

```
Administrator@P1600 /tmp
$ ./testmini LZO real-time data compression library (v1.08, Jul 12 2002).
Copyright (C) 1996-2002 Markus Franz Xaver Johannes Oberhumer
```

```
compressed 102400000 bytes into 401581 bytes
decompressed 401581 bytes back into 102400000 bytes
```

```
miniLZO simple compression test passed.
```

V prípade, že chcete tento príklad skompilovať pomocou Visual C++ 6.0, treba namiesto príkazu make použiť príkaz nmake s parametrom visualc, teda nmake visualc.

LZOP

LZOP je súborový kompresor, ktorý je veľmi podobný programu gzip (www.gzip.org).

Pre kompresiu používa kompresnú knižnicu LZO DATA. Medzi výhody oproti programu gzip patrí veľká kompresia a rýchla dekompresia (avšak za cenu pomerne veľkého kompresného pomeru). LZOP je distribuovaný pod licenciou GNU General Public License (GPL). To, že LZOP je rýchly kompresor a dekompresor, ukazuje aj tab. 1. Vo väčšine prípadov LZOP jasne vedie pred svojimi konkurentmi.

LZOP bol naprojektovaný s nasledovnými zámermi:

rýchlosť (pri kompresii aj dekompresii);

primeraná kompatibilita s gzip;

prenositeľnosť.

Hlavné rozdiely medzi LZOP a gzip:

Súbory komprimované pomocou LZOP sú zakončené príponou LZO.

LZOP štandardne nevymaže vstupný súbor.

NRV

Pomenovanie NRV pochádza zo slov Not Really Vanished. NRV je bezstratová dátová kompresná knižnica napísaná v jazyku C++, ktorá ponúka rýchlu kompresiu a veľmi rýchlu dekompresiu. Pri dekompresii nie je potrebná pamäť. NRV navyše obsahuje pomalé stupne kompresie, pomocou ktorých dosiahnete výborný kompresný pomer, hoci predsa len dekompresia má veľmi vysokú rýchlosť.

Funkcie:

NRV kompresný pomer je typicky lepší ako kompresný pomer zip, pkzip alebo gzip.

NRV dekompresia na mojom Intel Pentiu 133 dosahuje 13 MB/s.

NRV kód dekompresora môže byť menší ako 200 bajtov.

NRV nepotrebuje žiadnu dodatočnú pamäť pre dekompresiu - ani priestor na zásobníku.

NRV je základ veľmi komplikovaného kompresného prostriedku (compression engine), ktorý dovoľí jednoduchú adaptáciu pre špeciálne algoritmy a iné požiadavky.

NRV je ideálne riešenie pre embedded systems (vložené systémy) a podobné miesta.

Knižnica NRV je navrhnutá tak, aby pracovala na 16-bitových systémoch (ako je DOS, Windows 3.1, Atari ST a Pilot Palmtop) a tiež na 32-bitových a 64-bitových strojoch (ako Unix, VxWorks, Mac, Win32, BeOS, OS/2, 32-bit DOS extenders, Sony Playstation, Nintendo 64...).

UCL

UCL (UCL Compression Library) je bezstratová dátová kompresná knižnica napísaná v jazyku ANSI C. UCL implementuje niekoľko kompresných algoritmov, čím sa dosiahne vynikajúci kompresný pomer a veľmi rýchla dekompresia. Dekompresia nepotrebuje žiadnu dodatočnú pamäť. UCL je open source reimplementácia kompresného algoritmu NRV. Ak UCL porovnáme s LZO, UCL algoritmus dosahuje lepší kompresný pomer, ale dekompresia je pomalšia. Na mojom Pentiu 133 s nej boli dosiahnuté tieto časy:

memcpy(): ~60 MB/s;

UCL dekompresia v optimalizovanom asembleri: ~13 MB/s;

LZO dekompresia v optimalizovanom asembleri: ~20 MB/s;

LZO dekompresia v jazyku C: ~16 MB/s;

UPX

UPX (Ultimate Packer for eXecutables) je voľný, vysoko výkonný kompresný program pre kompresiu viacerých rozdielných spustiteľných formátov (podrobný prehľad nájdete v tab. 3). Posledná stabilná verzia programu UPX je označená 1.2.4 a nájdete ju na stránke <http://upx.sourceforge.net>. Kvality programu UPX možno charakterizovať dvomi vlastnosťami - excelentným kompresným pomerom a veľmi rýchlou dekompresiou. UPX je chránený licenciou GPL a pre kompresné služby používa kompresnú knižnicu NRV. Avšak kompatibilita tak trochu menej účinnej implementácie open source je dosahovaná prostredníctvom kompresnej knižnice UCL. UPX existuje v nasledujúcich variantoch: Win32 console version, 32-bit DOS version, Atari version, precompiled Linux version, source code (kvôli kompilácii však budete potrebovať kompresnú knižnicu UCL). Užívateľia Windows uprednostňujú verziu Win32, pretože lepšie podporuje dlhé názvy súborov. Pre Windows 95/98/Me/NT/2000/XP sa vyvíja veľmi pekné grafické rozhranie (GUI graphical frontends). Avšak už existuje (z inej dielne) funkčné grafické GUI rozhranie, ktoré vytvoril pán Dirk Paehl. Aktuálnu verziu programu nájdete na domovskej stránke www.dpaehl.de. Toto grafické rozhranie, aktuálne dostupné vo verzii 1.2.6 (obr. 1), má používateľom uľahčiť ovládanie programu UPX. Grafická nadstavba má v sebe vstavaný program UPX verzie 1.2.4, avšak máte možnosť si zvoliť aj externú verziu programu UPX. To pre prípad, že autor uvoľní novšiu verziu.

V tab. 2 nájdete niekoľko príkladov, podľa ktorých môžete posúdiť schopnosti UPX.

Záver

Uvedené knižnice a kompresné programy sa vyznačujú veľmi rýchlou kompresiou a dekompresiou a určite sami nájdete mnohé oblasti, kde by sa dali využiť. Pokiaľ vás niektorá z týchto kompresných metód zaujala, vedzte, že všetky uvedené kompresné programy a knižnice nájdete na Chip CD. Osobne doporučujem vyskúšať UPX.

Peter Gašparovič

Testy DOS/Windows				
Typ súboru	Najlepšia kompresia	Najrýchlejšia kompresia	Najrýchlejšia dekompresia	Celkove najlepši
Text	ENTROPY 0.5	LZOP 1.00w	LZOP 1.00w	SBC 0.860
Executable	RK 1.04.1	LZOP 1.00w	LZOP 1.00w	RAR (Win32) 3.00b5
EXE Packer (Win32)	ASPACK 2.00.1	UPX 0.99.2w	N/A	ASPACK 2.00.1
EXE Packer (DOS)	UPX 0.99.2w	UPX 0.99.2w	EXEPACK 4.06	UPX 0.99.2w
Zvuk (WAV)	MONKEY'S AUDIO 3.96	LZOP 1.00w	LZOP 1.00w	MONKEY'S AUDIO 3.96
Grafika (TIF)	ERI32 5.1fre	LZOP 1.00w	LZOP 1.00w	ERI32 5.1fre
Calgary Corpus	RK 1.04.1	LZOP 1.00w	LZOP 1.00w	SBC 0.968
Canterbury Corpus	COMPRESSIA 0.96	LZOP 1.00w	LZOP 1.00w	SBC 0.968
Worms 2	RK 1.04.1	LZOP 1.00w	LZOP 1.00w	SBC 0.968

Súbor	Formát	Executable	Originál	Komprimovaný	Pomer %
Netscape 4.06	win32/pe	netscape.exe	2 934 336	1 124 352	38,3
Descent 2	watcom/le	descent2.exe	1 448 873	652 832	45,5
MAME 0.36	djgpp2/coff	mame.exe	8 214 016	1 810 056	22,0
OneOhOne	dos/exe	101.exe	438 144	179 566	41,0
Emacs 20.2	linux/386	emacs-20.2	2 772 657	925 543	33,4
g++ 2.8.1	atari/tos	cc1plus.ttp	1 595 049	655 508	41,1

UPX	Pack	Unpack	Popis	Link
atari/tos	•	•	Atari TOS/MiNT executable	www.freemint.de
djgpp2/coff	•	•	DOS 32-bit COFF	www.delorie.com/djgpp
dos/com	•	•	DOS 16-bit. com súbor	
dos/exe	•	•	DOS 16-bit executable	
dos/sys	•	•	DOS 16-bit. sys súbor	
linux/386	•	•	Linux AOUT/ELF/script	www.linux.org
rtm32/pe	•	•	Windows 32-bit executable / DLL	
tmt/adam	•	•	DOS 32-bit executable	www.tmt.com
watcom/le	•	•	DOS 32-bit linear executable	
win32/pe	•	•	Windows 32-bit executable / DLL	www.openwatcom.org