

Lossless Compression Algorithms Comparison 2.0

Petr Šupina, 2004

Remarks

The document doesn't cover all available formats or issues such as compatibility or speed. All programs were set up for **the best compression** (including solid archiving), but there are some exceptions. ZIP, 7ZIP, GZIP, BZIP2 and maybe some others can be tuned even better, but not in standard tools or at a cost of a huge memory usage (e.g. over 1GB of virtual memory and expect no miracles). That tuning is possible with 7ZIP, Stuffit and so on (they can handle multiple formats).

Why use TAR for some formats? TAR offers no compression but it can join several files into one. GZIP, BZIP2 and UNIX Z cannot compress more than one file, so that is why. These formats are spread especially under non-Windows based operating systems.

I believe that LHA, UC2, ARC and ZOO haven't been updated for a pretty long time, but unpacking is still supported by many programs.

Algorithm	Files	Plain Text (C++)	Plain Text (MSI Setup log)	Binary File (winword.exe)	Binary File (mshtml.dll)	738 Plain Text Files (*.inf)	324 Binary Files (*.exe)
none	(*.*)	1 908 477	30 993 110	12 037 688	2 795 520	21 447 202	27 416 681
7ZIP	(*.7z)	269 067	796 759	4 754 936	1 118 350	1 343 198	8 521 589
ACE 2.0	(*.ace)	339 810	1 225 892	5 403 695	1 320 898	1 628 653	8 554 940
ARC 6	(*.arc)	882 542	8 411 599	8 951 798	2 186 962	6 429 233	17 284 047
ARJ 3	(*.arj)	417 976	1 522 763	6 520 850	1 580 124	2 661 228	12 282 143
Black Hole	(*.bh)	409 006	1 511 095	6 516 472	1 578 703	2 668 714	12 286 131
CAB 1.3	(*.cab)	325 451	1 248 959	5 297 826	1 284 907	1 651 058	8 613 163
LHA 6	(*.lzh)	435 169	1 566 618	6 616 733	1 593 367	2 675 734	12 400 874
RAR 2.9	(*.rar)	286 402	1 186 981	5 337 535	1 310 119	1 497 051	8 419 493
Stuffit	(*.sit)	333 104	999 232	6 311 840	1 468 809	2 339 761	11 897 781
Stuffit X	(*.sitx)	286 269	865 820	4 927 499	1 239 243	1 471 043	8 142 166
TAR+BZIP2	(*.tar.bz2)	390 911	1 263 512	6 526 943	1 546 012	2 115 397	12 014 453
TAR+GZIP	(*.tar.gz)	412 841	1 558 142	6 516 544	1 578 834	2 348 407	12 141 892
TAR+UNIX Z	(*.tar.z)	698 253	3 676 983	8 663 469	2 072 697	5 208 621	17 489 059
UC2	(*.uc2)	392 372	1 499 496	6 407 444	1 546 876	2 339 924	11 839 770
ZIP 2.0	(*.zip)	408 748	1 513 160	6 500 902	1 574 311	2 649 532	12 269 784
ZOO 2.1	(*.zoo)	811 421	6 878 598	8 722 087	2 131 525	5 858 101	16 942 589

* All numbers are filesizes in bytes.

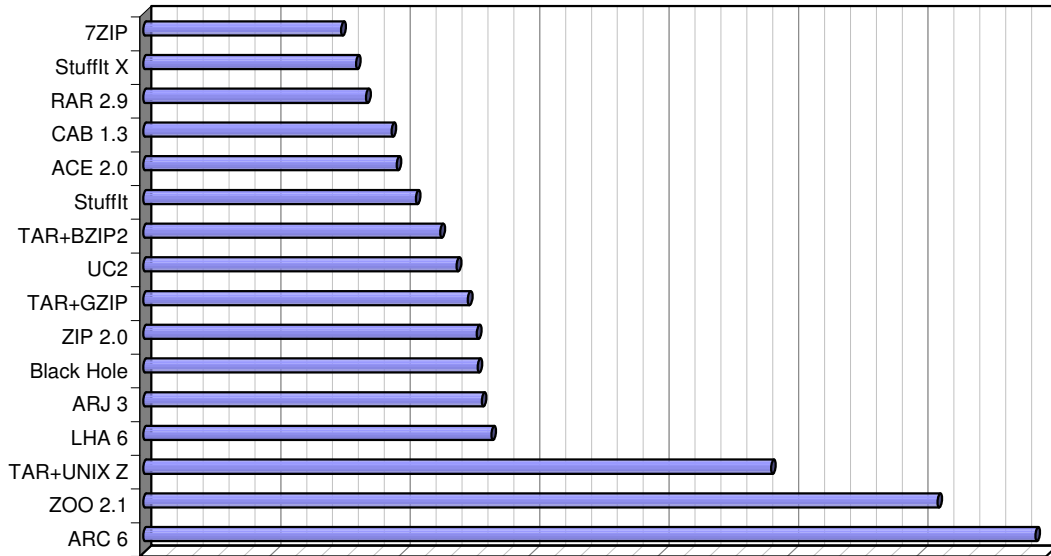
* The first line (blue) represents original sizes of files.

* The red marks are the smallest size in a comparison to others.

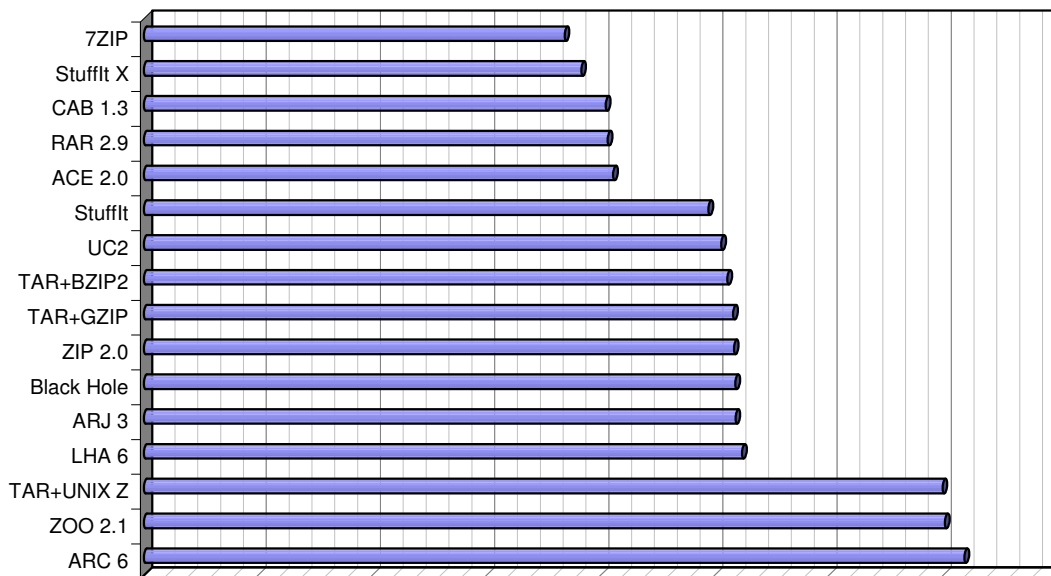
* Optimal/acceptable results are highlighted.

Note: 7ZIP and Stuffit X are slow and use a lot of memory with the setting I've chosen for the test.

**Relative compression ratio for plain text
(the shortest wins!)**



**Relative compression ratio for binary files
(the shortest wins!)**



**Overall relative compression ratio
(the shortest wins!)**

