

# File Compression Comparison

Adam Ferry, Nathan Kerns, Justin Niznik

## Table of Contents

Research Question .....	3
Methodology .....	3
Collected Data.....	4
WinZip ZIP Compression .....	4
WinRAR ZIP Compression .....	4
7-Zip ZIP Compression.....	5
Windows XP ZIP Compression .....	5
WinRAR RAR Compression .....	6
7-Zip 7-ZIP Compression .....	6
WinZip ZIP Compression (Using new JPG compression technology) .....	7
Findings.....	7
ZIP file format tests .....	7
Implications.....	7
ZIP, RAR, 7z File Format Findings.....	8
Implications.....	8
Scope and Limitations.....	8

## Research Question

The ZIP file format is one of the most commonly used formats in the grouping and compression of files. The basis of this paper is to determine which of the following free zip software utilities in the market place today does the best job in zip compression. The following criteria will be used in order to determine which utility does the best job such as file compression ratio, CPU load, and compression time. The second question that will be addressed is the capability of Win-Zip, WinRAR, and 7-Zip utilities when using their respective compression file formats of .ZIP, .RAR, and .7z. We will test how well the utility and format perform by using the same metrics as the ZIP format tests. More of the methodology will be explained further in the paper in addition to our recommendation on the best zip program to use under certain circumstances.

## Methodology

The way that we will be gathering data for our research question starts with first determining which ZIP utilities that will be used for our testing. Based upon market popularity, we have decided that we will test WinZip (v12), 7-Zip (v4.65), WinRAR (v1.7), and Windows XP's zip compression utility built into the operating system. The first three tools are the most commonly used tools in the market according to download.com's download figures<sup>1</sup>. The Windows XP ZIP compression tool was chosen because of its integration with Windows XP. We will use the ZIP file format in each of these programs in testing the compression of groups of files types. When testing the performance of the ZIP, RAR, and 7z compression formats, we will test each on its respective utility. We will compress a set of Audio Files consisting of MP3s equaling 497 MB , a 701 MB .AVI video file, picture files consisting of .JPGs equaling 500 MB, PDF files equaling 1.09 GB, and a 500 MB folder containing several different types of Microsoft Office file types.

The computer used for testing these utilities will be a Lenovo T500 with Windows XP as its operating system. The specs of the system are: Intel Core2 Duo P8400 @ 2.26 GHZ and 2 GB of DDR3 ram. Next, we will determine from each utility exactly how much each file was compressed from its original file size, how long it took to compress, and how much CPU usage it took to compress the file. We will use the default settings in each utility tested. Each individual test will be run three times for each utility and the average of the three tests will be reported.

To record the amount of time it takes to compress each type of file we will use the stopwatch function from an iPod touch and will use the performance monitor utility (perfmon.exe) in Windows XP in determining the average amount of CPU usage used for compression. With the performance monitor we will test the processor's usage percentage by measuring % processor time and the % user time. The % processor time measures "how much time the processor actually spends working on productive threads and how often it was busy servicing requests."<sup>2</sup> The % user time measures "the total amount of

---

<sup>1</sup> [http://download.cnet.com/WinZip/3000-2250\\_4-10003164.html](http://download.cnet.com/WinZip/3000-2250_4-10003164.html)

<sup>2</sup> <http://technet.microsoft.com/en-us/library/cc768048.aspx#XSLTsection130121120120>

non-idle time that was spent on User mode operations. This generally means application code.”

## Collected Data

### WinZip ZIP Compression

	WinZip JPGs	WinZip Avi	WinZip PDF	WinZip MP3	WinZip Docs
<b>Space savings (MB)</b>	2.37	8.04	39.65	16.72	89.99
<b>% Compressed</b>	0.47%	1.15%	3.91%	3.36%	17.98%
<b>Before Compression (MB)</b>	500.32	701.07	1,118.98	497.19	500.54
<b>After Compression (MB)</b>	497.95	693.02	1,075.20	480.47	410.55
<b>% Processor Average</b>	39.031	44.283	44.736	43.958	47.396
<b>% Processor Maximum</b>	49.219	54.688	53.906	52.344	55.469
<b>% User Time Average</b>	34.438	40.305	41.083	40.833	42.204
<b>% User Time Maximum</b>	39.844	49.219	45.313	45.313	49.219
<b>Time</b>	52.1 seconds	1min 3 sec	1 min 4 sec	44 sec	49 sec

### WinRAR ZIP Compression

	WinRAR JPGs	WinRAR Avi	WinRAR PDF	WinRAR MP3	WinRAR Docs
<b>Space savings (MB)</b>	2.36	8.03	39.49	16.68	89.73
<b>% Compressed</b>	0.47%	1.15%	3.53%	3.36%	17.93%
<b>Before Compression (MB)</b>	500.32	701.07	1,118.98	497.19	500.54
<b>After Compression (MB)</b>	497.96	693.04	1,079.50	480.51	410.81
<b>% Processor Average</b>	42.846	51.089	49.345	49.203	50.371
<b>% Processor Maximum</b>	54.688	57.813	53.125	60.769	56.250
<b>% User Time Average</b>	38.082	45.147	45.257	44.611	46.133
<b>% User Time Maximum</b>	47.656	50.781	50	50	50
<b>Time</b>	47.8 sec	1m 6 sec	1 min 48 sec	50 sec	41.3 sec

## 7-Zip ZIP Compression

	7-Zip JPGs	7-Zip Avi	7-Zip PDF	7-Zip MP3	7-Zip Docs
<b>Space savings (MB)</b>	2.11	6.83	39.34	15.40	90.31
<b>% Compressed</b>	0.42%	0.97%	3.52%	3.10%	18.04%
<b>Before Compression (MB)</b>	500.32	701.07	1,118.98	497.19	500.54
<b>After Compression (MB)</b>	498.21	694.24	1,079.64	481.79	410.23
<b>% Processor Average</b>	52.763	43.501	68.456	72.070	85.938
<b>% Processor Maximum</b>	71.094	50	96.875	97.656	99.219
<b>% User Time Average</b>	48.323	47.372	63.332	66.528	79.883
<b>% User Time Maximum</b>	63.281	53.906	92.969	92.969	92.188
<b>Time</b>	42 sec	1m 7 sec	1m 18 sec	33 sec	30 sec

## Windows XP ZIP Compression

	Win-XP Zip JPGs	Win-XP Zip Avi	Win-XP Zip PDF	Win-XP Zip MP3	Win-XP Zip Docs
<b>Space savings (MB)</b>	2.14	7.21	38.13	15.77	89.30
<b>% Compressed</b>	0.43%	1.03%	3.41%	3.17%	17.84%
<b>Before Compression (MB)</b>	500.32	701.07	1118.98	497.19	500.54
<b>After Compression (MB)</b>	498.18	693.86	1,080.86	481.42	411.24
<b>% Processor Average</b>	43.662	52.051	49.053	49.585	36.501
<b>% Processor Maximum</b>	56.25	66.938	62.5	57.813	52.344
<b>% User Time Average</b>	38.576	46.487	43.616	45.647	31.954
<b>% User Time Maximum</b>	46.875	51.563	50	50	48.438
<b>Time</b>	55.7 sec	1m 10 sec	1m 55 sec	50 sec	1m 4 sec

## WinRAR RAR Compression

	Rar JPGs	Rar Avi	Rar PDF	Rar MP3	Rar Docs
<b>Space savings (MB)</b>	1.51	7.42	47.53	18.75	94.49
<b>% Compressed</b>	0.30%	1.06%	4.25%	3.77%	18.88%
<b>Before Compression (MB)</b>	500.32	701.07	1118.98	497.19	500.54
<b>After Compression (MB)</b>	498.81	693.65	1,071.45	478.44	406.05
<b>% Processor Average</b>	83.562	93.924	90.949	91.162	83.681
<b>% Processor Maximum</b>	39.063	98.438	96.094	100	97.656
<b>% User Time Average</b>	69.287	84.635	81.282	83.247	73.682
<b>% User Time Maximum</b>	79.688	89.844	88.281	96.094	92.188
<b>Time</b>	2 m 42 sec	6 m 35 sec	9 m 53 sec	4m 16 sec	3m 29 sec

## 7-Zip 7-ZIP Compression

	7-Zip JPGs	7-Zip Avi	7-Zip PDF	7-Zip MP3	7-Zip Docs
<b>Space savings (MB)</b>	1.70	6.28	57.82	19.77	174.95
<b>% Compressed</b>	0.34%	0.90%	5.17%	3.98%	34.95%
<b>Before Compression (MB)</b>	500.32	701.07	1118.98	497.19	500.54
<b>After Compression (MB)</b>	498.62	694.79	1,061.16	477.42	325.59
<b>% Processor Average</b>	90.317	92.882	91.761	91.651	91.817
<b>% Processor Maximum</b>	96.094	100	99.219	98.438	100
<b>% User Time Average</b>	86.569	88.952	87.879	87.831	87.595
<b>% User Time Maximum</b>	96.094	95.313	96.094	96.094	99.219
<b>Time</b>	2 m 58 sec	4 m 8 sec	6 m 40 sec	2 m 52 sec	2 m 14 sec

## WinZip ZIP Compression (Using new JPG compression technology)

	WinZip JPGs
Space savings (MB)	109.99
% Compressed	21.98%
Before Compression (MB)	500
After Compression (MB)	390
% Processor Average	51.618
% Processor Maximum	58.594
% User Time Average	46.599
% User Time Maximum	53.125
Time	7 m 10 sec

## Findings

### ZIP file format tests

From our completed tests we found that WinZip and WinRAR are very similar when compressing files to ZIP format. Both have similar compression ratios, CPU usage, and compression speeds. However, of the two utilities WinZip does show that it generally performs faster and uses less CPU processing power. When compressing PDF files, WinRAR spent much more time compressing and was not as efficient at compressing either. The Windows XP ZIP file compressor utility performed similarly to WinZip and WinRAR but with slightly less compression, and significantly longer compression times.

The 7-ZIP utility performed slightly worse in the amount of file compression, and its CPU usage was a significant amount above the other utilities. For its two faults, it did however outperform the other utilities when speed was taken into account. The extra processing power was used to compress the files much more quickly than the other utilities.

### Implications

The implication of our findings is that WinZip is the overall best utility to use when compressing files to ZIP format. In a business environment, where enormous amounts of data are being compressed, WinZIP would show a clear advantage to WinRAR when measuring compression ratios, speeds, and CPU usage.

Conversely, if the user wishes to compress files to the ZIP format as quickly as possible without the regard for less compression and higher CPU usage, 7-Zip would be the best choice. However, this would not suit businesses very well because many times

systems need to run batch processes concurrently and 7-ZIP would take up a large amount of processing power, thus slowing down the other processes.

### **ZIP, RAR, 7z File Format Findings**

When comparing all three file formats it is necessary to look at each type of file that was compressed. When the .JPG files were compressed into .ZIP, they showed the highest amount of compression, much less CPU usage, and finished significantly earlier than the other compression formats. WinZip has recently released new JPG compression technology, which showed vast improvements over regular ZIP files and the other formats. The downside to using this technology is that the ZIP files produced can only be uncompressed with WinZIP 12.0. The .AVI file also showed significantly better performance in all three areas when compressed to ZIP format. 7z format was the best at compressing PDF files, ZIP was the fastest, and RAR took a considerably longer time to compress than the other two formats. MP3s were also compressed the most in the 7z format, but ZIP took the least amount of time once again. 7z also compressed Microsoft Word Documents drastically more than the other two file formats, exceeding them by 17%.

In terms of CPU usage the ZIP file format used the least amount of resources while the RAR and 7z format used a very high amount. Both the RAR and 7z formats took longer than the ZIP format.

### **Implications**

The 7z compression format is by far the best format when the amount of compression is of great importance. ZIP format delivered good compression, medium CPU usage, and took the least amount of time. RAR format did not seem to excel at anything.

In a business setting where compression is the most important aspect, the 7z format is the most preferable. If large amounts of documents and other file types are being compressed the 7z format will show the best performance. If the business wishes to balance compression with time and CPU usage, using the ZIP file format would be the best choice.

### **Scope and Limitations**

To accurately test the compression performance of each application the variables and conditions of this needed to be exactly the same at the start of each and every test. Being able to provide an environment where this was possible was challenging because of the ever changing conditions that a computer provides when in operation. We did our best to try and keep the environment as similar as possible for each test by trying to make sure that no extraneous applications were running in the background during our tests, and that the ones that were, were the same ones for each test run. We also made sure that we used the same computer for all of the testing and since it was on a laptop computer, we also made sure that it was plugged into an electrical outlet to insure maximum computing power. In addition, we were forced to use a manual stop watch in timing each utility as it

compressed and because of that there was the limitation of human error to be taken into account even though several tests were performed to average out any of this potential error. Another limitation was the software we used in order to test the amount of load that the CPU endured during our testing. It wasn't as complete as we had hoped in gathering detailed information about the current processing being done on the CPU in addition to giving a specific time frame of when the tests were taking place. We did our best to alleviate this possible error the same way as with the stop watch by taking an average over multiple tests.

As for the scope of the project, we tried to keep it within a limited set of options that a normal everyday user would encounter. There are several different settings and levels of compression that are available for each utility but to be able to test all of those settings and compression levels would be trivial when we are trying to determine what the best option would be for an everyday user. Therefore our main testing kept within the scope of using the .ZIP, compression file format and by also using the default settings for each of the utilities. We did use some of the other propriety formats (.RAR, and 7-ZIP) that some of the utilities offered for some comparison, but not to the level of testing that we did for just the .ZIP format. In addition, the types of files that were used in our testing consisted of files that someone might encounter on an everyday basis and did not get into larger more complex files that may benefit more from compression such as full hard drive backups or much larger uncompressed file types. This fact alone might have affected our results because we were trying to compress an already compressed file which is something that could be further studied in trying to improve the research being done. One last scope portion of the project to be taken into consideration was the overall file sizes of the test files we were compressing. We used general file sizes that we might expect the average user to be compressing and grouping together on an everyday basis. Had we more time and a more powerful computer, we really could have been able to test the full capacities and abilities of each compression utility and extraordinarily large data files.

Overall, given the scope and limitations of this project, we feel that we are able to determine from the data given the best options and recommendations when it comes to choosing a compression utility and for what types of use are most applicable. Our methods were well thought out and tested to the best of our abilities. While we can foresee the need for further testing and improvements that could be made with our methods, we feel that there is enough solid data and evidence created from our testing to provide the user with some good opinions when it comes to what file compression utility works the best.