

compression method for xlsx with 7z

Asked 10 years, 10 months ago Modified 8 months ago Viewed 7k times



Your privacy

By clicking "Accept all cookies", you agree Stack Exchange can store cookies on your device and disclose information in accordance with our [Cookie Policy](#).

Accept all cookies

Necessary cookies only

Customize settings



..Do some stuff..

```
7z a -tzip new ./decomp_xlsx/*
```

```
mv ./new.zip ./new.xlsx
```

From ECMA-376-2 Office Open Formats Part 2 (Packaging Conventions) The compression algorithm supported is DEFLATE, as described in the .ZIP specification. The package implementer shall not use any compression algorithm other than DEFLATE.

So, what switches do i need to use in 7z or other linux compatible program to get the job done without the warning? I've tried dropping the -tzip and using -m0=COPY, but excel can't even recover from that one.



guessing i'm not going to find a
n going to award that answer, and
ng. I'm not sure it handles the
5 / defF though.

Your privacy

By clicking "Accept all cookies", you agree Stack Exchange can store cookies on your device and disclose information in accordance with our [Cookie Policy](#).

Accept all cookies

Necessary cookies only

Customize settings

```

0 [Content_Types].xml
0 _rels/.rels
0 xl/_rels/workbook.xml.rels
0 xl/workbook.xml
0 xl/worksheets/sheet4.xml
0 xl/worksheets/_rels

0 xl/worksheets/_rels

0 xl/worksheets/sheet2.xml
0 xl/worksheets/sheet3.xml
00 xl/worksheets/sheet1.xml
00 xl/sharedStrings.xml

```

```
-rw---- 4.5 fat 1947 b- defS 80-Jan-01 00:00 xl/sharedStrings.xml
-rw---- 4.5 fat 22698 b- defS 80-Jan-01 00:00 xl/styles.xml
-rw---- 4.5 fat 7079 b- defS 80-Jan-01 00:00 xl/theme/theme1.xml
-rw---- 4.5 fat 220 b- defS 80-Jan-01 00:00 xl/prINTERSettings
/prINTERSettings2.bin
-rw---- 4.5 fat 464247 b- defS 80-Jan-01 00:00 xl/externalLinks
/externalLink1.xml
-rw---- 4.5 fat 338 b- defS 80-Jan-01 00:00 xl/externalLinks/_rels
/externalLink1.xml.rels
-rw---- 4.5 fat 220 b- defS 80-Jan-01 00:00 xl/prINTERSettings
/prINTERSettings1.bin
-rw---- 4.5 fat 593 b- defS 80-Jan-01 00:00 docProps/core.xml
-rw---- 4.5 fat 62899 b- defS 80-Jan-01 00:00 xl/calcChain.xml
-rw---- 4.5 fat 1031 b- defS 80-Jan-01 00:00 docProps/app.xml
20 files, 1392422 bytes uncompressed, 223675 bytes compressed: 83.9%
```

```
$ zipinfo new.xlsx
Archive: new.xlsx
Zip file size: 233180 bytes, number of entries: 20
-rw-r--r-- 3.0 unx 1031 tx defF 80-Jan-01 00:00 docProps/app.xml
-rw-r--r-- 3.0 unx 593 tx defF 80-Jan-01 00:00 docProps/core.xml
-rw-r--r-- 3.0 unx 62899 tx defF 80-Jan-01 00:00 xl/calcChain.xml
-rw-r--r-- 3.0 unx 464247 tx defF 80-Jan-01 00:00 xl/externalLinks
```

linux compression zip xlsx

Share Improve this question

edited Feb 24, 2013 at 17:21

asked Feb 22, 2013 at 21:39

Follow



jnewt

197 1 3 13

- 2 Run the unpack/pack with your modification in between. Then compare the "zip" contents with a diff tool. Are they really the same? Are there unintended changes? Is something missing? – [usr](#) Feb 22, 2013 at 21:41

```
7z x original.xlsx -o./original_decomp && 7z a -tzip new ./original_decomp/* && mv ./new.zip
./new.xlsx && 7z x new.xlsx -o./new_decomp && diff -r original_dec omp new_decomp && diff
original_decomp new_decomp This shows that the decompressed folders are identical, but the original / new
```

to-extract. – [Mark Adler](#) Feb 25, 2013 at



Your privacy

By clicking "Accept all cookies", you agree Stack Exchange can store cookies on your device and disclose information in accordance with our [Cookie Policy](#).

Accept all cookies

Necessary cookies only

[Customize settings](#)

Highest score (default)





8



For some strange reason, Microsoft is looking at the operating system encoding in the "version needed to extract" in the local file headers and central directory headers. It wants those to be zero, but 7z is setting them to 3 for Unix. If you are intent on using 7z, then you will need to patch the resulting file.

This program will do that:

```
/* needz.c - zero the operating system byte for "version needed to extract" in
the local and central headers of the zip files given on the command line.
Placed in the public domain by Mark Adler, 23 Feb 2013. */
```

```
#include <stdio.h>
#include <stdlib.h>

static void bail(char *why, char *what)
{
    fprintf(stderr, "needz error: %s%s\n", why, what);
    exit(1);
}

/* Read len bytes from offset as a little-endian integer. Negative offsets are
considered to be from the end of the file. */
static unsigned long peek(FILE *stream, off_t offset, int len)
{
    int ret, shift;
    unsigned long val;

    ret = fseeko(stream, offset, offset < 0 ? SEEK_END : SEEK_SET);
    if (ret)
        bail("not a zip file", "");
    val = 0;
    shift = 0;
    while (len-- > 0) {
        ret = getc(stream);
        if (ret == EOF)
            bail("not a zip file", "");
        val += (unsigned long)ret << shift;
        shift += 8;
    }
    return val;
}
```



```
ndian integer. Negative
e file. */
n, unsigned long val)
```

Your privacy

By clicking "Accept all cookies", you agree Stack Exchange can store cookies on your device and disclose information in accordance with our [Cookie Policy](#).

Accept all cookies

Necessary cookies only

Customize settings

```
END : SEEK_SET);
```

```
s assumes the classic zip
format (not Zip64), and no zip file comment. */
```

```

static void zip_zero_os(char *path)
{
    FILE *zip;
    unsigned entries;
    off_t central, local;

    zip = fopen(path, "r+b");
    if (zip == NULL)
        bail("could not open", path);
    if (peek(zip, -22, 4) != 0x06054b50)
        bail(path, " is not a zip file or has an end comment");
    entries = peek(zip, -12, 2);
    central = peek(zip, -6, 4);
    while (entries--) {
        if (peek(zip, central, 4) != 0x02014b50)
            bail(path, " has a structure error or is Zip64");
        poke(zip, central + 7, 1, 0);
        local = peek(zip, central + 42, 4);
        if (peek(zip, local, 4) != 0x04034b50)
            bail(path, " has a structure error or is Zip64");
        poke(zip, local + 5, 1, 0);
        central += 46 + peek(zip, central + 28, 2) +
            peek(zip, central + 30, 2) + peek(zip, central + 32, 2);
    }
    if (fclose(zip) == EOF)
        bail("could not close ", path);
}

int main(int argc, char **argv)
{
    while (--argc)
        zip_zero_os(++argv);
    return 0;
}

```

Share Improve this answer

Follow

edited Apr 11 at 13:56



cruizh

2,899 3 18 20

answered Feb 23, 2013 at 8:17



Mark Adler

105k 13 120 159



Your privacy

By clicking "Accept all cookies", you agree Stack Exchange can store cookies on your device and disclose information in accordance with our [Cookie Policy](#).

Accept all cookies

Necessary cookies only

[Customize settings](#)

well as a solution that would probably
 other questions, which I can move to
 ne with a pre packaged nix tool, if so
 sh (i'm calling a python script for my
 3, 2013 at 12:38

that. You can certainly convert the
 te, and seek on files. I don't think bash
 I don't know about. – [Mark Adler](#) Feb

es fix the problem, while still using 7z.
 it direct way. – [jnewt](#) Feb 25, 2013 at



You could use system `zip` and `unzip` instead. I often use something like the following.

4



Unzip the xlsx file to a directory:



```
$ unzip -o -d xlsx_dir Workbook1.xlsx
Archive:  Workbook1.xlsx
  inflating: xlsx_dir/[Content_Types].xml
  inflating: xlsx_dir/_rels/.rels
  inflating: xlsx_dir/xl/_rels/workbook.xml.rels
  inflating: xlsx_dir/xl/workbook.xml
  inflating: xlsx_dir/xl/sharedStrings.xml
  inflating: xlsx_dir/xl/theme/theme1.xml
  inflating: xlsx_dir/xl/styles.xml
  inflating: xlsx_dir/xl/worksheets/sheet1.xml
  extracting: xlsx_dir/docProps/thumbnail.jpeg
  inflating: xlsx_dir/docProps/core.xml
  inflating: xlsx_dir/docProps/app.xml
```

Then modify one or more of the XML files and re-zip them:

```
$ cd xlsx_dir

# Do something with the files like:
$ sed -i 's/Foo/Bar/' xl/sharedStrings.xml

$ find . -type f | xargs zip ../newfile.xlsx
```

The `find|zip` from within the directory isn't very pretty but it generates a file structure that matches the original without additional path stripping.

Share Improve this answer

edited Feb 24, 2013 at 0:48

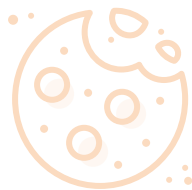
answered Feb 24, 2013 at 0:42

Follow



[jmcnamara](#)

39.2k 6 94 109



Your privacy

By clicking "Accept all cookies", you agree Stack Exchange can store cookies on your device and disclose information in accordance with our [Cookie Policy](#).

Accept all cookies

Necessary cookies only

[Customize settings](#)

I'm curious as to why it produces files maller than the original, yet when (es) shows they are the same. Is this 1:00

ression level. You could try the other Excel accepts all compression levels which would indicate that Excel doesn't atability with Excel it is better to focus matching the compression.

both are right. – [jnewt](#) Feb 25, 2013