

無圧縮ZIPアーカイブファイルの中を覗く

最近のWindowsでは圧縮フォルダとして扱われるZIPアーカイブファイルの中身を調べてみる。

ファイルフォーマット

以下のURLを参照しつつ中身を覗いてみる。

本家

<https://www.pkware.com/documents/casestudies/APPNOTE.TXT>

ZIPソフトウェア - 私的ZIPファイル研究所

<http://www.tnksoft.com/reading/zipfile/>

ZIP書庫ファイルフォーマット - 略して仮。

<http://www.tvg.ne.jp/menyukko/cauldron/dtzipformat.html>

巡回冗長検査

<http://ja.wikipedia.org/wiki/%E5%B7%A1%E5%9B%9E%E5%86%97%E9%95%B7%E6%A4%9C%E6%9F%BB>

FastHash

<http://hp.vector.co.jp/authors/VA033110/>

サンプルデータ

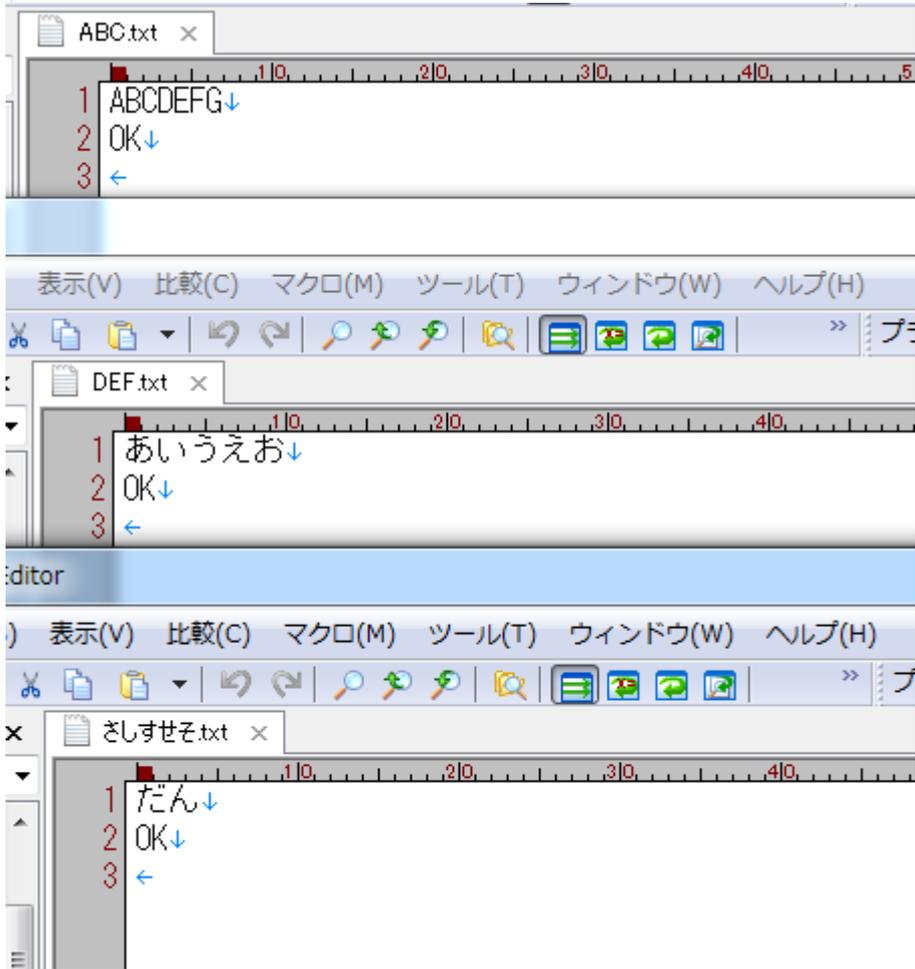
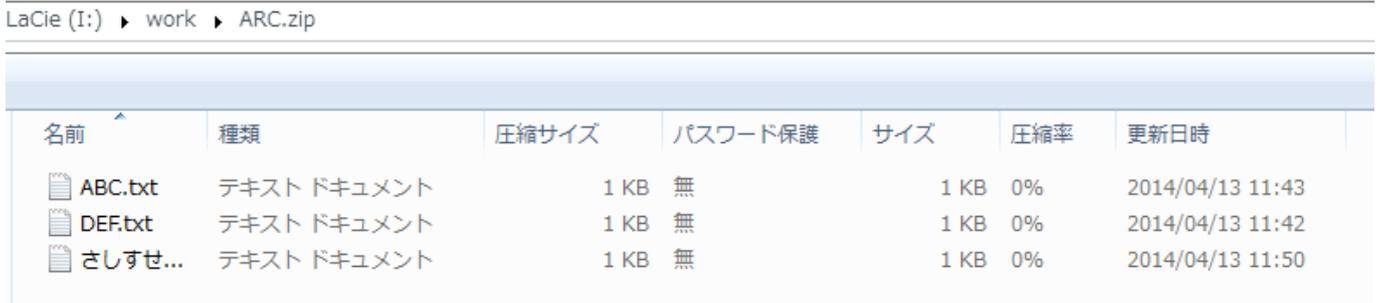
Windows7上で3つのファイルを圧縮格納したARC.zipを作り、その内容をバイナリエディタでダンプした。圧縮をさせないように小さなサイズのファイルを使っている。
使ったバイナリエディタはこちら。

Ag:Techsol - PowerWitch the Royal

<http://techsol.arcadiangarden.com/products/windows/pwtr.php>

このバイナリエディタ、起動時に会社で使うにはちょっと勇気が要るスプラッシュが出てくるかもしれない

以下は結果のキャプチャ画像。



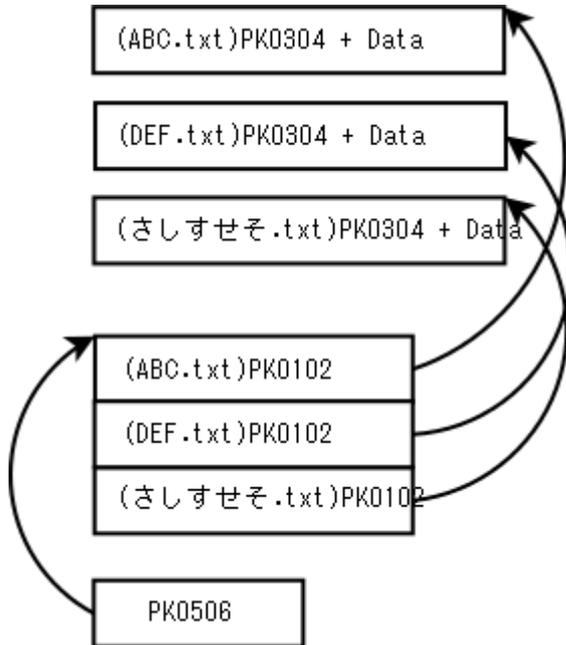
ダンプと構造確認

基本構造は ヘッダ(PKxxxx) データ、の繰り返し。

このサンプルだと、

1. PK0304(ABC.txt)
2. PK0304(DEF.txt)
3. PK0304(さしすせそ.txt)
4. PK0102(ABC.txt)
5. PK0102(DEF.txt)
6. PK0102(さしすせそ.txt)
7. PK0506

となっている。



フォーマットにあわせて整理するとこうなる。ヘッダの名称をここでは PKxxxx と呼ぶことにする。

"ABC.txt"のPK0304ヘッダ

- 00000000: 50 4B 03 04
- 00000004: 14 00 このZIPアーカイブファイルの展開に必要なZIPのバージョン。これは2.0を示す。
- 00000006: 00 00 オプション
- 00000008: 00 00 使用圧縮アルゴリズム。ゼロ(0)なら無圧縮。
- 0000000A: 6E 5D ← 時刻 hh:5bit + mm:6bit + ss:5bit □ssは実秒÷2の値
- 0000000C: 8D 44 ← 日付 YY:7bit + MM:4bit + DD:5bit □YYは1980からの経過年数
- 0000000E: 44 8F 7A 8D ← CRC32(圧縮前のファイルのCRC32)
- 00000012: 0D 00 00 00 ← 圧縮後のサイズ(13byte)
- 00000016: 0D 00 00 00 ← 圧縮前のサイズ(13byte)
- 0000001A: 07 00 ← ファイル名長さ(7byte)
- 0000001C: 00 00 ← コメントがあればその長さ。コメントはファイル名の直後に展開される。
- 0000001E: 41 42 43 2E 74 78 74 ← "ABC.txt"

```

00000025: 41 42 43 44 45 46 47 0D 0A 4F 4B 0D 0A ← "ABCDEFGH" \r\n "OK" \r\n
□圧縮されていれば圧縮後の内容

# "DEF.txt"のPK0304ヘッダ
00000032: 50 4B 03 04
00000036: 14 00
00000038: 00 00
0000003A: 00 00
0000003C: 55 5D
0000003E: 8D 44
00000040: 26 0E 50 21
00000044: 10 00 00 00
00000048: 10 00 00 00
0000004C: 07 00
0000004E: 00 00
00000050: 44 45 46 2E 74 78 74 ← "DEF.txt"
00000057: 82 A0 82 A2 82 A4 82 A6 82 A8 0D 0A 4F 4B 0D 0A ← "あいうえお" \r\n
"OK" \r\n □圧縮されていれば圧縮後の内容

# "さしすせそ.txt"のPK0304ヘッダ
00000067: 50 4B 03 04
0000006B: 14 00
0000006D: 00 00
0000006F: 00 00
00000071: 4D 5E 8D 44
00000075: E9 AF
00000077: 26 4E
00000079: 0A 00 00 00
0000007D: 0A 00 00 00
00000081: 0E 00
00000083: 00 00
00000085: 82 B3 82 B5 82 B7 82 B9 82 BB 2E 74 78 74 ← "さしすせそ.txt"
00000093: 82 BE 82 F1 0D 0A 4F 4B 0D 0A ← "だん" \r\n "OK" \r\n
□圧縮されていれば圧縮後の内容

# "ABC.txt"のPK0102ヘッダ
0000009D: 50 4B 01 02
000000A1: 14 00 ← このZIPアーカイブファイルの作成で使したZIPのバージョンとOS□これは2.0□MS-DOSを示す。
000000A3: 14 00 ← このZIPアーカイブファイルの展開に必要なZIPのバージョン。これは2.0を示す。
000000A5: 00 00 ← オプション
000000A7: 00 00 ← 使用圧縮アルゴリズム。ゼロ(0)なら無圧縮。
000000A9: 6E 5D ← 時刻 hh:5bit + mm:6bit + ss:5bit □ssは実秒÷2の値
000000AB: 8D 44 ← 日付 YY:7bit + MM:4bit + DD:5bit □YYは1980からの経過年数
000000AD: 44 8F 7A 8D ← CRC32(圧縮前のファイルのCRC32)
000000B1: 0D 00 00 00 ← 圧縮後のサイズ(13byte)
000000B5: 0D 00 00 00 ← 圧縮前のサイズ(13byte)
000000B9: 07 00 ← ファイル名長さ(7byte)
000000BB: 00 00 ← 拡張フィールド長さ。拡張フィールドはファイル名の直

```

後に展開される。

000000BD: 00 00

← コメントがあればその長さ。コメントは拡張フィールド

の直後に展開される。

000000BF: 00 00

← 分割されている場合、対応するPK0304ヘッダが格納さ

れたパートの番号。分割していないならゼロ(0)

000000C1: 01 00

← 対応するPK0304に格納したファイルの属性情報。これ

はテキストファイルを示す。

000000C3: 20 00 00 00

← ZIPアーカイブファイルを作成したOSで保持していた、

対象ファイルの属性情報。

000000C7: 00 00 00 00

← "ABC.txt"のPK0304ヘッダの位置

000000CB: 41 42 43 2E 74 78 74

← "ABC.txt"

"DEF.txt"のPK0102ヘッダ

000000D2: 50 4B 01 02

000000D6: 14 00

000000D8: 14 00

000000DA: 00 00

000000DC: 00 00

000000DE: 55 5D

000000E0: 8D 44

000000E2: 26 0E 50 21

000000E6: 10 00 00 00

000000EA: 10 00 00 00

000000EE: 07 00

000000F0: 00 00

000000F2: 00 00

000000F4: 00 00

000000F6: 00 00

000000F8: 20 00 00 00

000000FC: 32 00 00 00

← "DEF.txt"のPK0304ヘッダの位置

00000100: 44 45 46 2E 74 78 74

← "DEF.txt"

"さしすせそ.txt"のPK0102ヘッダ

00000107: 50 4B 01 02

0000010B: 14 00

0000010D: 14 00

0000010F: 00 00

00000111: 00 00

00000113: 4D 5E

00000115: 8D 44

00000117: E9 AF 26 4E

0000011B: 0A 00 00 00

0000011F: 0A 00 00 00

00000123: 0E 00

00000125: 00 00

00000127: 00 00

00000129: 00 00

0000012B: 00 00

0000012D: 20 00 00 00

00000131: 67 00 00 00

"さしすせそ.txt"のPK0304ヘッダの位置

00000135: 82 B3 82 B5 82 B7 82 B9 82 BB 2E 74 78 74 ← "さしすせそ.txt"

```
# PK0506ヘッダ
00000143: 50 4B 05 06
00000147: 00 00          分割していなければゼロ(0)。分割時はこのパートの番号。(1~65535)
00000149: 00 00          分割していなければゼロ(0)。分割時は最初のPK0304が格納されたパートの番
号。(1~65535)
0000014B: 03 00          ← 分割時はこのパートに格納されているファイル数。分割して
いないなら以下と同じ数。
0000014D: 03 00          ← ZIPアーカイブファイル全体で格納されているファイルの数
0000014F: A6 00 00 00  ← PK0102ヘッダ合計サイズ □ 0x35(ABC.txt) +
0x35(DEF.txt) + 0x3c(さしすせそ.txt) = 0xa6
00000153: 9D 00 00 00  ← PK0102ヘッダの開始位置
00000157: 00 00          コメントの長さ。コメントはPK0506ヘッダの直後に展開される。

00000159: A3 00 00 00 00 00 ←ゴミ？
```

各ヘッダ内の数値はリトルエンディアンとなっているので、計算時はx86アーキテクチャでの表現に合わせる。

```
□LB      MB □ MB      LB
□A6 00 00 00 → 00 00 00 A6
□E9 AF 26 4E → 4E 26 AF E9
□
□LB MB      MB LB
□03 00      → 00 03
```

ZIPアーカイブファイルはアーカイブ媒体が分割される場合を想定している□FDやテープが使われていた頃に生まれたものなので、現状、わざわざZIPアーカイブファイルを分割することはまずないので、単一ファイル(単一媒体)として扱う。

無圧縮のため、ファイルの圧縮前後の値が同一。異なれば圧縮が行われているとわかる□CRC32は圧縮前のファイルのCRC32を計算したもので、圧縮を解除した後正しく復元できたかをこれで検証している模様。

PK0304ヘッダとPK0102ヘッダでは重複した値がある。おそらく歴史的理由、という奴であろうと推測する。

PK0506ヘッダの情報から推測すると、

- PK0102ヘッダは連続して並んでいる必要がある。
- PK0304ヘッダ□PK0102ヘッダはPK0506ヘッダのあとでもかまわない。

はず。

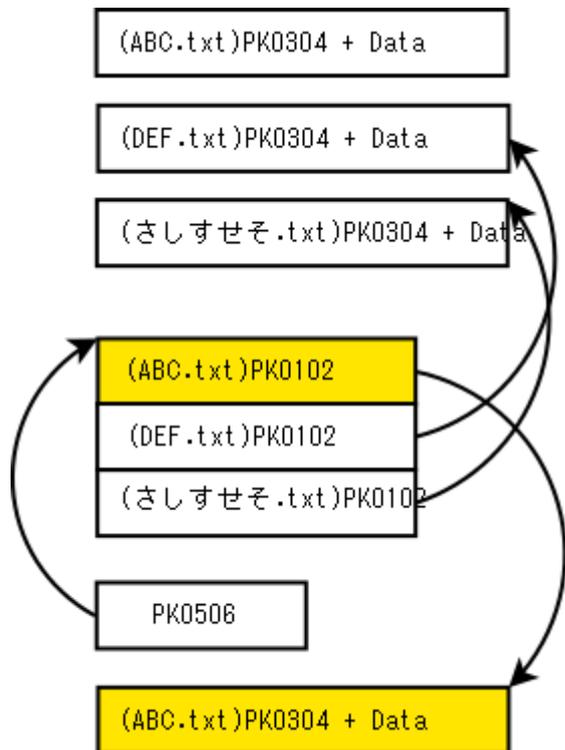
PK0506ヘッダ直下、ロケーション 0x159以下の値は何だろう。何度か作るとつかない場合もあるし、値が異なる場合もある□7-Zipだとつかない。

ファイル名に関し、SHIFT-JISのコードがビッグエンディアンで出力されている。おそらくマルチバイトな文字が使われる事を想定していないし規定もないためと推測される。ファイル名にマルチバイト文字を避けたほうが良い理由はこの辺りにあるのかもしれない。

実験□ PK0506ヘッダの下にPK0304ヘッダを配置する

ARC.zip をコピーした ABC2.zipを改変□PK0506ヘッダのあとにPK0304ヘッダとデータをを追加する。

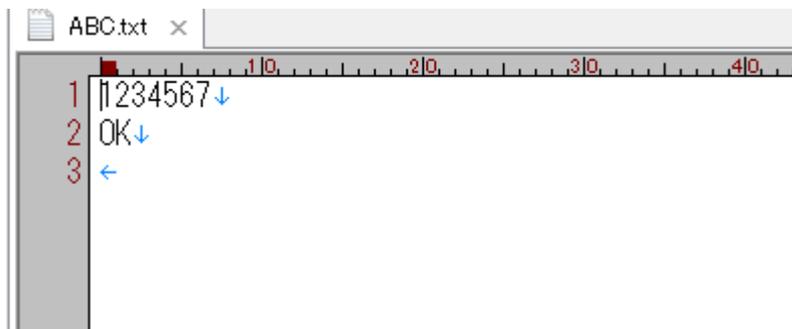
The screenshot shows the PowerWitch the Royal hex editor interface. The main window displays a hex dump of a file named 'ABC2.zip'. The columns represent hex addresses and bytes, with corresponding ASCII characters shown on the right. A red arrow points to the hex value '00' at address 00000130. Several rows of hex data are highlighted in yellow, including addresses 00000080, 00000090, 000000A0, 000000B0, 000000C0, 000000D0, 000000E0, 000000F0, 00000100, 00000110, 00000120, 00000130, 00000140, 00000150, 00000160, 00000170, and 00000180. The ASCII view shows various characters and file signatures like 'PK', 'DEF.txt', and 'ABC.txt'. The status bar at the bottom shows 'メモリ利用率 9%'.



追加のPK0304ヘッダで

- ABC.txtの中身を “ ABCDEFG\r\nOK\r\n”から“1234567\r\nOK\r\n”に置き換え
- ABC.txt対応ヘッダのCRC32書き換え
- ABC.txt対応PK0102ヘッダにあるPK0304ヘッダロケーションを追加のPK0304ヘッダロケーションに書き換え

し、解凍してABC.txtを開いてみる。



ABC.txtの内容が置き換わっていることを確認できた。つまりWindows圧縮フォルダはPK0506ヘッダ以降も構造を認識していることがわかる。

実験□ PK0506ヘッダの下に改変PK0506ヘッダを追加する

ARC.zip をコピーした ABC3.zipを改変□PK0506ヘッダのあとにPK0102ヘッダをひとつ飛ばしたPK0506ヘッダを追加する。

I:\work\ABC3.zip - PowerWitch the Royal

ファイル(F) 編集(E) 検索(S) 表示(V) 比較(C) ウィンドウ(W) ツール(T) ヘルプ(H)

ADDRESS	+0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+A	+B	+C	+D	+E	+F	SUM	Shift-JIS	
00000000	50	4B	03	04	14	00	00	00	00	00	00	6E	5D	8D	44	44	8F	25	PK n]好D循
00000010	7A	8D	0D	00	00	00	0D	00	00	00	00	07	00	00	00	41	42	AB	· AB
00000020	43	2E	74	78	74	41	42	43	44	45	46	47	0D	0A	4F	4B	5E	C.txtABCDEFG OK	
00000030	0D	0A	50	4B	03	04	14	00	00	00	00	00	55	5D	8D	44	50	PK U]好	
00000040	26	0E	50	21	10	00	00	00	10	00	00	00	07	00	00	00	CC	& P!	
00000050	44	45	46	2E	74	78	74	82	A0	82	A2	82	A4	82	A6	82	73	DEF.txtあいうえお	
00000060	A8	0D	0A	4F	4B	0D	0A	50	4B	03	04	14	00	00	00	00	26	OK PK	
00000070	00	4D	5E	8D	44	E9	AF	26	4E	0A	00	00	00	0A	00	00	9C	M^好魅&N	
00000080	00	0E	00	00	00	82	B3	82	B5	82	B7	82	B9	82	BB	2E	59	さしすせそ.	
00000090	74	78	74	82	BE	82	F1	0D	0A	4F	4B	0D	0A	50	4B	01	77	txtだん OK PK	
000000A0	02	14	00	14	00	00	00	00	00	6E	5D	8D	44	44	8F	7A	13	n]好D循	
000000B0	8D	0D	00	00	00	00	00	00	00	07	00	00	00	00	00	00	AE	·	
000000C0	00	01	00	20	00	00	00	00	00	00	00	41	42	43	2E	74	89	ABC.t	
000000D0	78	74	50	4B	01	02	14	00	14	00	00	00	00	55	5D	64	xtPK U]		
000000E0	8D	44	26	0E	50	21	10	00	00	10	00	00	00	07	00	9D	好& P!		
000000F0	00	00	00	00	00	00	00	20	00	00	00	32	00	00	00	00	52	2	
00000100	44	45	46	2E	74	78	74	50	4B	01	02	14	00	14	00	00	23	DEF.txtPK	
00000110	00	00	00	4D	5E	8D	44	E9	AF	26	4E	0A	00	00	00	0A	9C	M^好魅&N	
00000120	00	00	00	0E	00	00	00	00	00	00	00	00	00	20	00	00	2E		
00000130	00	67	00	00	00	82	B3	82	B5	82	B7	82	B9	82	BB	2E	B2	g さしすせそ.	
00000140	74	78	74	50	4B	05	06	00	00	00	00	03	00	03	00	A6	B2	txtPK 7	
00000150	00	00	00	9D	00	00	00	00	00	50	4B	05	06	00	00	00	43	· PK	
00000160	00	02	00	02	00	71	00	00	00	D2	00	00	00	00	00	00	47	q x	
00000170	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	00		
00000180	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	00		
00000190	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	00		
000001A0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	00		
000001B0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	00		
000001C0	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	00		

EC 43 76 79 CA E4 C9 85 2F E5 22 3F D4 49 E1 3A

Address 0000016Fh (367) Bit ----,----

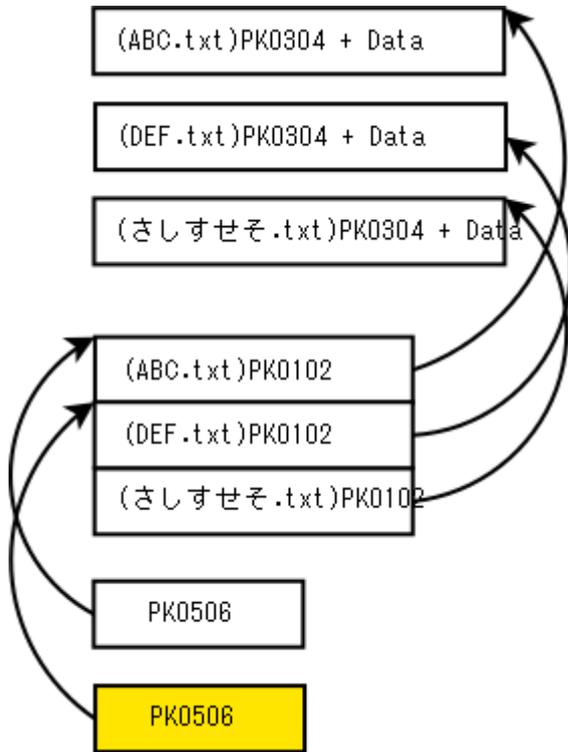
File Size 0000016Fh (367) Byte --h ()

Range Size 00000000h (0) Word ----h ()

Double Word -----h ()

ヘルプ 一括設定 検索 置換 ジャンプ 選択 窓選択 メニューバー 開く 保存

変更 上書 単純 拡張 ▶ メモリ利用率 8%



追加のPK0506ヘッダで

- PK0102ヘッダ合計サイズを “ DEF.txt”のPK0102ヘッダと “ さしすせそ.txt”のPK0102ヘッダ合計サイズに書き換え
- PK0102ヘッダ開始ロケーションを “ DEF.txt”のPK0102ヘッダロケーションに書き換え
- PK0102ヘッダ個数を2個に変更

しABC3.zipを解凍してみる。

名前	種類	圧縮サイズ	パスワード保護	サイズ	圧縮率	更新日時
DEF.txt	テキストドキュメント	1 KB	無	1 KB	0%	2014/04/13 11:42
さしすせそ.txt	テキストドキュメント	1 KB	無	1 KB	0%	2014/04/13 11:50

ABC.txtの内容が見えなくなった。つまりWindowsの圧縮フォルダでは最後に現れたPK0506が有効になる事がわかる。なお7-Zipでは壊れたアーカイブと判定されてしまうPK0506はひとつだけ存在させるのが正解かもしれない。

この情報を使った嘘吐きなzipコマンドを作ってみた。

技術資料, ZIP

From:
<https://wiki.hgotoh.jp/> - 努力したWiki

Permanent link:
<https://wiki.hgotoh.jp/documents/other/memo01/other-017>

Last update: 2025/11/20 09:34



