# 次世代アーカイバソフト(圧縮形式名ｙｚ２)の開発

Development of next-generation archiver-software( lossless data compression type name "yz2" )

山崎　敏 [1)]
Satoshi YAMAZAKI

1)株式会社 日立インフォメーションテクノロジー
（〒259-0157 神奈川県足柄上郡中井町境４５６番地　E-mail: syamaza@hitachi-it.co.jp ）

**ABSTRACT.** Nowadays, The Internet is using widely and the amount of information is increasing by leaps and bounds, so it will be very important term that "transmitting / saving data (contents, software, etc.) at the high speed and high compression rate". The LZH compression form that is used widely in Japan and the ZIP compression form that is used widely in the world are developed ten or more years ago, so new high performance compression form is strongly desired. The author developed the "yz2" compression form which uses Range Code coding process at entropy process and optimized dictionary structure. The author also compared the performance of the "yz2", the LZH, and the ZIP form, and verified that the "yz2" compression form is achieved the high performance exceeding the existing compression form at every compression rate, compressing speed, and decompressing speed.

## １．背景

インターネットの時代に突入した今、「いかにして高速にかつ効率良くコンテンツやソフトウェアを転送/保存するか」が重要な課題となっている．

この課題に対する 1 つの解答として、アーカイバソフト(ファイル圧縮)がある．現在の日本では、日本製の LZH 圧縮形式がもっとも広く使われている．また、世界では ZIP 圧縮形式が広く使われている．しかし、LZH 圧縮形式/ZIP 圧縮形式ともに開発されてから十数年が経過しており、新しい圧縮形式が望まれている．そのことを裏づけるように近年、新たな圧縮形式がいくつかデビューしている．

## ２．目的

そこで今回の開発では、LZH 圧縮形式に代る次世代の日本製のアーカイバソフト(圧縮形式名 yz2)を開発する．主な目的は、yz2 圧縮形式をアーカイバソフトのデファクトスタンダードにすることである．日本国内はもとより，全世界でのデファクトスタンダードになることを目的とする．

yz2 圧縮形式は、提案者(山崎)が 1997 年に独自に開発した yz1 圧縮形式に改良を加え、さらなる圧縮/解凍速度の向上と圧縮率の向上を図る形式である．

## ３．開発の概要

### （１） 開発方針

提案者（山崎）が 1997 年に独自に開発した、yz1 形式に改良を加え、さらに圧縮速度/解凍速度の向上と圧縮率を向上させる．yz1 形式は、LZ 系の圧縮アルゴリズムに独自の（辞書の構造を新たに考案し）改良を加えることで、従来の LZ 系に比べて主に大幅な高速化と高い圧縮率を実装したアルゴリズムを適用した形式である．現在の yz1 形式は、最後のエントロピー符号化部分を独自の Huffman 符号化によって行っている．これを yz2 形式では、最新の疑似算術符号形式である RangeCode 符号 に変更/対応することで、速度の向上および圧縮率の向上を狙う．

また、Huffman 符号化に最適化している現在の辞書構造の見直しを行い、RangeCode 符号への最適化も試みる．この辞書構造の見直しが圧縮速度および圧縮率どの程度有効であるか不明であるが、 Huffman 符号から RangeCode 符号 への変更が圧縮速度および圧縮率に対して有効であることはすでに実測して確認済みである．

### （２）用語の解説

用語の解説を表１に示す．

**表１ 用語の解説**

| No. | 用語 | 説明 |
|---|---|---|
| 1 | アーカイバ | 複数のファイルを１つにまとめることでファイルの利便性を向上させるためのソフトウェア．UNIX 系では、tar 形式が有名．ただ、tar 形式ではファイルの圧縮は行わない．Windows/PC 系では LZH,ZIP 圧縮形式が有名．<br>アーカイバソフト全般に関しては「統合アーカイバ・プロジェクト」が詳しい．<br>→ http://www.csdinc.co.jp/archiver/ |
| 2 | LZH 圧縮形式 | 日本国内でスタンダードなアーカイバ/圧縮形式．<br>LZH 形式の歴史に関しては「奥村晴彦のホームページ」が詳しい．<br>→ http://www.matsusaka-u.ac.jp/~ okumura/ |
| 3 | ZIP 圧縮形式 | 海外（世界的）でスタンダードなアーカイバ/圧縮形式．<br>圧縮アルゴリズムに関しては「comp.compression Newsgroup FAQs（英語）」が詳しい．<br>→ http://www.faqs.org/faqs/by-newsgroup/comp/comp.compression.html |
| 4 | yz1 圧縮形式 | 開発者(山崎)が 1997 年に公開した独自の圧縮形式．<br>LZH,ZIP 圧縮形式より、圧縮速度が高速であるのが特徴．圧縮率も良い．<br>yz1 形式 に関してはアーカイバソフト「DeepFreezer」を参照のこと．<br>また DeepFreezer に関しては提案者（山崎）のホームページ「やまざき@ BinaryTechnology のページ」が詳しい．<br>→ http://member.nifty.ne.jp/yamazaki/ |
| 5 | Huffman 符号/算術符号 | エントロピー符号のアルゴリズム一種．<br>Huffma 符号は古くから知られているスタンダードな符号化アルゴリズム．<br>算術符号もわりと古くから知られている符号化アルゴリズムではあるが、多くの特許が取られているためあまりスタンダードではない．<br>エントロピー符号のアルゴリズムに関しては「comp.compression Newsgroup FAQs（英語）」が詳しい．<br>→ http://www.faqs.org/faqs/by-newsgroup/comp/comp.compression.html |
| 6 | RangeCode 符号 | エントロピー符号のアルゴリズム一種．<br>算術符号に似た符号化アルゴリズム．算術符号より高速で特許も取られていないのが特徴．まだ新しい（論文:1979)アルゴリズムなのでそれほどスタンダードではない．<br>RangeCode 符号に関しては「Range encoder Homepage（英語）」が詳しい．<br>→ http://www.compressconsult.com/rangecoder/ |

**（３）システム構成**

**今回の開発では、以下の表２に示す２つのプログラムを作成した．**

**表２ 開発プログラム一覧**

| プログラム名 | 機能 | 概要 |
|---|---|---|
| yz2enc.exe | yz2 エンコーダ | ファイルやフォルダ（ディレクトリ）内のファイル群を圧縮し、１つのファイル（拡張子名 yz2）を作成するプログラム． |
| yz2dec.exe | yz2 デコーダ | yz2 ファイルを解凍し、圧縮前のファイル群に戻すためのプログラム． |

**また、これらのプログラムが動作するための動作環境を表３に示す．**

**表３ 動作環境**

| 項目 | 環境 |
|---|---|
| ソフトウェア | Windows 95 / 98 / NT / 2000 |
| ハードウェア | 上記 WindowsOS が動作する DOS/V 系 PC、推奨環境としては「CPU:300MHz 以上、RAM:32MB 以上、HDD:2GB 以上」 |

**（４）他の圧縮アルゴリズムとの比較**

**他の圧縮アルゴリズム LZH/ZIP と yz2 の内部構成の違いを表４と図１に示す．**

**表４ LZH/ZIP と yz2 の内部構成の比較**

| 圧縮形式名 | 辞書構成 | 符号化アルゴリズム |
|---|---|---|
| LZH/ZIP | LZ77 スライド辞書 | Huffman 符号 |
| yz2 | LZ77 スライド辞書 ×２５６個 | RangeCode 符号 |

**まず、エントロピー符号化アルゴリズムに関しては yz2 では RangeCode 符号を使っている．これは、Huffman 符号より、高速でかつ高圧縮である理由から選択している．次に辞書構成であるが、LZ77 のスライド辞書の考え方はそのまま使い辞書の数を２５６個に増やした．これにより、辞書検索速度が上がり、また、辞書への単語の登録数も増やしたため圧縮率も向上している．**

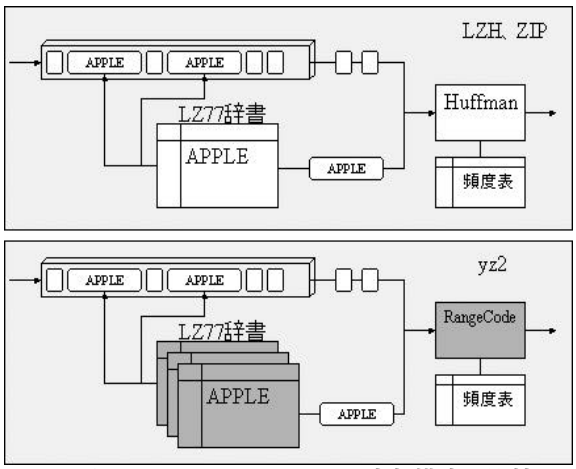


**図１　LZH/ZIP と yz2 の内部構成の比較**

## ４．開発の結果

開発終了時の結果と評価については、インターネット上で公開している．ここでは操作方法と画面の概要、そして関連するインターネットアドレスとを示す．

### （１）操作方法

フォルダ hoge を圧縮し、hoge.yz2a を作成するにはコマンドラインから

```
C:¥> yz2enc.exe -j hoge
```

と入力する。

hoge.yz2a を解凍し、hoge フォルダを作成するにはコマンドラインから

```
C:¥> yz2dec.exe -j hoge.yz2a
```

と入力する。

### （２）実行画面

開発したプログラムは Windows 用のコマンドラインプログラムである．それぞれの実行画面を図３、図４に示す．

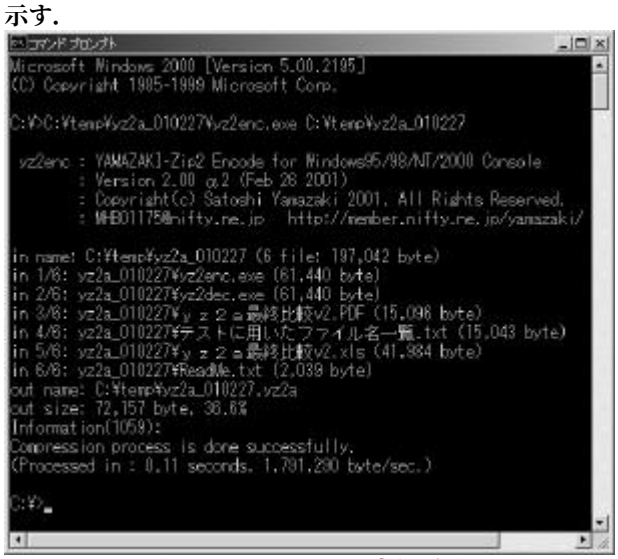


**図３　yz2enc.exe の実行画面**

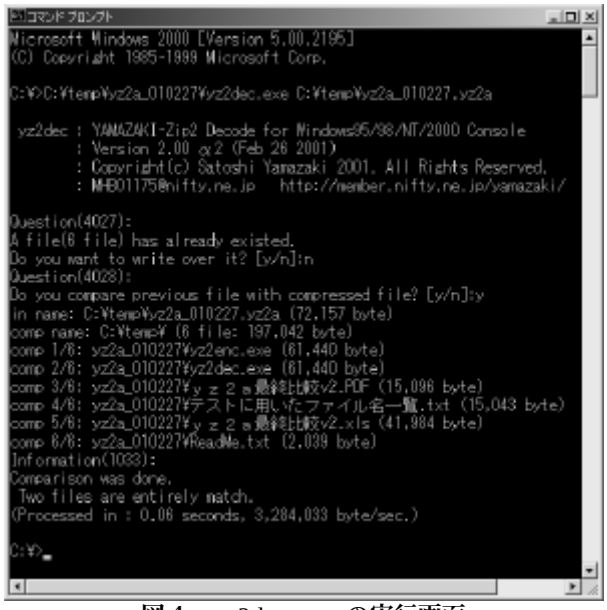


**図４　yz2dnc.exe の実行画面**

### （３）オープンソース

開発結果と評価結果の詳細については、表４に示すインターネット上のアドレスにて公開している．なお、開発終了時のプログラムも同じファイル内に入っている．また、プログラムのソースコードも同じアドレスにオープンソースとして公開している．このことは目的として上げていた「デファクトスタンダードへの道」の一歩を踏み出したと認識している．

**表４ インターネット上の公開アドレス**

| 項目 | アドレス/ファイル名 |
|---|---|
| URL | http://member.nifty.ne.jp/yamazaki/yz2/ |
| FileName | yz2_010227.yz1 (77,444 byte)<br>yz2_010227.zip (74,373 byte) |
| Source code | yz2source_010226.yz1 (185,578 byte)<br>yz2source_010226.zip (245,566 byte) |

### （４）性能

圧縮率、圧縮速度、解凍速度についての計測結果を LZH の性能を１００として ZIP の性能と yz2 の性能を比較したグラフを図５に示す．

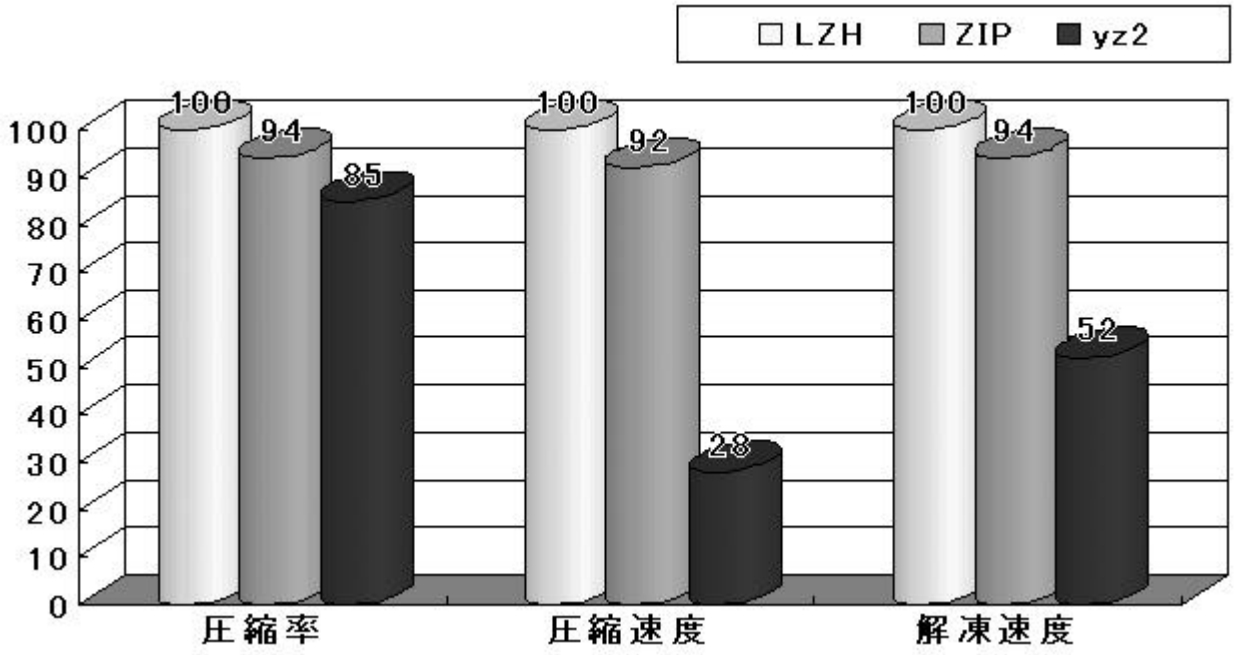


図５　性能の比較

このグラフからわかるように、yz2 圧縮形式は既存の LZH や ZIP 形式と比べ、圧縮率、圧縮速度、解凍速度全ての比較項目で上回っている．特に圧縮速度においては大幅に上回る性能であることを確認できた．

（５）まとめ

本開発において完成した yz2 圧縮アルゴリズムは、圧縮率に関しても、圧縮速度に関しても、良い結果を出すことができた．既存の広く使われているアーカイバと比較しても遜色の無い性能が出ていると思う．また、開発中にいくつかの改良すべき点を見つけたが、それらについては今後とも改良を続けていきたいと思っている．

## ５．参加企業及び機関

（株）日立インフォメーションテクノロジー

## ６．参考文献

［1］ 植松友彦／著, 「文書データ圧縮アルゴリズム入門」, ＣＱ出版, ISBN4-7898-3672-X

［2］ 韓 太舜、小林欣吾／著, 「情報と符号化の数理」, 培風館, ISBN4-563-00599-1