

次世代アーカイバ(yz2)の拡布と市販ソフトへの適用

Next-generation archiver("yz2") is spread and application to commercial software.

山崎 敏¹⁾
Satoshi YAMAZAKI

酒井 智巳²⁾
Tomomi SAKAI

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4 5 6 番地

2) 有限会社 エレクトリックシーブ
〒463-0037 名古屋市守山区天子田 3-302 ロイヤルタウン天子田 201

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 RangeCode 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

「IT革命」といわれたこの数年で非常に大きく変わったことは、デジタルデータやデジタルコンテンツの爆発であるといえる。これらの多量のデータ群を「いかに効率よく保存し、また高速に転送するか？」という問題は、現在も、常に大きく問われつづけている。

画像や音声などの不可逆なデータに関しては JPEG や MP3 などの強力でスタンダードな圧縮形式が開発されている。しかし、一方、ユニバーサルな可逆データの圧縮形式に関しては ZIP や LZH などの十数年前の古い技術をいまだに使い続けているのが現状である。

そこで、ZIP や LZH などの古い圧縮形式に変わる新しい(高速で高圧縮率の)圧縮アルゴリズムを採用した yz2 形式を開発すべく、前年度の「H12 未踏ソフト創造事業」に対して「次世代アーカイバソフト(圧縮形式名 yz2)の開発」を提案した。その後、提案は採択され、yz2 形式のアーカイバソフトは無事に開発を完了した。

参考: yz2 に関しては
<http://member.nifty.ne.jp/yamazaki/yz2/index.html> を参照のこと。

参考: ZIP, LZH など、汎用アーカイバソフトに関しては <http://csdinc.co.jp/archiver/> を参照のこと。

2. 目的

「H12 未踏ソフト創造事業」にて開発した yz2 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。yz2 形式がデファクトスタンダードになるためには、労力や費用を必要とせず、yz2 形式が自らのもつポテンシャルによって広がっていける状態に押し上げることが必要となる。この「押し上げ」が本提案の核となる目的である。

より具体的には「yz2 形式の Windows 版 DLL (yz2.dll)の開発」「サンプルGUIプログラム (DeepFreezer2)の開発」「市販ソフト(ギズモ I.Q.)への組み込み販売」の3つの項目になる。

この3つの項目を達成することができれば、yz2 形式がデファクトスタンダードへの道へ踏み出し、将来的には、多くの人が公平に利用できるアーカイバ(圧縮アルゴリズム)を手にすることができ、IT革命による本当の意味の恩恵を受けることができると考えている。

参考: DeepFreezer に関しては、
<http://member.nifty.ne.jp/yamazaki/DeepFreezer/index.html> を参照のこと。

3．開発の概要

(1) 開発方針

(a) yz2 形式の改良と Windows 用 DLL(yz2dlg.dll)の開発

- ・ Windows 用 DLL (yz2dlg.dll) の作成
- ・ 圧縮率の向上 (PPM アルゴリズムの適用)
- ・ 圧縮速度の向上 (マルチスレッドによる並列処理の適用)
- ・ Linux (gcc)への対応

(b) サンプル GUI プログラム (DeepFreezer2) の開発

(c) 市販ソフト (ギズモ I.Q.) への組み込み販売

(2) 成果物

今回の開発では、以下の表 1 に示す 2 つのプログラムを作成した。

表 1 開発プログラム一覧

プログラム名	概要
DeepFreezer 2	圧縮・展開プログラム 一般にアーカイバと呼ばれる種類のソフトウェア Windows 用の GUI を持つ。フリーソフト。
ギズモ I.Q.	プログラムを組んで対戦する形のゲーム。Windows 用の市販ソフトとして販売される。

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 2 に示す。

表 2 動作環境

項目	環境
ソフトウェア	Windows 98 / 2000
ハードウェア	上記 Windows98/2000 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: 1 GB 以上」

(3) yz2 アルゴリズムの内部構成

yz2 の内部構成の LZH/ZIP と比較して図 1 に示す。
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。

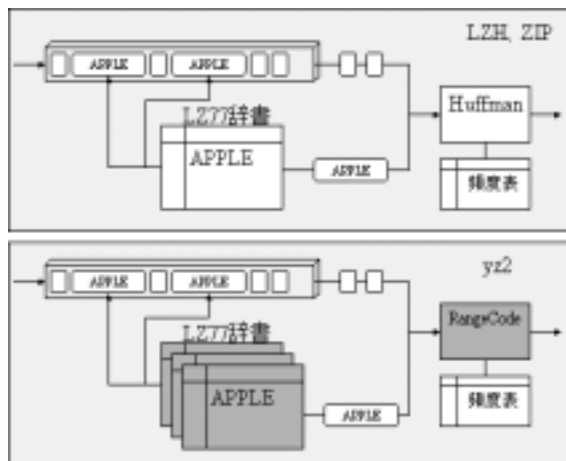


図 1 LZH/ZIP と yz2 の内部構成の比較

(4) yz2 アルゴリズムの改良概略

yz2 の圧縮率の向上として PPM のアルゴリズムを適用した。概略図を図 2 に示す。

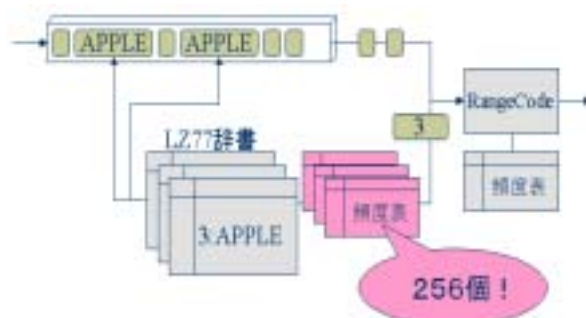


図 2 PPM アルゴリズムの適用図

yz2 の圧縮速度の向上としてマルチスレッドのアルゴリズムを適用した。概略図を図 3 に示す。

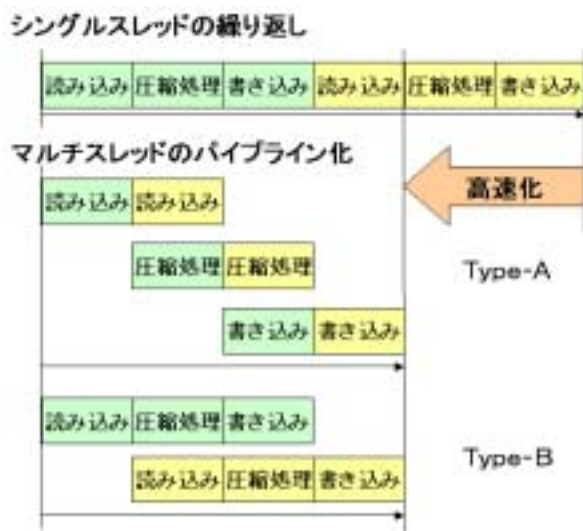


図 3 マルチスレッドの適用図

4．開発の結果

開発終了の結果として、実行画面、インターネット上で公開場所、自己評価を示す。

(1) 実行画面

DeepFreezer2 と Gizmo.I.Q.の実行時の画面を図3、図4に示す。



図3 DeepFreezer2 の実行画面

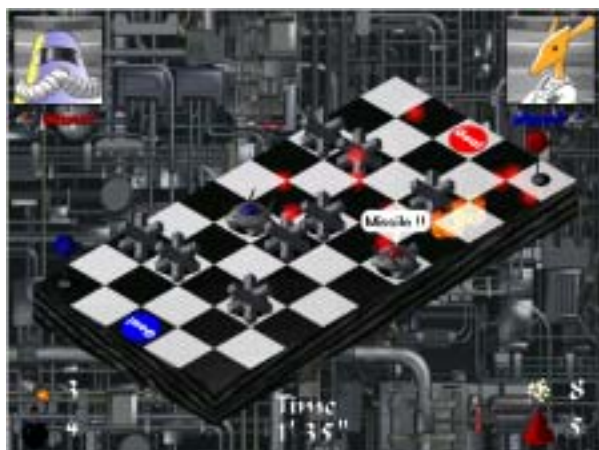


図4 Gizmo.I.Q. の実行画面

(2) オープンソース

DeepFreezer2 のプログラムとソースファイルについては、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。

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	yz2source_2002_02_21.yz1 (2002/01/13 版) [1.82 MB (1,917,216 バイト)]
	yz2source_2002_02_21.zip (2001/02/21 版) [2.50 MB (2,625,181 バイト)]

ギズモ I.Q.は CD-ROM の形態で市販する。有限会社エレクトリックシープの
Website <http://www.electricsheep.co.jp/>
にて通信販売される予定。

(3) 評価

(1-1) yz2 形式の改良と Windows 用 DLL(yz2dlg.dll)の開発

・ Windows 用 DLL (yz2dlg.dll) の作成

DLL の名称は yz2dlg.dll へと変更したが開発は完了している。今後の課題としては、ユーザの意見を吸収し、できるかぎり長い間サポートしていきたい。

・ 圧縮率の向上 (PPMアルゴリズムの適用)

PPM のアルゴリズムも適用済みである。ただし、結論としては速度は遅くなるだけで、肝心の圧縮性能は上がらなかった。この原因については、PPMアルゴリズムによる効果は出ているものの頻度表の数を増やしたことによって頻度表が適正な値に更新されるまでに多くのデータが必要になってしまったことが圧縮率にマイナスに働いたためと思われる(圧縮するデータ量が多くなれば圧縮率が良くなっている数値からの判断)。今後の課題としては、頻度表の初期値を一定比率ではなく、圧縮データの内容を反映した値で初期化できるように頻度データを渡す方法を検討すべきと思われる。

・ 圧縮速度の向上 (マルチスレッドによる並列処理の適用)

マルチスレッドのコードも適用済みである。当初、中間発表の時点では期待したほどの性能が得られなかったが、その後、マルチスレッドの処理方法を2種類実装して比較したり、バッファサイズを可変にしたり、複数のマシンで実測を重ねたところ、バッファサイズが小さく、対象ファイル数が比較的多い時にはマルチスレッドによる並列処理の効果が発揮されることが確認できた。

・ Linux (gcc)への対応

yz2source_011116.yz1 (2001/11/16 版) alpha4 にて Linux に対応し公開した。その後、最新の Linux への適用は進んでいない。努力目標ではあったが、時間が足りなくなってしまったのは残念である。

(4) 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。

今後の課題として、yz2 の拡布に関して、ネット上にて公開する他にも継続的な広報活動が必要であると思われる。

マルチスレッドによる圧縮速度の向上に関しては、実行するハードウェアの環境やデータ量に左右されやすいことが確認できた。マルチスレッドの有効性の判断するにはまだデータが少ないと思われる。今後もこのあたりの測定結果には注意していきたい。

PPMアルゴリズムの圧縮率の向上については、ほとんど変わらない結果となってしまった。頻度テーブルの初期値をデータに埋め込むことで若干の改善の余地があるかと思われるが、今後の課題としたい。

5．参加企業及び機関

株式会社 日立インフォメーションテクノロジー

有限会社 エレクトリックシーブ

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」, C Q 出版, ISBN4-7898-3672-X